

**Rationalisation of aluminium alloys
using machine learning and Artificial
Intelligence**

**A Thesis Submitted for the
Degree of Doctor of Philosophy**

By

Tanu Tiwari

**Brunel Centre for Advanced
Solidification Technology,
Brunel University London**

2025

DECLARATION BY CANDIDATE

I hereby declare that this thesis, submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, is my original work and has not been submitted previously, in whole or in part, for any degree or diploma at this or any other institution.

To the best of my knowledge, all sources of information used or quoted have been acknowledged appropriately. This work was carried out under the supervision of Prof. Dmitry Eskin and Dr. Chamini Mendis at the BCAST, Brunel University of London.

Tanu Tiwari

28 November 2025

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisors, **Prof. Dmitry Eskin** and **Dr. Chamini Mendis**, for their continuous support, guidance, and encouragement throughout my PhD journey. Their insight and expertise have been invaluable in shaping both this thesis and my development as a researcher.

I am especially thankful to **Prof. Hamid Assadi** and **Prof. Issac Chang**, for their fruitful discussions, which enhanced the direction of this research.

I would also like to extend my appreciation to my research group, colleagues, and the academic and technical staff of the **BCAST, Brunel University of London** for their support, collaboration, and the intellectually stimulating environment they provided.

Special thanks to my parents for their unwavering support, throughout this challenging and rewarding journey.

Finally, I gratefully acknowledge the financial support provided by Circular Metals Centre funded by an UKRI/EPSRC grant EP/V011804/1, which made this research possible.

ABSTRACT

Aluminium alloys are widely used across various sectors of engineering due to their lower density combined with higher strength compared to many existing alloys of other metals. These unique characteristics have led to an increased demand for and discovery of new aluminium alloys with targeted properties and compositions. Traditional methods of designing new materials with desired properties, such as trial-and-error and reliance on domain experts' experience, are time-consuming and expensive. These techniques also expand the search area for suitable alloys.

In this research, we propose a machine learning-based design system to reduce the number of grades across all series of age-hardenable and non-age-hardenable aluminium alloys. The system collects features based on chemical composition, mechanical properties, corrosion resistance, weldability, and thermal and electrical properties under different tempering and hardening conditions for machine learning modelling. A combination of PCA (Principal Component Analysis) and K-means clustering is applied for clustering and sub-clustering similar alloys based on their compositional and property profiles into clusters and sub-clusters. Next, an optimisation algorithm, namely a multi-property decision-making method, i.e., TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), identifies the optimum alloys within each sub-cluster. These selected alloys exhibit a balanced set of properties that effectively represent the range of characteristics found among other alloys in the same sub-cluster.

Subsequently, a recycling algorithm is applied to predict the mixing ratio based on closeness scores generated by the optimisation algorithm. This process mixes the optimum alloy in each sub-cluster with the remaining alloys in the sub-cluster, resulting in a single optimised alloy as determined by the optimisation algorithm. This method significantly reduces the number of alloy grades while maintaining key material properties and enhancing recyclability, which has a metallurgical basis.

This design system is enhanced and developed into a dedicated recycling software application, offering a practical tool for the aluminium industry. It supports sustainable development and improves recycling efficiency, aligning alloy manufacturing with the principles of the circular economy.

LIST OF PUBLICATIONS

The following publications have resulted from the research presented in this thesis:

- Tiwari, T., Jalalian, S., Mendis, C., & Eskin, D. Classification of T6 Tempered 6XXX Series Aluminum Alloys Based on Machine Learning Principles. JOM 75, 4526–4537 (2023). <https://doi.org/10.1007/s11837-023-06025-9>.
- Tiwari, T., Mendis, C. & Eskin, D. Facilitating Recycling of 6xxx Series Aluminum Alloys by Machine Learning-Based Optimisation. J. Sustain. Metall. 11, 2323–2334 (2025). <https://doi.org/10.1007/s40831-025-01112-4>.
- Tiwari, T., Mendis, C., & Eskin, D. (**Under review**). Machine Learning-Based Recycling Strategies for Internal Industrial Scrap of Aluminum Alloys: Focus on 3xxx, 5xxx, and 6xxx series alloys (2025).

LIST OF CONFERENCES AND SYMPOSIUMS

The following is the list of conferences attended during this research:

- **LIME Conference, BCAST 2023:** Speaker and poster on “Accelerated discovery of light alloys using ML/AI”.
- **Brunel Poster Conference, Brunel university of London 2023:** Highest voted poster on “Accelerated discovery of light alloys using ML/AI”.
- **FEMS Junior EUROMAT Conference, Manchester 2024:** Oral presentation on “Optimisation of 6XXX series aluminium alloys using ML and recyclability principles”.
- **BCAST PGR Symposium, BCAST (2023–2025):** Annual speaker on “Rationalisation of aluminium alloys using ML/AI”.

LIST OF TABLES

Table 2.1. Physical properties of aluminium [1]	4
Table 2.2. Types of wrought alloy series	10
Table 2.3. Wrought alloy series – composition (wt%) and yield strength (MPa) ranges [3], [7].....	11
Table 2.4. Libraries used in project.....	21
Table 3.1. Silhouette Scores for all Clustering Algorithm comparison	71
Table 5.1. Composition and property variability across clusters	118
Table 5.2. Actual and normalised values (in parentheses) of properties of alloys within sub-cluster 01 of Cluster 0	122
Table 5.3. Properties and applications of optimum alloys in each cluster	127
Table 5.4. Final recycled alloy composition and properties for sub-cluster 01	131
Table 5.5. Comparison of predicted vs. reported properties of optimum alloys for Sub-cluster 01.....	132
Table 6.1. Range of properties and compositions in low Cu alloys.....	145
Table 6.2. Actual and normalised (in parentheses) values of properties of low-Cu alloys within Sub-cluster 01 of Cluster 0	147
Table 6.3. Optimum low-Cu 7xxx series alloys and their corresponding applications	150
Table 6.4. Final Composition and Properties of the Recycled Alloy (Sub-cluster 01).....	152
Table 6.5. Comparison of predicted vs reported properties of optimum alloys for Sub-cluster 01 of cluster 0 for low Cu containing alloys.....	153
Table 6.6. Range of properties and compositions in high Cu alloys.....	160
Table 6.7. Optimum high-Cu 7xxx series alloys and their corresponding applications	163
Table 7.1. Chemical compositions and properties range's for 2xxx, 3xxx and 5xxx series data	173
Table 7.2. Optimum alloys for 2xxx series aluminium alloys	175
Table 7.3. Mixing ratios and key property improvements within the sub-clusters.....	177
Table 7.4. Optimum alloys for 5xxx series aluminium alloys	185
Table 7.5. Mixing ratios of 5xxx series alloys	186
Table 7.6. Optimum alloy for 3xxx series aluminium alloys.....	192
Table 7.7. Mixing ratios of 3xxx series sub-clusters.....	193
Table 7.8. Composition and property ranges for Set-1 alloys.....	198
Table 7.9. Composition and property ranges for Set-2 alloys.....	198
Table 7.10. Composition and property ranges within each cluster in Figure 7.4.....	200
Table 7.11. Composition and property ranges of alloys within each cluster in Figure 7.5...	201
Table 7.12. Comparison of optimised recycled alloys from Set-1 and Set-2.....	203
Table 8.1. USER defined 6xxx series alloys	210
Table 9.1. Reduced Set of Optimum Aluminium Alloys Across All Series.....	228
Table 11.1. Composition and properties range in 2xxx series clusters	257
Table 11.2. Property and composition ranges in clusters.....	264
Table 11.3. Composition and properties range's in 3xxx series clusters	269
Table 11.4. Preliminary Analysis Raw Dataset Overview	272
Table 11.5. Preliminary Analysis Processed Dataset Overview	273
Table 11.6. Non-Age-Hardenable Processed Dataset Overview	277
Table 11.7. Age-Hardenable Alloys Raw Dataset Overview.....	280
Table 11.8. 6xxx Series Raw Dataset Overview	282
Table 11.9. 6xxx Series Processed Dataset Overview	282
Table 11.10. 7xxx Series Raw Dataset Overview	287
Table 11.11. 7xxx Low-Cu Processed Dataset Overview.....	287

Table 11.12. 7xxx High-Cu Processed Dataset Overview	289
Table 11.13. Parameter values for other aluminium alloy series	293
Table 11.14. Phase formation in aluminium alloy series	298

LIST OF FIGURES

Figure 2.1. Ashby diagram: strength vs cost [2]	5
Figure 2.2. Al-Mg-Si phase diagram isopleths: (a) Al-xMg-2.4Si alloy, with Si fixed at 2.4 wt.% and Mg varying; (b) Al-7.5Mg-xSi alloy, with Mg fixed at 7.5 wt.% and Si varying [11].....	13
Figure 2.3. The main steps in ML	22
Figure 2.4. Box plot	26
Figure 2.5. Dendrogram illustrating the agglomerative hierarchical clustering process for five observations. The y-axis represents the linkage distance at which clusters are combined. Observations 1 and 2 form the first cluster, observations 3 and 4 merge next, and observation 5 subsequently joins this cluster. Applying a cut-off distance of approximately 6 (dashed line) yields two final clusters: {1,2} and {3,4,5}	35
Figure 2.6. Conceptual illustration of DBSCAN clustering and cluster validity distance measures. The ϵ -radius defines the local neighborhood used to classify points as core, border, or noise. For a representative data point, the intra-cluster distance (a_i) denotes the mean distance to all other points within the same cluster (cluster cohesion). The arrows labeled p_1 , p_2 , and p_3 illustrate example distances from the selected point to points in the nearest neighboring cluster; in practice, distances to all points in that neighboring cluster are considered when computing the inter-cluster distance (b_i), defined as their mean. These measures are commonly used to evaluate cluster cohesion and separation (i.e., silhouette analysis; described in Section 2.2.3.7.1).	37
Figure 2.7. Artificial neural network	42
Figure 3.1. Preliminary analysis of algorithms for methodology selection	67
Figure 3.2. Clustering results for 6xxx series alloys based on chemical composition (Si, Fe, Cu, Mn, Mg) and mechanical properties at T6 tempering condition (yield strength, UTS, elongation, and modulus of toughness). (a) K-means clustering applied to the standardised multi-dimensional dataset; the axes represent abstract scaled feature dimensions and do not correspond to a single physical parameter. (b) PCA + K-means clustering, where Principal Component 1 and Principal Component 2 are linear combinations of the same composition and mechanical variables. The plots therefore represent clustering based on all variables simultaneously	71
Figure 3.3. Hierarchical clustering: Dendrogram diagram	73
Figure 3.4. DBSCAN clustering	74
Figure 3.5. Clustering of data using (a) Gaussian Mixture Models (GMM), and (b) Mean Shift algorithm	76
Figure 3.6. MCDMA algorithms comparison.....	77
Figure 3.7. Convergence of Bayesian optimisation showing predicted yield strength (MPa) versus iteration number for Cluster 2 (6xxx series) alloys. The objective function corresponds to yield strength predicted by the OLS-based regression model within the defined compositional bounds.	79
Figure 3.8. Iterative Bayesian optimisation trajectory showing predicted yield strength (MPa) versus iteration number for Cluster 2, illustrating sequential model updates and convergence toward the optimal composition.....	80
Figure 3.9. Training loss (mean squared error, MSE) versus epoch number for the artificial neural network (ANN) models trained separately on each PCA-K-means-derived alloy cluster. The curves show the progressive reduction in MSE during backpropagation using the Adam optimiser, indicating stable convergence behaviour across all five clusters despite the limited dataset size.	82

Figure 3.10. Design framework	83
Figure 4.1. (a) PCA-plot, (b) Elbow plot and (c) Silhouette plot for all age-hardenable alloy series	97
Figure 4.2. Clustering visualisation of age-hardenable alloy series	98
Figure 4.3. (a) PCA-plot, (b) Elbow plot, and (c) Silhouette plot for all non-age-hardenable alloy series	100
Figure 4.4. Clustering visualisation of non-age-hardenable alloy	101
Figure 5.1. Histogram analysis	106
Figure 5.2. Outlier detection in dataset	107
Figure 5.3. Pair-wise analysis of dataset	109
Figure 5.4. Correlation matrix of properties	110
Figure 5.5. Scatter plot: (a) Yield strength vs Cu and (b) Yield strength vs Mg:Si	112
Figure 5.6. Violin plot for yield strength distribution	113
Figure 5.7. PCA variance plot (a), Elbow plot (b), and Silhouette plot (c) for determining the optimal number of clusters	115
Figure 5.8. Clustering results for (a) T5, (b) T6, and (c) T7 tempering conditions	117
Figure 5.9. Sub-clustering analysis of clusters: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, (e) Cluster 4.	122
Figure 5.10. TOPSIS closeness scores of alloys	127
Figure 5.11. LIME plots of YS and modulus of toughness: (a) Cluster 2, (b) Cluster 0, (c) Cluster 1, (d) Cluster 3, and (e) Cluster 4	134
Figure 5.12. Volume phase formation within all cluster of 6xxx series	137
Figure 6.1. A PCA variance plot (a), Elbow plot (b), and Silhouette plot (c), for determining the optimal number of clusters in low-Cu 7xxx series alloys	142
Figure 6.2. Clustering of low Cu 7xxx series aluminium alloys at (a) T5, (b) T6, and (c) T7 tempering condition.	144
Figure 6.3. Sub-clusters of low-Cu 7xxx series alloy: (a) Cluster 0, sub-clusters, and (b) Cluster 1, sub-clusters, and (c) Cluster 2, sub-clusters	147
Figure 6.4. TOPSIS closeness scores of alloys A, B, C, and D	150
Figure 6.5. LIME results for YS and modulus of toughness for (a) Cluster 0, (b) Cluster 1, and (c) Cluster 2	155
Figure 6.6. PCA variance plot (a), Elbow plot (b), and (c) Silhouette score plot for high Cu 7xxx series alloys	157
Figure 6.7. Clustering of high Cu 7xxx series aluminium alloys at (a) T5, (b) T6, and (c) T7 tempering condition	159
Figure 6.8. Sub-clusters of high-Cu 7xxx series alloy: (a) Cluster 0, sub-clusters, (b) Cluster 1, sub-clusters, (c) Cluster 2, sub-clusters, and (d) Cluster 3, sub-clusters	162
Figure 6.9. LIME results for (a) Cluster 0 (Medium properties), (b) Cluster 1 (Low properties), and (c) Cluster 2 (High properties)	165
Figure 7.1. Thermo-Calc analysis of 2xxx series clustering	184
Figure 7.2. Thermo-Calc analysis of clusters of 5xxx series alloys	191
Figure 7.3. Thermo-Calc analysis of clustering of 3xxx series	197
Figure 7.4. Clustering of mixed scrap aluminium alloys	199
Figure 7.5. Clustering of optimum alloys of 3xxx, 5xxx, and 6xxx series	201
Figure 8.1. Design of recycling software application	207
Figure 8.2. Interface for input of (a) Number of alloys, (b) Chemical composition and Properties of alloy 1, (c) Chemical composition and Properties of alloy 2 (d) Chemical composition and Properties of alloy 3	211
Figure 8.3. Tempering condition selection: T6	212

Figure 8.4. Recommended recycled alloy: (a) Composition & (b) Mixing ratios and properties and similar alloy grade.....	213
Figure 8.5. Figure 8.57: Demonstration of the extended alloy recycling software application interface: (a), (b) & (c) Composition of the USER input alloys, (d) Mixing ratios of each USER input alloys, (e) Recycled alloy composition, and (f) Recycled alloy properties.....	218
Figure 11.1. Processing conditions 6xxx series: (a) Forged, (b) Rolled, (c) Sheet, and (d) Plate.....	254
Figure 11.2. Plots to detect the number of clusters in 2xxx-series alloys (a) PCA variance plot, (b) an Elbow plot and, (c) Silhouette score plot	256
Figure 11.3. Clustering analysis of 2xxx series alloys.....	257
Figure 11.4. Sub-clustering analysis of 2xxx series alloys: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, (e) Cluster 4, and (f) Cluster 5	261
Figure 11.5. Plots to detect the number of clusters in 5xxx-series alloys (a) PCA variance plot (b) an Elbow plot and, (c) Silhouette score	263
Figure 11.6. Clustering of 5xxx series alloys.....	264
Figure 11.7. Sub-clustering of 5xxx series alloys: (a) Cluster 0, (b) Cluster 1, and (c) Cluster 2.....	266
Figure 11.8. Plots to detect the number of clusters in 3xxx-series alloys (a) PCA variance plot, (b) an Elbow plot and, (c) Silhouette score	268
Figure 11.9. Clustering of 3xxx series alloys.....	269
Figure 11.10. Sub-clustering of 3xxx series alloys: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, and (d) Cluster 3.....	272

TABLE OF CONTENTS

DECLARATION BY CANDIDATE	ii
ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
LIST OF PUBLICATIONS	v
LIST OF CONFERENCES AND SYMPOSIUMS.....	vi
LIST OF TABLES.....	vii
LIST OF FIGURES	ix
TABLE OF CONTENTS.....	xii
CHAPTER 1: INTRODUCTION.....	1
1.1 Project Background.....	1
1.2 Main Goals of the Project	2
1.3 Key Objectives of the Study	2
1.4 Outline of the Thesis.....	3
CHAPTER 2: LITERATURE REVIEW.....	4
2.1 Aluminium Alloys Overview.....	4
2.1.1 General Characteristics and Importance of Aluminium and Aluminium Alloys [1].....	4
2.1.2 Effect of Alloying Elements and Impurities in Aluminium Alloys [1], [3].....	6
2.1.3 Classification of Aluminium Alloys	10
2.1.4 Processing Conditions in Aluminium Alloys [1], [12]	14
2.1.5 Effects of Hardening and Tempering Treatments in Aluminium Alloys [1], [10]	15
2.1.6 Recyclability and Related Issues in Aluminium Alloys	17
2.1.7 Need for Optimisation and Reducing the Number of Alloy Grades.....	18
2.2 Machine Learning (ML) and Artificial Intelligence (AI)	19
2.2.1 Overview of ML/AI	19
2.2.2 Types of Machine Learning [40], [46].....	23
2.2.3 Machine Learning Algorithms.....	24
2.2.3.1 Exploratory Data Analysis (EDA) Techniques [48]	24
2.2.3.2 Feature Engineering [49]	25
2.2.3.2.1 Handling Outliers [50]	25
2.2.3.2.2 Scaling Techniques [51], [52]	26
2.2.3.3 Feature Selection Techniques: Correlation Analysis [53]	27
2.2.3.4 Regression Algorithm [54]	28
2.2.3.5 Classification Algorithms [57],[56].....	29
2.2.3.6 Dimensionality Reduction Algorithm: Principal Component Analysis (PCA) [58]	31
2.2.3.7 Clustering Algorithm.....	32

2.2.3.7.1 K-means Clustering Algorithm [59], [60].....	33
2.2.3.7.2 Hierarchical Cluster Analysis (HCA) [60].....	34
2.2.3.7.3 Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [61]	36
2.2.3.7.4 Comparative Discussion of Clustering Approaches.....	37
2.2.3.7.5 Gaussian Mixture Model (GMM) [62].....	38
2.2.3.7.6 Mean Shift Clustering [62].....	39
2.2.3.8 Optimisation Algorithms	39
2.2.3.8.1 Bayesian Optimisation Algorithm [63]	39
2.2.3.8.2 Artificial Neural Network (ANN) [47], [64].....	41
2.2.3.8.3 TOPSIS (Technique for Order Preference by Similarity to the Ideal Solution) [65], [66].....	43
2.2.3.8.4 AHP (Analytic Hierarchy Process) [67].....	45
2.2.3.8.5 VIKOR (VIšekriterijumsko KOmpromisno Rangiranje) [67]	46
2.2.3.9 Explainable AI technique: LIME Algorithm (Local Interpretable Model-agnostic Explanations) [68]	47
2.3 Implementation of Machine Learning (ML) and Artificial Intelligence (AI) in Materials Science	48
2.4 ML/AI Implementation in Aluminium Alloy Development.....	61
2.5 Conclusions	64
2.5.1 Aluminium and Aluminium Alloys, and Recyclability Demands	64
2.5.2 Machine Learning/Artificial Intelligence	65
CHAPTER 3: METHODOLOGY OF STUDY	67
3.1 Preliminary Analysis for Model Selection.....	67
3.2 Design Framework.....	82
3.3 Data Collection and Pre-processing	83
3.3.1 Data Collection	83
3.3.2 Data Pre-processing	84
3.4 Combined Approach of PCA and K-means Clustering	86
3.5 Sub-clustering of Clusters.....	86
3.6 Optimisation Algorithm	88
3.7 Alloy Mixing in Clusters: Recycling Algorithm	89
3.8 LIME Algorithm	91
CHAPTER 4: RESULTS AND DISCUSSION: HYBRIDISING ALL ALUMINIUM ALLOY SERIES: VALIDATING CLUSTERING ALGORITHMS.....	92
4.1 Data Collection	92
4.1.1 2xxx, 6xxx, 7xxx Age-Hardenable Alloy Series	92
4.1.2 1xxx, 3xxx, 5xxx and 8xxx Non-Age-Hardenable Alloy Series	93

4.2 Data Cleaning and Pre-processing	94
4.2.1 Predicting Missing Values in the Dataset	94
4.2.2 Data Pre-processing	94
4.3 Combined Approach of PCA and K-means Clustering	95
4.3.1 Clustering of all Age-Hardenable Alloy Series	95
4.3.2 Clustering of All Non-Age-Hardenable Alloy Series	98
4.4 Interpreting Hybridisation Results: New Classification Based on Properties	101
4.4.1 Age-Hardenable Alloys Series Hybridisation: Interpretation.....	101
4.4.2 Non-Age-Hardenable Alloys Series Hybridisation: Interpretation.....	102
4.5 Conclusion: Hybridisation of Age-Hardenable and Non-Age Hardenable Alloy Series	103
CHAPTER 5: RESULTS AND DISCUSSION – OPTIMISATION OF 6xxx SERIES	
ALUMINIUM ALLOYS [146], [147].....	104
5.1 Data Collection and Pre-processing.....	104
5.2 Data Visualisation to Understand the Dataset	105
5.3 Combined Approach of PCA and K-means Clustering: Clustering of 6xxx Series Dataset.....	113
5.4 Sub-clustering of Clusters.....	119
5.5 Optimisation Algorithm.....	122
5.6 Alloy Mixing within Sub-clusters: Recycling Algorithm.....	128
5.7 Validation of Mixed Alloy Composition	131
5.8 Understanding the Clustering Process: LIME Algorithm.....	132
5.9 Metallurgical Reasoning behind Clustering.....	134
5.10 Thermo-Calc Analysis as a Basis for Cluster Classification	136
5.11 Conclusion on 6xxx Series Optimisation.....	137
CHAPTER 6: RESULTS AND DISCUSSION: OPTIMISATION OF 7xxx-SERIES	
ALUMINIUM ALLOYS	139
6.1 Data Collection	139
6.2 Optimisation of Low-Cu 7xxx Series Aluminium Alloys	140
6.2.1 Clustering of Low Cu 7xxx Series Alloys	140
6.2.2 Sub-clustering of Clusters.....	145
6.2.3 Optimum Alloy Selection Algorithm.....	147
6.2.4 Alloy Mixing Within Sub-clusters: Recycling Algorithm.....	151
6.2.5 Validation of Recycled Alloy Properties	153
6.2.6 LIME Analysis to Explain the Clustering Results.....	154
6.3 Optimisation of High Cu-Containing 7xxx Series Alloys	155
6.3.1 Clustering of High Cu-Containing 7xxx Series Alloys	156
6.3.2 Sub-clustering of High Cu Alloys Clusters	160

6.3.3 Optimum Alloy Selection and Recycling Analysis	162
6.3.4 Understanding the Clustering Process: LIME Algorithm.....	164
6.4 Metallurgical Reasoning Behind Clustering of 7xxx Series Alloys	165
6.4.1 Reasoning for Low-Cu 7xxx Series Alloys	166
6.4.2 Reasoning Behind High-Cu 7xxx Series Alloys.....	168
6.5 Conclusions on 7xxx Series Optimisation	170
CHAPTER 7: CASE STUDY – OPTIMISATION OF 2xxx, 3xxx, AND 5xxx SERIES	
ALUMINIUM ALLOYS AND MIXED SCRAP RECYCLING COMPARISON	172
7.1 Data Collection for 2xxx, 3xxx, and 5xxx Series Aluminium Alloys.....	173
7.2 Optimisation of 2xxx series Aluminium Alloys	174
7.2.1 Optimum Alloy Selection in 2xxx Series	174
7.2.2 Recycling Algorithm for Alloy Mixing of 2xxx Series Alloys	176
7.2.3 Metallurgical Reasoning Behind Clustering of 2xxx Series Alloys	178
7.2.4 Thermo-Calc Analysis as a Basis for Cluster Classification	183
7.3 Optimisation of 5xxx Series Aluminium Alloys	184
7.3.1 Optimum Alloys Selection for the 5xxx Series Alloys.....	184
7.3.2 Recycling Algorithm for Alloy Mixing of 5xxx Series Alloys	186
7.3.3 Metallurgical Reasoning Behind Clustering of 5xxx Series Alloys	187
7.3.4 Thermo-Calc Analysis as a Basis for Cluster Classification	190
7.4 Optimisation of 3xxx Series Aluminium Alloys	191
7.4.1 Optimum Alloy Selection for the 3xxx Series.....	191
7.4.2 Recycling Algorithm for Alloy Mixing of the 3xxx Series Alloys	193
7.4.3 Metallurgical Reasoning Behind Clustering of 3xxx Series Alloys	193
7.4.4 Thermo-Calc Analysis as a Basis for Cluster Classification	197
7.5 Comparative Study – Internal Industrial Scrap Alloys vs Optimised Alloy Recycling	197
7.5.1 Data Collection	198
7.5.2 Alloys Sorting for Both Sets 1 and 2	199
7.5.2.1 Alloy Sorting of Set 1: Mixed Internal Industrial Scrap Alloys	199
7.5.2.2 Alloy Sorting of Set-2: Optimum Alloys.....	201
7.5.3 Recycling Comparison between Set-1 and Set-2 Alloys	202
7.6 Conclusion on Optimisation of 2xxx, 3xxx, and 5xxx Series Alloys and Mixed Scrap Recycling Comparison.....	203
CHAPTER 8: PRACTICAL APPLICATION: ALUMINIUM ALLOY RECYCLING SOFTWARE 206	
8.1 Overview of Recycling Software Application.....	206
8.2 Design of Recycling Software Application	207
8.3 Interface of the 6xxx Series-Based Alloy Recycling Software Application.....	207

8.3.1 User Input Parameters.....	207
8.3.2 Calculation of Alloy Mixing Ratios, Composition, and Properties.....	208
8.3.3 Similar Alloy Grade Search.....	209
8.4 Demonstration of the Alloy Recycling Software Application for 6xxx Series	210
8.4.1 User Input Compositions, Properties, and Processing Conditions	210
8.4.2 Optimum Alloy Calculation.....	212
8.4.3 Mixing Ratio Calculation.....	213
8.4.4 Finding the Closest Standard Alloy Grade	213
8.5 Extension of Recycling Software Application to Mixed-Series Recycling.....	214
8.5.1 Model Enhancements and Performance.....	214
8.5.2 User Interface and Demonstration	216
8.6. Conclusion on Alloy Recycling Software Application.....	218
CHAPTER 9: SUMMARY, CONCLUSIONS AND FUTURE WORK.....	220
9.1 Conclusions.....	220
9.1.1 It is Feasible to Reduce Standard Alloy Grades Across Several Alloy Series	220
9.1.2 Machine Learning Algorithms are Effective for Optimising Alloy Grade Systems	221
9.1.3 Optimisation Framework Yields Robust and Representative Alloy Selections and Recycling	222
9.1.4 Metallurgical Interpretation Validates Data-Driven Results	225
9.1.5 Summary of Selected Optimal-Performing Alloys Across Aluminium Series	228
9.1.6 Validation of Optimisation Framework and Mixed-Series Recycling	229
9.1.7 User-Friendly Recycling Software Application Enables Optimisation for Practical Use	230
9.2 Challenges and Limitations.....	231
9.3 Future Work.....	232
9.4 Practical Implications.....	233
9.5 Concluding Comments.....	233
CHAPTER 10: REFERENCES	235
CHAPTER 11: APPENDIX.....	253
11.1 Graphs for Processing Conditions of 6xxx Series Aluminium Alloys	253
11.2 Supplementary Material: Clustering and Sub-Clustering Analysis for the Case Study Chapter.....	255
11.2.1 Clustering and Sub-clustering of 2xxx Series Alloys.....	255
11.2.2 Clustering and Sub-clustering of 5xxx Series Alloys.....	261
11.2.3 Clustering and Sub-clustering of 3xxx Series Alloys.....	266
11.3 Datasets and Reproducibility Workflow.....	272
11.3.1 Dataset 1: Preliminary Analysis.....	272

11.3.1.1 Raw Data	272
11.3.1.2 Processed Data	273
11.3.1.3 Reproducibility Notes for Preliminary Analysis	274
11.3.2 Dataset 1: Non-Age-Hardenable Alloys	277
11.3.2.1 Raw Data	277
11.3.2.2 Processed Data	277
11.3.2.3 Reproducibility Notes for Non-Age-Hardenable Alloys	278
11.3.3 Dataset 2: Age-Hardenable Alloys	280
11.3.3.1 Raw Data	280
11.3.3.2 Processed Data	281
11.3.3.3 Reproducibility Notes	281
11.3.4 Dataset 3: 6xxx Series Alloys	281
11.3.4.1 Raw dataset	282
11.3.4.2 Raw dataset	282
11.3.4.3 Reproducibility Notes for 6xxx Series Alloys	284
11.3.5 Dataset 3: 7xxx Series Alloys	286
11.3.5.1 Raw Dataset	287
11.3.5.2 Processed Dataset – Low-Cu 7xxx Series	287
11.3.5.3 Processed Dataset – High-Cu 7xxx Series	289
11.4 Codes for the Project	290
11.4.1 Optimisation Algorithm	291
11.4.1.1 Combined Approach of PCA and K-means clustering	291
11.4.1.2 Sub-clustering and Optimisation Algorithm	293
11.4.1.3 Recycling Algorithm	295
11.4.2 Metallurgical Reasoning behind Clustering	297
11.4.2.1 LIME Analysis on Clusters	297
11.4.2.2 Thermo-Calc Validation of Clusters	298
11.4.3 Recycling Software Application Code for Sorted 6xxx series alloys	299
11.4.4 Extended Recycling Software Application Code for Sorting and Recycling of 2xxx, 3xxx, 5xxx, 6xxx, 7xxx, 8xxx	302

CHAPTER 1:INTRODUCTION

1.1 Project Background

Traditionally, a hill-climbing approach has been employed in the development and design of aluminium alloys, leading to gradual improvements over time. Drawing on lessons learned from human experience and practical applications, aluminium alloy design has evolved into an iterative process. However, extensive alloy development in aluminium metallurgy remains labour-intensive, costly, and often underutilises the vast amount of existing data on alloy properties. Testing billions of potential alloy combinations is clearly impractical.

Empirical alloy discovery through conventional techniques, such as mixing alloying elements and analysing their structures using methods like X-ray diffraction (XRD) and scanning electron microscopy (SEM), is both lengthy and tedious. A classic example of this empirical progression can be seen in the evolution from Al–Cu alloys to Al–Zn–Mg, Al–Zn–Mg–Cu, and Al–Cu–Li alloys, driven by the changing demands of structural applications.

When the potential solution space becomes extremely large, rapid fabrication and testing are not feasible. The likelihood of producing alloys with the desired properties decreases, making it essential to reduce the solution space. Consequently, the number of alloys within each aluminium series should be condensed into a smaller set of optimised compositions. This would not only improve overall performance but also enhance recycling efficiency by simplifying material separation and reuse.

Promising solutions lie in advanced computational techniques such as machine learning (ML), artificial intelligence (AI), and deep learning (DL). Machine learning algorithms capable of learning patterns and constructing statistical models that analyse data and predict future outcomes. These algorithms can learn from existing datasets and establish inference models that describe the relationship between targeted properties and material descriptors. As a result, the properties of unexplored alloys can be estimated, enabling rational experimental design and more accurate predictions.

By leveraging large datasets on aluminium alloys in combination with ML and AI algorithms, it becomes possible to consolidate the wide classifications of aluminium alloys into a smaller number of grades with optimised compositions and superior properties.

1.2 Main Goals of the Project

The primary goal of this research is the development and application of ML and AI algorithms to optimise aluminium alloy compositions. This optimisation aims to facilitate full metal circulation and slow down the resource loop by reducing the number of alloy grades while maintaining performance.

To achieve this, the research addresses the following key aims:

- Development and application of techniques and algorithms for determining optimised alloy compositions, simplifying alloy systems by reducing the number of grades within each system.
- Application of these approaches to standardise alloy compositions, using commonly available alloying elements and avoiding elements that limit recyclability. Explore recycling options.
- Validation of the developed approaches through reference data, considering alloy composition, thermomechanical history, performance levels, and fields of application.

1.3 Key Objectives of the Study

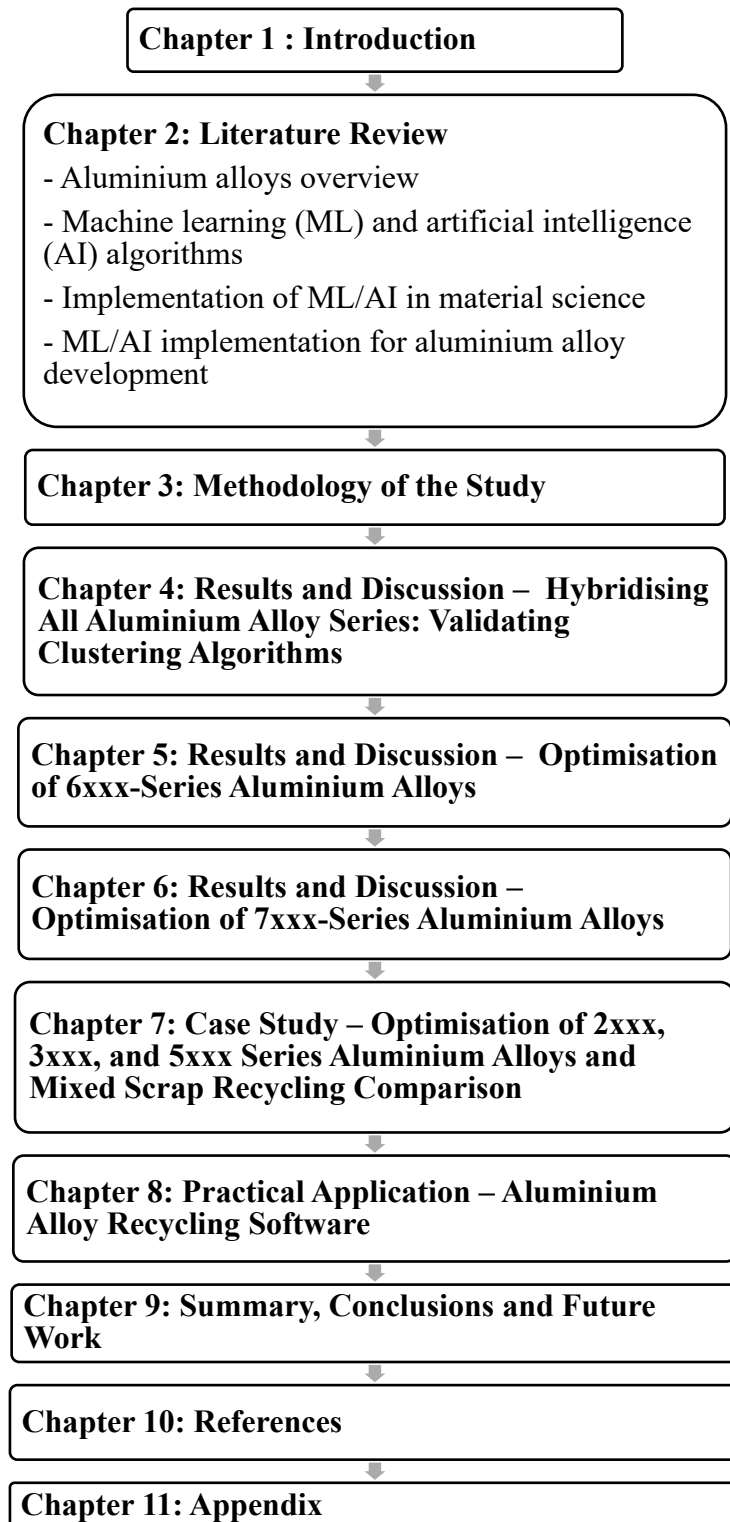
The specific objectives of this research are:

- **Data Collection:** To develop large datasets for each aluminium alloy series (1xxx–8xxx) based on chemical composition, properties, thermomechanical processing, and processing history, by compiling all possible grades from standards, literature, and reliable online sources.
- **Alloy System Simplification:** To condense the dataset for each alloying system into a smaller number of representative classes and sub-classes with equivalent properties, enabling machine learning (ML) models to predict the compositions and properties of optimised alloys.
- **Optimisation and Recycling:** To develop a new algorithm based on the recyclability principle, facilitating controlled mixing of alloys within the same sub-classes to achieve

properties comparable to those of the optimal alloys. The reduced set of optimum alloys should collectively cover the same property range as the larger dataset.

- **Validation:** To validate the optimised alloys using an independent (unseen) dataset.

1.4 Outline of the Thesis



CHAPTER 2: LITERATURE REVIEW

2.1 Aluminium Alloys Overview

2.1.1 General Characteristics and Importance of Aluminium and Aluminium Alloys [1]

Aluminium and its alloys combine low weight with high strength making them highly versatile and economically attractive for numerous applications. They range from extremely ductile, soft foils used in packaging to alloys capable of withstanding demanding engineering conditions. Following steel, aluminium alloys are among the most widely used structural metals. Table 2.1 summarises the general physical properties of aluminium.

Table 2.1. Physical properties of aluminium [1]

Property	Value
Atomic Number	13
Atomic Weight (g/mol)	26.98
Valency	3
Crystal Structure	FCC
Melting Point (°C)	660.2
Boiling Point (°C)	2480

Although aluminium is generally more expensive than steel, as shown by the blue arrow in the Ashby plot (Figure 2.1) [2], its combination of properties is particularly advantageous for applications where mobility and weight reduction are critical, including aerospace, marine, and various land vehicles [1].

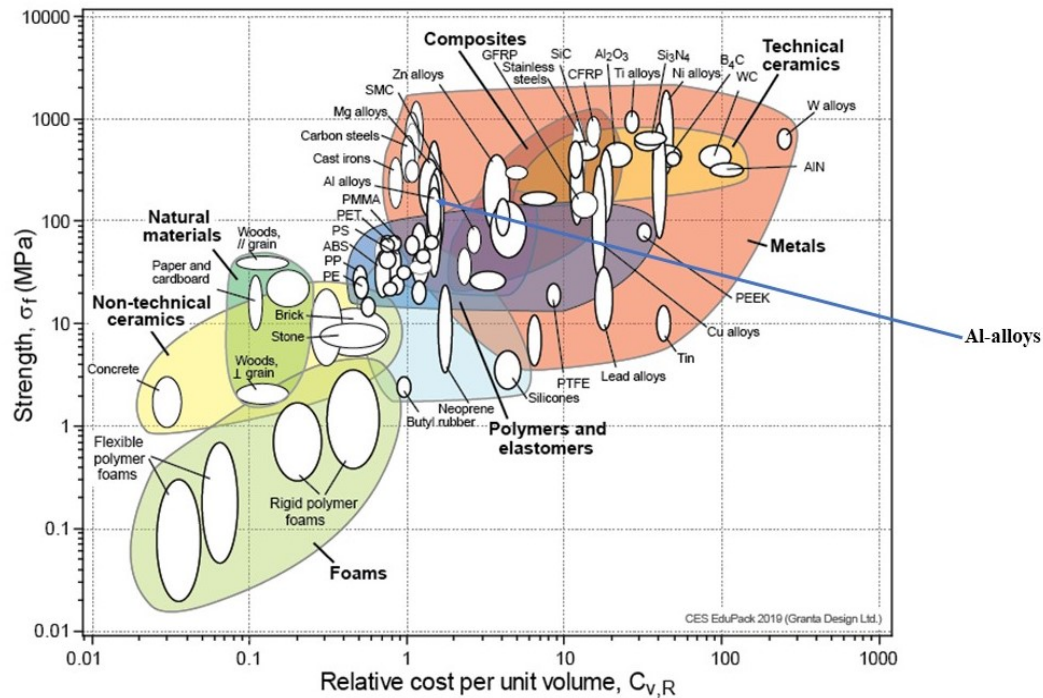


Figure 2.1. Ashby diagram: strength vs cost [2]

Aluminium exhibits excellent resistance to corrosion, which is a problem in steel and causes it to rust away. This resistance in aluminium arises from the spontaneous formation of a thin, transparent, and colourless aluminium oxide layer on its surface when exposed to oxygen. This layer adheres to the surface and resists further corrosion. Aluminium also shows strong resistance to various salts, water, and a wide range of environmental and chemical agents. It also possesses high thermal and electrical conductivity, although certain alloys also exhibit high electrical resistance. Due to its non-ferromagnetic nature, it is particularly suitable for electronic and electrical applications. Aluminium is also non-pyrophoric, except in fine powdered form, which makes it ideal for environments where flammability is minimized. It has wide applications in the food and beverage packaging industry due to its non-toxic nature, and it is often coated with a thin polymer layer for additional protection. Aluminium has a naturally attractive appearance with a soft, lustrous, shiny finish, and when anodised, it can acquire almost any colour or a bright metallic surface [1].

Due to its ease of fabrication, it can be cast using standard foundry techniques, rolled or thinned down to a thickness even thinner than paper, hammered, forged, or extruded into almost any shape or profile. These exceptional properties of aluminium allow it to compete with less

expensive materials, which show limited workability, making it the most widely used metal in a wide range of industries [1].

2.1.2 Effect of Alloying Elements and Impurities in Aluminium Alloys [1], [3]

The properties of the aluminium alloys are governed by the addition of the alloying elements. The primary alloying elements added to aluminium are copper, magnesium, silicon, zinc, and manganese. In addition to these, additional alloying elements and impurities are also added, including iron, chromium, titanium, nickel, lithium, silver, zirconium, bismuth, and vanadium. Trace elements that are also present or intentionally added include boron and beryllium. These elements significantly govern the properties of aluminium alloys. In the present work, only these alloying elements are considered, as discussed below [1], [3].

Copper: Cu is a key alloying element in both wrought and cast aluminium alloys. With 2–7% Cu and the addition of other alloying elements, it forms the important families of aluminium alloys (i.e., 2xxx). These alloys respond to solution heat treatment and aging. These heat treatments enhance the strength and hardness of these alloys due to the formation of the metastable Al_2Cu phase. The maximum strength of aluminium alloys is achieved with 4–6% Cu; this value can vary depending on the presence of other alloying elements.

Copper-Magnesium: The addition of magnesium to aluminium-copper alloys provides significant benefits by increasing strength through the formation of an additional hardening phase, Al_2CuMg [4]. Solution heat treatment and quenching of magnesium-added alloys result in a greater increase in strength compared to Al-Cu alloys. In certain wrought alloys of this type, high ductility is observed alongside increased strength upon room-temperature aging. Artificial aging further increases yield strength but at the cost of reduced tensile elongation.

Copper-Magnesium and Additional Elements: Aluminium-copper-magnesium alloys that contain iron exhibit superior bearing performance, dimensional stability, and maintain high hardness and strength at elevated temperatures. In heat-treated alloys, even small amounts of iron, such as 0.5% in an Al-4%Cu-0.5%Mg wrought alloy, can reduce tensile properties by forming AlFeSi phases. AlFeSi phases will be formed upon solidification so will present even in non-heat-treated alloys. Role of Fe and Ni in Al-Cu-Mg alloys is to produce AlFeNi compounds (sometimes with Cu or Si) to improve high-temperature properties.

Gallium: Gallium is typically found in aluminium as a minor impurity at concentrations ranging from 0.001 to 0.02%. Its influence on the mechanical properties of aluminium is minimal. However, at a higher concentration of around 0.2%, gallium can alter the alloy's response to etching, affect corrosion behaviour, and modify certain brightening characteristics.

Boron: Boron is used in aluminium as a grain refiner and to improve conductivity by promoting the precipitation of elements such as Cr, Ti, V, and Mo. During solidification, boron can be used as a grain refiner at levels of 0.005 to 0.1%. When used with an excess of titanium, it becomes more effective. Commonly, titanium to boron is added in commercial grain refiners in 5-to-1 or 3-to-1 ratios.

Beryllium: To reduce oxidation at elevated temperatures, Be is used in aluminium alloys that contain magnesium. For the improvement of aluminium film adhesion, up to 0.1% Be is used in aluminising baths for steel, which helps resist deleterious iron-aluminium complex formation.

Bismuth: To create free-machining alloys, low-melting and soft metals such as lead, bismuth, and cadmium are added to aluminium. In solid aluminium, these elements have restricted solubility and form a low-melting, soft phase that helps lubricating the cutting tool and promotes chip breaking. Upon solidification, the expansion of bismuth compensates for the shrinkage of lead, which is a major advantage in avoiding casting defects. In some Al-Mg-Si and Al-Cu alloys, a 1-to-1 lead-bismuth ratio is used to counteract the detrimental effect of sodium on hot cracking. Small additions of bismuth (i.e., 20 to 200 ppm) can be made to aluminium-magnesium alloys.

Chromium: In commercially pure aluminium, chromium is present as a minor impurity, typically ranging from 5 to 50 ppm. The addition of chromium has a significant effect on the electrical resistivity of aluminium alloys. Higher chromium levels, particularly in the presence of other elements such as iron, manganese, and titanium, can lead to the formation of coarse constituent phases. It is commonly added to various alloys, including Al-Mg-Si, Al-Mg, and Al-Mg-Zn, usually not exceeding 0.35%. In these alloys Cr forms fine, dispersed phases with slow diffusion rates, which in turn inhibit nucleation and grain growth during recrystallisation.

Iron: In aluminium, the most common impurity is iron. In the molten state, iron has a higher solubility in aluminium, and therefore, at all molten stages of production, it can easily be

dissolved. However, in the solid state, the solubility of iron is very low (up to ~0.04%), and most of the iron appears as an intermetallic phase, present in combination with aluminium and often other alloying elements. There are, however, a number of casting and wrought alloys where Fe is added as an alloying element to promote high-temperature stability or to improve separation from the die in high-pressure die casting [5].

Magnesium: In the 5xxx series of alloys, magnesium is the major alloying element. The maximum solid solubility of magnesium in aluminium is about 17.4%, but in current wrought alloys, the magnesium content does not exceed 5.5%. The strength of aluminium is significantly increased by the addition of magnesium, without decreasing ductility. Additionally, the weldability and corrosion resistance of aluminium alloys are improved by magnesium additions.

Manganese: In primary aluminium, manganese is typically present at concentrations of 5 to 50 ppm and is a common impurity in aluminium alloys. Its addition reduces electrical resistivity, while in solid solution or as finely dispersed intermetallic phases, manganese significantly enhances alloy strength. Importantly, it does not negatively affect corrosion resistance. Although manganese has limited solubility in solid aluminium containing typical impurities, it remains in solution during chill casting. Consequently, most of the added manganese is largely retained, even in large ingots. Aluminium–manganese alloys form the basis of the 3xxx series.

Magnesium-Manganese: In the work-hardened condition, the presence of the Mg-Mn combination in wrought alloys provides high strength, good welding characteristics, and high resistance to corrosion. However, the difficulty of alloy fabrication increases with the amounts of either Mn or Mg, and during hot rolling, the risk of cracking rises, especially in the presence of trace amounts of sodium.

Magnesium-Silicon: Up to 1.5% each of silicon and magnesium are present in 6xxx series wrought alloys in an approximate ratio of 1.73:1 to form the Mg_2Si phase. Upon age hardening, very fine metastable precipitates are formed, which efficiently harden these alloys.

Molybdenum: Molybdenum occurs in aluminium as a very low-level impurity, typically ranging from 0.1 to 1.0 ppm. Historically, at concentrations around 0.3%, it was employed as a grain refiner and to modify iron-containing phases. However, molybdenum is no longer used for these purposes.

Lithium: As an impurity, lithium is typically present at a few ppm, usually less than 5 ppm. Under humid conditions, lithium can promote aluminium foil discolouration (blue corrosion). The presence of traces of lithium greatly increases molten aluminium's oxidation, and it also alters the surface characteristics of wrought aluminium products. However, there are alloy grades where lithium is the main alloying element, producing high strength, high elastic modulus, and low density, such as Al-Mg-Li and Al-Cu-Li alloys. These alloys are especially used in aerospace and military applications.

Nickel: In aluminium, nickel has a maximum solid solubility of 0.04% and typically exists as an insoluble intermetallic compound with copper and iron. While nickel decreases ductility, it enhances the strength of high-purity aluminium. Although binary aluminium–nickel alloys are not commonly used, a distinct class of aluminium–nickel casting alloys has been developed. At elevated temperatures, nickel is also added to aluminium–silicon and aluminium–copper alloys to improve strength and hardness, as well as to reduce their coefficient of thermal expansion.

Silicon: In electrolytic commercial-grade aluminium, silicon is present as the highest-level impurity (0.01 to 0.15%), after iron. At levels up to 1.5%, silicon is used with magnesium in heat-treatable 6xxx series wrought alloys to produce the Mg_2Si phase. Silicon is also the major component (3 to 18%) of most Al-based casting alloys.

Vanadium: In commercial-grade aluminium, vanadium is typically found at concentrations ranging from 10 to 200 ppm. Because it reduces electrical conductivity, vanadium is usually precipitated from alloys intended for conductive applications, often together with boron.

Zirconium: Zirconium (Zr), in the range of 0.1 to 0.3%, is added to form a fine intermetallic precipitate, Al_3Zr . This addition inhibits recrystallization and recovery, particularly in Al-Zn-Mg-Cu and some Al-Cu-Mg alloys.

Zinc and Its Alloy Systems: Aluminium–zinc alloys have long been used, but wrought forms are prone to stress-corrosion cracking, while casting alloys are susceptible to hot cracking, limiting their applications. The addition of magnesium to aluminium–zinc alloys significantly enhance their strength, particularly in compositions containing 3–7.5% zinc. This improvement is largely due to the formation of the $MgZn_2$ phase, which responds strongly to heat treatment and contributes to increased tensile and yield strengths. In cold-water-quenched 1.6 mm sheets,

the MgZn_2 content ranging from 0.5% to 12% progressively raises strength, with further improvements observed when magnesium is added in excess of the stoichiometric amount required to form MgZn_2 .

Incorporating copper, together with minor amounts of manganese and chromium, into the aluminium–zinc–magnesium system produces some of the highest-strength commercially available aluminium alloys. In this system, the aging response is largely governed by magnesium and zinc, while copper accelerates aging by increasing supersaturation and promoting the nucleation of phases such as CuMgAl_2 and Al_6CuMg_4 . The presence of copper also increases quench sensitivity during heat treatment and generally reduces resistance to general corrosion and stress-corrosion cracking. Minor alloying additions, including zirconium and chromium, can further influence both mechanical properties and corrosion resistance, optimising performance for demanding applications [3].

2.1.3 Classification of Aluminium Alloys

The physical, and mechanical properties of aluminium alloys depend on the microstructure and composition [6]. The properties of aluminium are significantly enhanced by the addition of selected elements to pure aluminium. Aluminium alloys are mainly classified as wrought alloys, casting alloys, and rapidly solidified/powder alloys, which are further subdivided into age-hardenable alloys and work-hardenable alloys. In the present work we are mainly focusing on wrought aluminium alloy series.

The main alloying elements used with aluminium are magnesium, manganese, copper, silicon, and zinc. Up to 2% of these elements can be present in aluminium alloys. Additionally, there are minor alloying elements added to aluminium alloys, but they are usually present in amounts $< 0.5\%$. The important impurities and alloying elements are listed here, along with a concise summary of their major effects, to produce a list of wrought composition families in Tables 2.2 and 2.3 [1], [6], [3].

Table 2.2. Types of wrought alloy series

Alloy group	Principal alloying element	Aluminium Content / Notes
1xxx	Unalloyed aluminium	Purity of 99.0% and above
2xxx	Copper and magnesium	Heat treatable alloys
3xxx	Manganese	Non heat treatable alloys

4xxx	Silicon	Low melting point alloys
5xxx	Magnesium	Non heat treatable alloys
6xxx	Magnesium and silicon	Heat treatable alloys
7xxx	Zinc, magnesium, copper	Heat treatable alloys
8xxx	Other elements	Variable

Table 2.3. Wrought alloy series – composition (wt%) and yield strength (MPa) ranges [3], [7]

Se- ries	Si (%)	Fe (%)	Cu (%)	Mn (%)	Mg (%)	Zn (%)	Cr (%)	Ti (%)	Zr (%)	YS (MPa)
1xxx	0–0.30	0–0.40	0.003– 0.35	0– 0.05	0–0.30	0–0.5	0– 0.005	0–0.10	—	23–92
2xxx	0.05– 1.3	0.05– 1.6	0.9–6.3	0–1.3	0–1.6	0–1	0–0.05	0–0.15	0–0.08	140– 462
3xxx	0.08– 0.7	0.05– 0.5	0.03– 0.9	0.15– 1.5	0–1.5	0–1.0	0–0.2	0– 0.175	—	61–170
4xxx	1.2–14	0.5–0.8	0.2–1	0.1– 1.2	0.1– 1.4	0.1– 1.3	0.01– 0.25	0.005– 0.2	0.01	105– 350
5xxx	0–0.7	0–0.4	0–0.6	0–1.4	0.4– 5.6	0–2.8	0–0.25	0–0.1	—	160– 310
6xxx	0.35– 1.35	0.075– 0.5	0.1– 0.95	0– 0.85	0.35– 1.4	0.1–1	0.03– 0.35	0.01– 0.2	0.05– 0.25	160– 430
7xxx	0.06– 0.5	0.04– 0.3	0.02– 2.4	0–0.8	0.75– 3.15	4– 11.5	0– 0.175	0–0.25	0–0.3	290– 730
8xxx	0.05– 1.2	0.12– 9.3	0.05– 2.2	0.01– 2.7	0.05– 1.0	0.02– 1.8	0.005– 1	0.005– 1	—	280– 560

For wrought alloy series a four-digit system and the main characteristics are given as follows [1]:

1xxx Series: Aluminium grades with a purity of 99% or higher are widely used, particularly in chemical and electrical applications. These alloys are typically characterised by high thermal and electrical conductivity, excellent corrosion resistance, and outstanding workability, but they exhibit relatively low mechanical strength. Strength can be moderately enhanced through strain hardening. The principal impurities in these grades are silicon and iron.

2xxx Series: In the 2xxx series alloys, copper is the principal alloying element, often added with magnesium. To obtain optimum properties, these alloys require solution heat treatment. Their mechanical properties are similar to, and sometimes even exceed, those of low-carbon steel in the heat-treated condition. In most instances, precipitation heat treatment (aging) is employed to further increase mechanical properties, which results in increases in hardness and yield strength, though with some loss in elongation. These alloys do not show good corrosion

resistance, unlike most other aluminium alloys, and may be subject to intergranular corrosion under certain conditions. A high-purity aluminium and a magnesium-silicon alloy from the 6xxx series, or those containing 1% Zn in the form of sheet, are usually clad on each side of 2xxx alloy rolled sheet. The coating thickness, usually 2 to 5%, provides galvanic protection to the core material, greatly enhancing corrosion resistance. Alloys in the 2xxx series are particularly well-suited for parts and structures requiring high strength-to-weight ratios. They are commonly used in the manufacture of aircraft wheels, truck and suspension parts, wing skins, aircraft fuselage, structural parts, and components that operate at temperatures up to 150°C.

3xxx Series: In the 3xxx series alloys, manganese is the major alloying element. They are non-heat-treatable alloys, but their strength is about 20% higher compared to the 1xxx series. Up to about 1.5% manganese can be effectively added to aluminium; however, only a limited percentage is used as a major element in most alloys.

4xxx Series: In the 4xxx series alloys, silicon is the major alloying element, and up to 12% silicon can be added to significantly lower the melting range without producing brittleness. These alloys are commonly used as welding wire and brazing alloys for joining, where a lower melting range than that of the base metal is required. Most alloys in this series are non-heat-treatable; however, when they are used in welding heat-treatable alloys, they may absorb some alloying elements. Consequently, they can exhibit a limited response to heat treatment. Alloys containing appreciable amounts of silicon develop a dark grey to charcoal appearance when anodised, making them suitable for architectural applications.

5xxx Series: In the 5xxx series alloys, magnesium is the major alloying element. These alloys are work-hardenable and provide moderate-to-high strength when magnesium is the primary alloying element or when used in combination with manganese. This series of alloys provides good corrosion resistance in marine environments and has excellent welding characteristics. However, to minimize susceptibility to stress-corrosion cracking, the amount of cold work and the operating temperatures of higher-magnesium alloys (those containing more than ~3.5% Mg) should be limited, especially for applications above ~65°C. Scandium (Sc) is especially effective in this series, enabling their dispersion hardening.

6xxx Series: The 6xxx series alloys are heat-treatable, formed by the combination of magnesium and silicon in 1.73:1 proportion (wt%) to create magnesium silicide (Mg_2Si), as shown in Figure 2.2 [8], [9], [10]. These alloys show good formability, weldability, machinability, and

corrosion resistance, but with medium strength. Heat treatment in the T6 temper, which involves precipitation hardening, strengthens these alloys, while in the T4 temper, the alloys are solution heat-treated but not precipitation hardened (see **Section 2.1.5** for a detailed explanation of temper designations) [10].

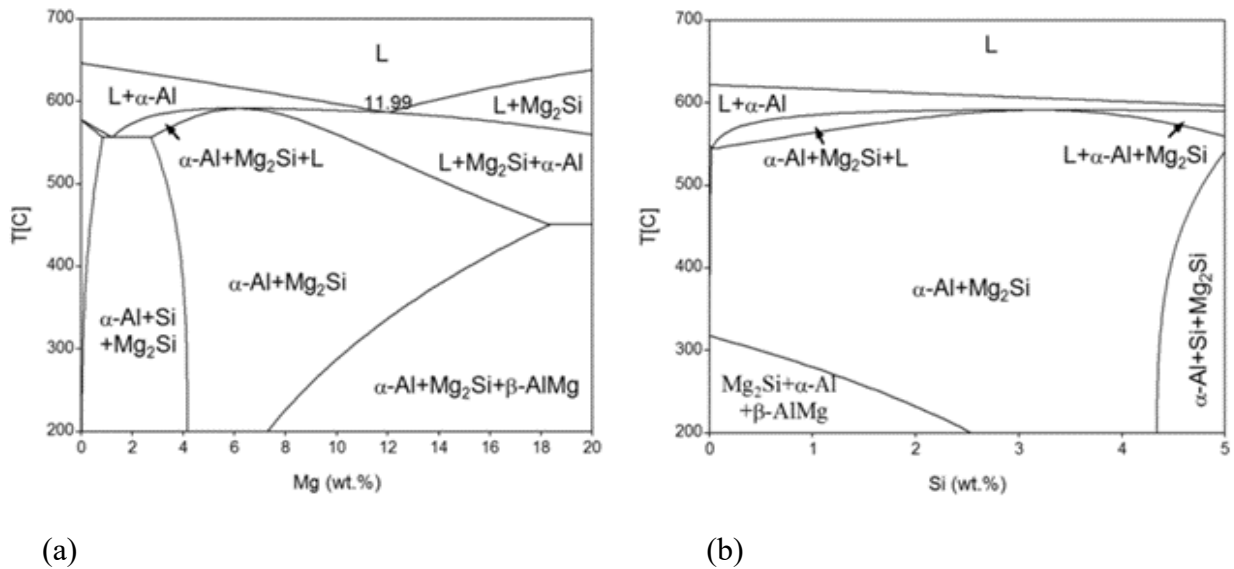


Figure 2.2. Al-Mg-Si phase diagram isopleths: (a) Al-xMg-2.4Si alloy, with Si fixed at 2.4 wt.% and Mg varying; (b) Al-7.5Mg-xSi alloy, with Mg fixed at 7.5 wt.% and Si varying [11]

7xxx Series: In Al-Zn-Mg 7xxx series alloys, zinc (1 to 8%) is the major alloying element. These alloys are heat-treatable and offer moderate to high strength. They exhibit good corrosion resistance, and weldability. Another group of 7xxx alloys, Al-Zn-Mg-Cu, provides very high strength but poor stress-corrosion resistance and are not weldable. Small quantities of other elements, such as chromium and zirconium, are added to improve properties, and dilute additions of scandium are also made. Alloys from the 7xxx series are used in mobile equipment, airframe structures, and other highly stressed parts. Higher-strength 7xxx alloys exhibit reduced resistance to stress-corrosion cracking, and to achieve a better combination of strength, fracture toughness, and corrosion resistance, they are often utilised in a slightly overaged tempering condition.

8xxx Series: The 8xxx series alloys encompass a wide range of chemical compositions and basically contain the alloys that do not fit in other series. For example, powder metallurgy-processed alloys, such as dispersion-strengthened Al-Fe-Ce alloys or Al-Fe-V-Si alloys,

demonstrate improved performance at elevated temperatures. Lithium-containing alloys offer higher stiffness and lower density.

2.1.4 Processing Conditions in Aluminium Alloys [1], [12]

Depending on the desired properties and intended applications, aluminium alloys can be processed in various ways. These processing conditions affect the mechanical performance and surface finish of the alloys. The most common processing methods are discussed below:

Extrusion: In this process, the aluminium alloy is forced through a die to produce uniform cross-sectional profiles. It often produces shapes such as beams, rods, and bars. It enhances the strength of the alloys through a technique known as strain hardening. It can also create complex shapes with precise dimensions.

Rolling: In this process, aluminium alloys are formed into plates, sheets, or foils. This is achieved by reducing the thickness and lengthening the alloy using a pair of rollers. Rolling enhances the strength, thickness uniformity, and surface finish of the alloy. The rolling process is divided into two types: hot rolling and cold rolling [12].

- **Hot Rolling:** This process is carried out at a high temperature, typically above 300°C, to increase the flowability of the alloy and make shaping easier.
- **Cold Rolling:** This process hardens the material by performing the operation at room temperature, strengthening the alloy through work hardening. It is mostly utilised for producing thinner, precision-shaped products such as sheets.

Forging: The process of forging involves the shaping of aluminium alloys by the application of compressive force. This is typically carried out at elevated temperatures and can lead to better mechanical properties, particularly strength, through the refined grain structure. Forging is done by either open-die or closed-die operations. It results in higher fatigue strength, tensile strength, and toughness due to the fine-grained microstructure developed in the process of deformation.

Aluminium plates and sheets, extrusions, forgings, and stampings are rolled and heat-treated in subsequent steps. Aluminium plates are thick and mainly used in structural forms, whereas sheets are thin and widely used in the production of stamping and forming processes. The

thickness as well as the heat treatment influence their strength, corrosion resistance, and general formability. There is much room for improvement in mechanical properties with respect to control over grain size and formation of phase under heat treatment.

2.1.5 Effects of Hardening and Tempering Treatments in Aluminium Alloys [1], [10]

Aluminium alloys are available in a wide range of temper designations, which indicate the combination of thermal and mechanical processing used to achieve specific properties. Common designations include F (as-fabricated), O (annealed), T1–T10 (various heat treatments for age-hardenable alloys), and H1x–H4x (strain-hardened conditions for non-age-hardenable alloys). Although many tempers exist, in this study we focus only on those relevant to the alloys and processing conditions investigated: T5, T6, and T7 for age-hardenable 2xxx, 6xxx, and 7xxx series alloys, and H14 for 1xxx series and H32 for 3xxx and 5xxx series non-age-hardenable alloys as discussed below in detail.

T Tempers for Age-Hardenable Alloys

Age-hardenable alloys are strengthened through heat treatment, which promotes the formation of precipitates that increase their hardness and strength. In this study, we focus only on the 2xxx, 6xxx, and 7xxx series alloys, presenting the tempering conditions for these age-hardenable systems.

- **T4:** The alloy is solution heat-treated and then naturally aged at room temperature. After quenching from the solution treatment temperature, the material is allowed to age without artificial heating. The T4 temper provides moderate strength with good ductility and formability and is often used when subsequent forming operations are required.
- **T5:** After a high temperature shaping process (usually extrusion), the alloy is cooled and then artificially aged without intermediate cold work. The material is held at a sufficiently high temperature (e.g., 8 hours at 175°C or 24 hours at 120°C) to achieve precipitation hardening. Minor straightening or flattening may be applied to meet dimensional tolerances, but this does not significantly affect mechanical properties. This temper is used when moderate strength and good dimensional stability are required.
- **T6:** Solution heat treatment is applied, followed by artificial aging to achieve maximum strength without remarkable cold working. This is a common temper for 2xxx, 6xxx, and 7xxx series alloys, widely used in various industrial applications.

- **T7:** Following solution heat treatment, the alloy is aged to an overaged condition. Peak strength is slightly reduced compared to T6, but the temper improves dimensional stability and resistance to stress corrosion, making it suitable for applications where long-term stability under load is critical.

H Tempers for Non-Age-Hardenable Alloys [10]

Non-age-hardenable aluminium alloys cannot be strengthened through precipitation heat treatment and therefore rely on mechanical deformation (cold working) to increase strength. In this study, strain-hardened conditions for 1xxx, 3xxx, and 5xxx series alloys are considered.

The H temper designation follows a systematic nomenclature that describes both the type of strain-hardening process applied and the degree of cold work introduced. The letter H indicates strain hardening. The first digit following “H” specifies the processing condition:

- **H1x** – Strain hardened only. The alloy is strengthened exclusively through cold working without subsequent thermal treatment.
- **H2x** – Strain hardened and partially annealed. After cold work, the alloy is partially softened through controlled annealing to achieve a specified strength level.
- **H3x** – Strain hardened and stabilised. The alloy undergoes a low-temperature stabilisation treatment after cold working to reduce strain ageing effects and improve dimensional stability. This designation is common for magnesium-containing alloys, such as the 5xxx series.
- **H4x** – Strain hardened and lacquered or painted. The alloy is cold worked and subsequently coated, where the thermal effect of coating slightly modifies mechanical properties.

The second digit indicates the degree of strain hardening, with increasing numbers corresponding to higher levels of cold deformation and therefore increased strength and reduced ductility. Common designations include: H12–Quarter-hard, H14–Half-hard, H16–Three-quarter hard, H18–Full hard.

This structured nomenclature allows clear identification of both the mechanical processing history and the expected mechanical behaviour of non-age-hardenable aluminium alloys. Since strain-hardened tempers are included in the present dataset, understanding the H1x–H4x

system is essential for correctly interpreting variations in strength and ductility among these materials.

The specific strain-hardened tempers considered in this study are:

- **H14:** The material is cold worked (strain hardened) solely to increase strength, with no further heat treatment. This results in a good balance of strength and ductility and is typical for 1xxx series alloys.
- **H32:** The alloy is strain-hardened and then stabilised through a low-temperature treatment, improving corrosion resistance without significantly reducing strength. This temper is typical for 3xxx and 5xxx series alloys, which are widely used in marine, roofing, and chemical processing applications [1], [10].

2.1.6 Recyclability and Related Issues in Aluminium Alloys

Traditionally, primary aluminium production, derived from bauxite ore through the Bayer process and subsequent Hall–Héroult electrolysis, is highly energy-intensive and generates significant CO₂ emissions. The overall efficiency of this process is only about 25%, making aluminium production costly and environmentally demanding. Expanding the use of secondary (recycled) aluminium is therefore one of the most effective strategies for reducing the life-cycle impact of aluminium and supporting circular economy and net-zero targets [13], [14], [15].

Recycling requires only about 5% of the energy needed to produce primary aluminium, saving roughly 8 tonnes of bauxite and 14,000 kWh of energy per tonne of metal. The carbon footprint is also greatly reduced, from about 12 tonnes of CO₂ per tonne of primary aluminium to only 0.35 tonnes for recycled aluminium [16]. Currently, recycled aluminium accounts for roughly one-third of global use, and this share is expected to reach 50% by 2050 [17], [18].

In Europe, nearly 90 percent of automotive aluminium scrap is recovered and reused annually, and recycling rates for beverage cans exceed 75% [19], [18],[20]. Scrap availability depends strongly on product life cycles: beverage cans have an average life of ~0.2 years, vehicles about 15 years, and building components up to 50 years [21]. Aluminium scrap is generally classified into new scrap (clean process waste such as trimmings or defective castings) and old scrap (post-consumer materials like cans, automotive parts, and building elements). New scrap usually has a controlled composition and can be easily remelted, while old scrap is often

contaminated with coatings, lubricants, or other foreign materials and requires additional cleaning, de-coating, and sorting before remelting [21].

The main challenge in producing wrought aluminium alloys from recycled material lies in achieving the correct chemical composition with minimal use of primary aluminium. Mixed scrap batches often contain excessive amounts of critical elements such as Fe, Cu, Mn, Mg, Zn, and Si, which must be diluted or carefully sorted prior to melting [22], [23]. Clean and well-sorted scrap helps maintain acceptable impurity levels but is costlier and less available.

Recycling efficiency is also limited by the inability of wrought alloys to tolerate unintended elements. To control impurities, industry often relies on dilution with primary metal, which increases cost and produces surplus low-grade scrap [22], [23]. These issues highlight the importance of developing alloys with higher impurity tolerance and improved sorting technologies.

Conventional separation methods, such as density-based, eddy current, magnetic, or air classification, struggle to distinguish similar alloys and are prone to losses due to corrosion or contamination [24] [25], [26]. Advanced sensor-based technologies, including X-ray fluorescence (XRF), X-ray transmission (XRT), and laser-induced breakdown spectroscopy (LIBS), enable more accurate alloy identification but remain costly. Meanwhile, secondary aluminium production produces salt slag, a hazardous byproduct requiring special treatment or recycling [27], [28], [29] [30], [31].

Overall, aluminium recycling provides significant energy, economic, and environmental benefits, lowering production costs and emissions while conserving natural resources. However, the persistence of impurity control, alloy compatibility, and byproduct management (such as slags) challenges underscores the need to simplify alloy compositions.

2.1.7 Need for Optimisation and Reducing the Number of Alloy Grades

Aluminium (Al) alloy design over the last century has largely been an iterative and empirical process, based on lessons learnt from human experience and in-service use. In the field of Al-alloy metallurgy, incremental alloy development is slow and costly and does not fully leverage the data that exists [1]. In many engineering applications, metal alloys are essential, and for our ‘metals-based civilisation’, they play a crucial role. Most metals do not exist in their natural

(metallic) state in the Earth's crust, with a few exceptions (e.g., silver, gold, and mercury). Therefore, almost all metallic alloys are anthropogenic, they are extracted and designed by humans. This implies that the ultimate "design" of alloys is a human endeavour, not constrained by inherent material performance limitations [32].

Phase diagrams have traditionally been used to discover new aluminium alloys, relying on thermodynamic information [32]. However, it is very difficult to reveal and interpret the relationship between changes in input variables and the target property without using computational and numerical methods.

More recently, 'ab initio' methods have been employed to discover aluminium alloys by performing structural calculations from "scratch." However, for alloy design, this approach cannot be generalised due to the non-linearity and high dimensionality of property variations with composition. Additionally, existing knowledge about alloys is not fully utilised by ab initio methods [33].

When the solution space has already been narrowed and combinatorial methods become viable, improved fabrication rates can be highly beneficial. However, in cases where the potential solution space is vast, even more rapid testing and fabrication will not be suitable. The probability of fabricated alloys having the desired properties will be low, necessitating a narrowing of the solution space.

2.2 Machine Learning (ML) and Artificial Intelligence (AI)

2.2.1 Overview of ML/AI

In 1956, John McCarthy coined the term "Artificial Intelligence (AI)" at the Dartmouth Conference, which was built upon the work of Alan Turing on machine intelligence in the 1950s [34]. During the early development phase, Newell and Simon in 1956 demonstrated that machines can perform problem-solving like humans can [35]. Although during the AI winters, the progress was quite slow during the 1970s and late 1980s due to the low availability of computational power and high expectations from AI. Interest in AI was renewed with advances in machine learning (ML), a subset of AI defined by Arthur Samuel (1959) [36] is widely recognised for defining ML. He described it as "the field of study that gives computers the ability to learn without being explicitly programmed. Key contributions include Hinton et al. (2006) [37],

who introduced deep belief networks, and LeCun et al. (2015) [38], who reviewed the transformative impact of deep learning across fields such as computer vision, speech recognition, and natural language processing (NLP), marking a new era of intelligent systems and applications. Many other researchers have further contributed to AI and ML, shaping their rapid development over the decades. Let us now explore the definitions of these methods [39], [40], [41].

Artificial Intelligence (AI) [42]

Artificial intelligence refers to a computer system's ability to imitate certain cognitive processes performed by humans, such as learning and problem-solving. A computer system can replicate human reasoning to learn from new knowledge and make judgements based on that information.

Machine Learning (ML) [42]

Machine learning is a subset of AI. It involves using mathematical data models to assist a computer in self-directed learning. As a result, a computer system continuously learns new skills and patterns within a dataset and improves on its own.

AI enables “intelligent” computers to mimic human thought and perform autonomous tasks. The process through which a computer system acquires intelligence is known as machine learning.

Capabilities of AI and ML [40], [43], [44]:

- **Predictive Analytics:** This capability enables the prediction of trends and behavioural patterns by identifying cause-and-effect relationships within data.
- **Recommendation Engines:** Recommendation engines analyse data to generate personalised product suggestions.
- **Speech Recognition and Natural Language Processing (NLP):** A computer system can recognise spoken words and interpret meaning from both written and spoken language using NLP.
- **Sentiment Analysis:** Sentiment analysis helps identify and classify positive, neutral, and negative sentiments expressed in text.

- **Better and Faster Decision Making:** Businesses use AI to reduce human error and machine learning to improve data integrity, resulting in better decisions based on data.
- **Image and Video Processing:** This enables features such as visual search and recognition of faces, objects, and actions in images and videos.

Python Libraries Used in Machine Learning [41], [45]

Python is one of the most widely used programming languages, due to its extensive library package, and offers many advantages over other programming languages. The following Python libraries are commonly used in machine learning:

- **NumPy:** A well-known Python library for processing large multidimensional arrays and matrices with sophisticated mathematical operations. It is highly useful for basic scientific ML computations.
- **Pandas:** It is the most widely used Python packages for data analysis and preparation. While not used directly for machine learning, these packages are essential during data preparation prior to model training.
- **Matplotlib:** A Python data visualisation library, Matplotlib offers a variety of graphing possibilities, such as histograms, error plots, and bar plots, and more.
- **Scikit-learn:** Scikit-learn is a NumPy and SciPy-based library. It supports the majority of supervised and unsupervised learning algorithms.
- **TensorFlow:** TensorFlow is an open-source high-performance numerical computation library created by the Google Brain team. It enables the definition and execution of tensor-based computation and can be used for deep neural network training and deployment.

In this project, we used the following libraries, as shown in Table 2.4, and common ones are NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn.

Table 2.4. Libraries used in project

Library	Purpose/Usage
NumPy	Used for performing mathematical operations efficiently
Pandas	Used for data manipulation and analysis
Matplotlib	Used for creating interactive visualisations of data
Seaborn	Used for making statistical plots and visualisations
Scikit-learn	Used for ML tasks such as regression, clustering, and data pre-processing to build predictive models

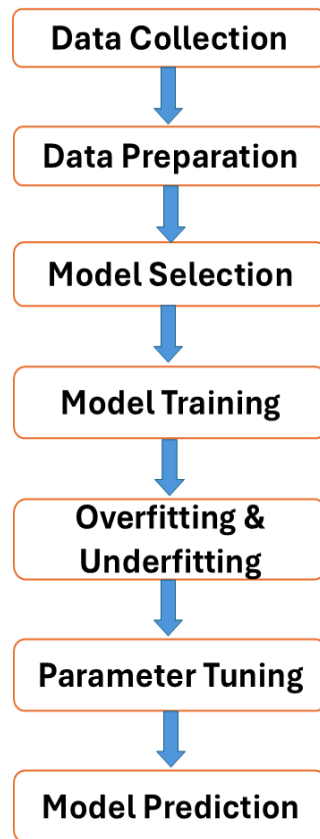


Figure 2.3. The main steps in ML

Step 1: Collecting Data – As shown in Figure 2.3, machine learning begins with the data provided to it. It is crucial to obtain trustworthy data so that the ML model can identify the appropriate patterns. The accuracy of a model depends largely on the quality of the data it receives.

Step 2: Preparing the Data – The data must be cleaned of irrelevant information, duplicate values in rows and columns, missing values, and other issues. The dataset may also need to be reorganised, with rows and columns potentially being rearranged or indexed.

Step 3: Model Selection – The outcome of applying a machine learning algorithm to the collected data depends on the choice of the ML model. It is important to choose a model suited to the specific task at hand.

Step 4: Model Training – The dataset is usually divided into an 80:20 split, with 80% used for training and 20% used for testing. During the training phase, the selected machine learning

algorithm learns patterns from the training data by adjusting its internal parameters (such as weights and biases in neural networks). This is typically achieved by minimising a loss (or error) function, which measures the difference between the predicted output and the actual output.

An optimisation algorithm (e.g., gradient descent) is used to iteratively update the model parameters to reduce this error. This process continues for multiple iterations (epochs) until the model converges or reaches satisfactory performance [47].

After training, the model's performance must be evaluated. This is done by testing the model with data it has never seen before. The testing set, obtained during the data separation process, serves as the unseen data. If testing is performed using the same data as training, the accuracy would be very high, since the model has already identified those patterns. The model improves its predictive accuracy over time with more training.

Step 5: Underfitting and Overfitting – Underfitting usually occurs if a model fails to accurately capture the complexity of the data or if there are insufficient features to represent it. As a result, the model produces significant training errors. Conversely, an overfit model treats some noise in the training data as useful information, making it unable to make reliable predictions on new, unseen data.

Step 6: Parameter Tuning – The model parameters are variables typically chosen by the programmer. The accuracy of the model will be highest at specific parameter values. Finding these optimal settings is referred to as parameter tuning.

Step 7: Making Predictions – After completing all the previous steps, the model is ready to be used on unseen data to make accurate predictions [40].

2.2.2 Types of Machine Learning [40], [46]

Machine learning (ML) can be broadly divided into four main types; however, this study focuses on supervised and unsupervised learning, as outlined below:

Supervised Learning: Supervised machine learning is characterised by the use of labelled datasets. These datasets are intended to "supervise" or "train" algorithms to correctly classify data or predict outcomes. Labelled inputs and outputs allow the model to monitor its accuracy and

improve over time. Supervised learning is divided into two categories: classification and regression.

Unsupervised Learning: In unsupervised learning, machine learning algorithms are used to analyse and group unlabelled datasets. These algorithms are referred to as ‘unsupervised’ because they identify hidden patterns in the data without human intervention. The three main tasks that unsupervised learning algorithms are used for are clustering, association rule learning, and dimensionality reduction.

2.2.3 Machine Learning Algorithms

The most common ML methods used in our analysis are briefly introduced and discussed in this section. We begin with exploratory data analysis (EDA) and feature engineering. Next, we cover decision tree-based techniques, such as random forests (RFs) for regression and classification. We then discuss clustering algorithms, which enable the identification of patterns and groupings in data without prior labels. Additionally, we include optimisation algorithms, such as Bayesian optimization, artificial neural networks (ANNs), and multi-criteria decision-making (MCDM) algorithms, which are particularly useful for balancing competing objectives in materials design.

2.2.3.1 Exploratory Data Analysis (EDA) Techniques [48]

Exploratory data analysis (EDA) is used to understand the dataset’s structure prior to machine learning model training. The following are some EDA techniques utilised in present study:

Histogram: Histograms are plots for continuous variables, generated using Seaborn’s `histplot()`, to examine the overall shape of the dataset. This aids in decisions regarding potential transformations and outlier handling.

Pair Plot: The `pairplot()` function generates scatter plots and histograms for each pair of variables, illustrating relationships between features to examine correlations, and multivariate patterns.

Scatter Plot: The `scatterplot()` function generates scatter plots to examine interactions among features, identify potential trends, and correlations, thereby informing feature selection.

Violin Plot: The `violinplot()` function combines information about distribution and density to compare features. It serves as a visual tool to highlight variations within the data.

These visual techniques provide critical insights into distributions, associations, and variations, thereby facilitating decisions related to feature selection and pre-processing.

2.2.3.2 Feature Engineering [49]

Feature engineering is a machine learning pre-processing step that extracts features from unprocessed data. It helps in better communicating a fundamental issue to predictive models, improving the model's accuracy on unobserved data. Four main operations make up feature engineering in machine learning: feature creation, feature transformations, feature extraction, and feature selection. These methods are typically used because adding more features increases the dimensionality of the data, which unnecessarily raises the model's complexity. After

removing irrelevant features, we often end up with a simpler feature space, making the outcomes easier to interpret.

2.2.3.2.1 Handling Outliers [50]

Outlier handling involves identifying and removing outliers from a dataset to achieve a more accurate representation of the data. This process can be applied at different scales and significantly affects the predictions of the model. One commonly used method for detecting outliers is box plot analysis, illustrated in Figure 2.4. [50]:

A box plot is a graphical representation of a data distribution that highlights its central tendency, spread, and outliers. The steps involved in box plot analysis are as follows:

Step 1: Calculate the Interquartile Range (IQR) – The box plot quartiles are Q1 (lower quartile, 25th percentile), median (50th percentile), and Q3 (upper quartile, 75th percentile). The IQR is the difference between Q3 and Q1, representing the central 50% of the data.

Step 2: Identify Outliers – Outliers are values that lie more than 1.5 times the IQR below Q1 or above Q3.

Step 3: Interpret the Plot –

- **Box:** The central rectangle spans Q1 to Q3, showing the middle 50% of the data.
- **Median:** A line inside the box represents the 50th percentile of the dataset.
- **Whiskers:** Lines extending from the box to the highest and lowest values within $1.5 \times$ IQR from the quartiles as shown in Figure 2.4.
Lower whisker: Extends to the smallest non-outlier value.
Upper whisker: Extends to the largest non-outlier value.
- **Outliers:** Points beyond the whiskers, usually plotted as circles or dots.
- **Symmetry:** Suggested by roughly equal whisker lengths and a median line near the centre of the box.
- **Skewness:** Indicated by uneven whiskers, where one whisker is longer than the other.
- **Spread:** A larger IQR represents greater variability in the data.

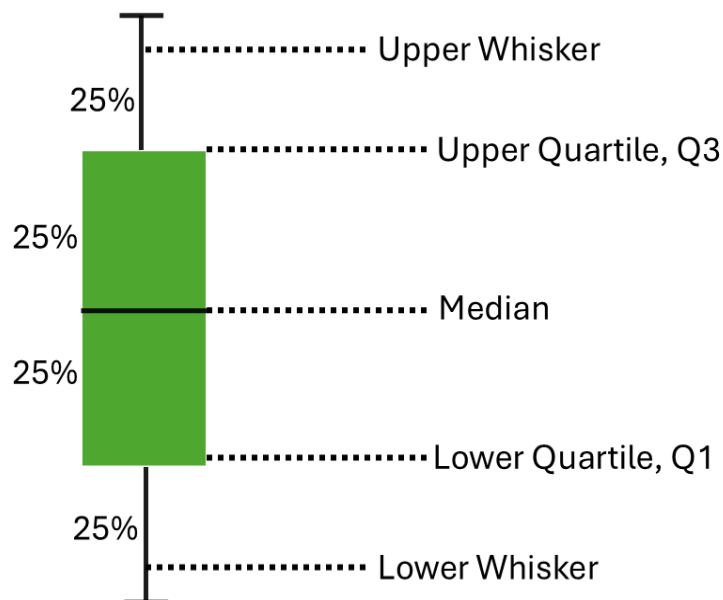


Figure 2.4. Box plot

2.2.3.2.2 Scaling Techniques [51], [52]

Feature scaling is a pervasive and crucial issue in machine learning. Data must be scaled appropriately to train a predictive model. The following are common feature scaling techniques:

Normalisation: Normalisation scales all values within a predetermined range between 0 and 1 (also known as min-max normalisation). This processing reduces the effect of outliers but does not alter the feature's distribution.

Standardisation: Standardisation scales the data by considering the standard deviation (also known as z-score normalisation). This method diminishes the impact of outliers by producing a distribution with a mean of 0 and a variance of 1. All data points are subtracted from their mean, and the result is divided by the variance of the distribution.

2.2.3.3 Feature Selection Techniques: Correlation Analysis [53]

Correlation analysis measures the direction and strength of the relationship between two or more variables. The correlation coefficient ranges from +1 to -1, where a value of ± 1 indicates a perfect correlation and a value closer to 0 suggests a weaker relationship. The sign of the coefficient indicates the relationship's direction: a + sign denotes a positive relationship, while a - sign indicates a negative one. The four common types of correlation are Pearson correlation, Kendall rank correlation, Spearman correlation, and Point-biserial correlation. Here, we focus on Pearson correlation.

Pearson's correlation coefficient r is the most widely used statistic for evaluating the strength of a linear relationship between variables, with the strength determined by the magnitude (absolute value) of r . The absolute value of r indicates the degree of linear association, with values closer to 1 representing stronger relationships. For instance, in an alloy dataset, Pearson's r can be used to measure the relationship between compositions and properties. A Pearson correlation coefficient r close to ± 1 indicates a strong linear relationship between the two variables. The correlation is carried out using the Pearson correlation formula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

Where n is the number of pairs of variables, $\sum xy$ is the sum of the product pairs, $\sum x$ is the sum of the x scores, $\sum y$ is the sum of the y scores, $\sum x^2$ is the sum of the squared x scores, and $\sum y^2$ is the sum of the squared y scores.

In simple linear regression with a single predictor, the Pearson correlation coefficient r between the predictor and response variable is numerically equivalent to the regression correlation coefficient R , and therefore R^2 equals r^2 . However, Pearson's r^2 offers a distinct interpretative advantage. While R indicates the strength and direction of the linear relationship, r^2 quantifies the proportion of variance in the response variable that is explained by the predictor. This makes r^2 more directly interpretable, as it expresses explanatory power in terms of

variance explained rather than merely the magnitude of association. Furthermore, r^2 can be computed directly between two variables without the need to construct a regression model. This property makes it particularly advantageous in preliminary feature evaluation and selection, where the objective is to assess the explanatory contribution of individual variables prior to model development.

2.2.3.4 Regression Algorithm [54]

Regression creates relationship between different variables. It typically considers a phenomenon of interest and has numerous observations for it. Each observation includes at least two features. Regression creates a relationship between the features under the premise that at least one of them is dependent on the others. The dependent variables or responses are referred to as the dependent features. The independent variables or predictors are names for the independent features. There is typically only one continuous, unbounded dependent variable in regression analysis. However, the inputs can be categorical, discrete, or continuous data. It is customary to use $\mathbf{y}_{rf}$ to represent the outputs and $\mathbf{x}_{rf}$ to represent the inputs. The vector $\mathbf{x}_{rf} = (x_1, \dots, x_i)$, where i is the number of inputs, can represent two or more independent variables. Although there are many regression techniques, in this study we focus on the random forest regression approach [54].

Random Forest Regression [55], [56]

Random Forest Regression is an ensemble learning technique that predicts a continuous outcome by using multiple decision trees (DTs). A decision tree is a tool used for decision support, represented as a tree-like graph that illustrates potential outcomes. This method employs conditional expressions within its algorithm. They assist in identifying a plan of action, particularly for decision analysis, to help achieve objectives more effectively.

In general, it is more efficient in cases of complex regression when the relationship between independent and dependent variables is too nonlinear or too intricate. It relies on exploiting the strength of multiple decisions by averaging the predictions to have a stronger and more accurate result.

If we consider a dataset with independent variables $x_1, x_2, \dots, x_i$, the Random Forest Regression model trains T decision trees using different random subsets of data and features. For a new

input, each tree in the forest predicts an outcome $y_t(x_1 \dots x_i)$, and the final regression output is given by:

$$\hat{y}(x_1, \dots x_i) = \frac{1}{T} \sum_{t=1}^T y_t(x_1, \dots x_i) \quad (2)$$

In this case, the averaging reduces the variance caused by each individual tree so that the resulting prediction surface is smoother and more general.

The goal of Random Forest Regression is to minimise the prediction error by training multiple decision trees on randomly bootstrapped samples and averaging their predictions. The randomness in sampling and feature selection ensures that the model reduces overfitting while maintaining a high level of accuracy.

The performance of Random Forest Regression can be approximated with the coefficient of determination, also called R^2 or Mean Squared Error (MSE). The R^2 value is an expression of how much of the variations in the target variable y can be explained by the model.

$$R^2 = 1 - \frac{SSR}{SST} \quad (3)$$

where SSR is the sum of squared residuals (the squared differences between predicted and observed values), and SST is the total sum of squares (the total variation in y).

A higher R^2 value (closer to 1) indicates that the model explains more of the variation. An $R^2 = 1$ corresponds to a perfect fit.

2.2.3.5 Classification Algorithms [57],[56]

Classification algorithms are a fundamental component of supervised machine learning. Their purpose is to predict categorical outcomes from input data by learning patterns and relationships within labelled datasets. A classification task involves assigning each data point to one of several predefined categories or classes based on its features. These algorithms analyse the input features of the training data to learn decision boundaries or patterns that distinguish the classes and then apply this learned knowledge to make predictions on unseen data. Classification algorithms are of various types; in this study, we focus on random forest classification.

Random Forest Classification

The Random Forest Classification model is an ensemble learning method widely used for various classification tasks. It generates multiple decision trees, with each tree producing its own prediction. The final class prediction is determined by a majority vote among all the trees. The mechanism of random forest classification is as follows:

A number of decision trees are created using random selections of subsets of features and observations. For a given input $\mathbf{x}_{rf} = (x_1, x_2, \dots, x_i)$, where each x_i represents a feature and i is the number of inputs, the process proceeds as follows:

Step 1: Training of Trees – For each individual tree, a bootstrap sample D_b is drawn from the training dataset D , and the tree is trained using this sample. The bootstrap sample is performed by randomly choosing data points with replacement; that is, for a single sample, some of the data points may appear several times while some are not included. At each node of the decision tree, only a random subset of features is considered for splitting. This adds diversity among trees, reduces correlation between trees, and avoids overfitting. Further splitting of the tree proceeds until a stopping criterion, such as maximum depth or some threshold value for the number of samples in a leaf, is met.

Step 2: Prediction for a Single Tree – Once the tree is constructed, it maps the input $\mathbf{x}$ to one of the classes $C_1, C_2, \dots, C_q$, where q is the number of classes. This prediction is made by traversing the tree based on the feature values in $\mathbf{x}_{rf}$.

Mathematically, for the t -th tree, the prediction is:

$$h_t(\mathbf{x}_{rf}) = \arg \max_{C_j} (P(C_j | \mathbf{x}_{rf}, T_t)) \quad (4)$$

where T_t is the model of the t -th tree, and $P(C_j | \mathbf{x}_{rf}, T_t)$ is the probability that $\mathbf{x}_{rf}$ belongs to class C_j as determined by the tree.

Step 3: Final Prediction – The Random Forest averages the predictions from all T trees with majority voting. The final predicted class $\hat{y}$ is:

$$\hat{y} = \arg \max_{C_j} \left(\sum_{t=1}^T I(h_t(\mathbf{x}_{rf}) = C_j) \right) \quad (5)$$

where $I(.)$ is an indicator function, which is 1 when the condition is satisfied, and 0 otherwise.

2.2.3.6 Dimensionality Reduction Algorithm: Principal Component Analysis (PCA) [58]

PCA is a statistical method for reducing the number of dimensions in a dataset while maintaining as much information or variance as possible. It converts a large set of correlated variables into a smaller set of uncorrelated variables called principal components. PCA is commonly used in machine learning to reduce dataset complexity while preserving essential information. The steps involved in PCA are as follows:

Step 1: Standardising the Data – PCA is sensitive to the scale of the data, so when features have different units or ranges, it is essential to standardise them to ensure each contributes equally. This is done by subtracting the mean and dividing by the standard deviation for each feature.

$$X' = \frac{X - \mu}{\sigma} \quad (6)$$

Where: X' is the standardised value, X is the original value, μ is the mean of the feature, and σ is the standard deviation of the feature.

Step 2: Calculate the Covariance Matrix – The covariance matrix is of importance in understanding one variable in relation to another. This matrix shows the magnitude of joint variation between two variables. A very high covariance between two variables indicates extremely similar paths of variation.

$$\text{cov}(\mathbf{x}_{pca}, \mathbf{y}_{pca}) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) \quad (7)$$

Where: $\text{cov}(\mathbf{x}_{pca}, \mathbf{y}_{pca})$ is the covariance between two variables $\mathbf{x}_{pca}$ and $\mathbf{y}_{pca}$, X_i and Y_i are the values of the variables $\mathbf{X}$ and $\mathbf{Y}$, and μ_X and μ_Y are the means of $\mathbf{X}$ and $\mathbf{Y}$.

Step 3: Compute the Eigenvalues and Eigenvectors – The covariance matrix is then used to compute eigenvalues and eigenvectors. These eigenvectors represent the directions of the maximum variance (principal components), and the eigenvalues tell us the magnitude of the variance along these directions.

$$cov(X) \cdot v = \lambda \cdot v \quad (8)$$

Where: $cov(X)$ is the covariance matrix, v is the eigenvector, and λ is the eigenvalue.

Step 4: Arrange the Eigenvectors and Eigenvalues – After calculating the eigenvalues, the eigenvectors are arranged in descending order based on their corresponding eigenvalues.

The eigenvectors linked to the largest eigenvalues represent the principal components these are the directions in feature space along which the data variance is maximised.

Step 5: Select the Top w Principal Components – We choose the top w eigenvectors (principal components) from the sorted eigenvalues. The value of w determines the number of dimensions to retain. For example, if you want to preserve 80% of the total variance, you will select the top w components that together account for that proportion of the variance.

Step 6: Transform the Data – Finally, we can project the original data onto the new set of principal components (eigenvectors), reducing the dimensions of the data.

$$X_{new} = X' \cdot W \quad (9)$$

Where: X_{new} is the transformed data (reduced dimensionality), X' is the standardised data, and W is the matrix of the top w eigenvectors (principal components).

2.2.3.7 Clustering Algorithm

Clustering is the process of grouping variables with similar characteristics, which helps simplify complex datasets and makes them easier to interpret. The primary goal of clustering is to uncover the hidden structure within the data. Since clustering can be subjective, various techniques have been developed to address different challenges. Each problem has its own criteria for measuring the similarity between data points, so it is essential to select a method that aligns with the specific clustering objective. In this study we only focused on some of the most common clustering techniques as discussed below.

2.2.3.7.1 K-means Clustering Algorithm [59], [60]

K-means clustering is a widely used clustering technique that partitions data into distinct groups based on similarity. As a non-hierarchical method, it assigns data points to clusters so that points with similar features are grouped together, while dissimilar points are placed in separate clusters. The algorithm performs this classification by calculating the distance between each data point and the centroid of every cluster, assigning each point to the nearest cluster. The process consists of the following steps:

Step 1: Initialise Cluster Centres – Randomly select K distinct data points as the initial cluster centres. Then, assign each of the remaining $n-K$ data points to the cluster with the nearest centre.

Step 2: Update Cluster Centres – Using the current cluster assignments, calculate the new cluster centres.

Step 3: Reassign Patterns – Assign each of the n data points to the closest cluster centre.

Step 4: Convergence Check – If the assignment of patterns to clusters does not change after two iterations, stop; otherwise, return to **Step 2**.

The Number of Clusters can be Determined Using the Elbow Method:

The Elbow method is one of the most widely used techniques for determining the optimal number of clusters. This method uses the “total variation within a cluster”, measured by the Within-Cluster Sum of Squares (WCSS). The WCSS value is calculated using the following formula:

$$\text{WCSS} = \sum_{o=1}^K \sum_{i \in C_o} \| \mathbf{Z}_i - \mathbf{U}_o \|^2 \quad (10)$$

Where: K = number of clusters, C_o = set of points in cluster o , $\mathbf{Z}_i$ = feature vector of point i , $\mathbf{U}_o$ = centroid of cluster o

Silhouette Method: The Silhouette Method is used to evaluate the quality of clustering algorithms. For each data point, the average intra-cluster distance (a_i) and the average distance to the nearest cluster (b_i) are used to calculate the Silhouette Score (SC):

$$SC = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (11)$$

Where a_i is the average distance between a data point and all other points within the same cluster, and b_i is the average distance between that data point and all points in the nearest cluster to which it does not belong.

The Silhouette Score (SC) ranges from -1 to 1: A value of 1 indicates well-separated, distinct clusters, and a value of -1 suggests incorrect cluster assignments. To determine the optimal number of clusters, the average SC across all samples can be computed and used as a metric. Higher scores indicate better-defined clusters.

2.2.3.7.2 Hierarchical Cluster Analysis (HCA) [60]

Hierarchical Cluster Analysis (HCA) is also an unsupervised ML algorithm which is used to organise unlabelled datasets into clusters. This method builds a hierarchy of clusters, often visualised as a dendrogram, a tree-like structure, as shown in Figure 2.4. Steps involved in hierarchical clustering are as follows:

Step 1: Initial Clusters – Treat each data point as an individual cluster. If there are N data points, there will initially be N clusters.

Step 2: Combine Nearest Clusters – Identify the two closest data points or clusters and combine them into a single cluster. This results in $N-1$ clusters.

Step 3: Repeat Merging – Continue merging the two nearest clusters to form a single cluster. After each step, the number of clusters decreases by one.

Step 4: Stop Condition – Repeat the merging process until only one cluster remains.

Step 5: Construct the Dendrogram – Once all the clusters have been combined into one large cluster through hierarchical agglomerative clustering, the next step is to construct a dendrogram. The dendrogram visually represents the hierarchical relationship of the clusters and

shows how individual data points are progressively grouped together. It is useful for understanding the underlying structure of the data and determining the optimal number of clusters. In present case, the dendrogram was generated using the `linkage()` function from the SciPy library with the 'ward' method, which minimises the variance within clusters. A horizontal line was added to the dendrogram at a specified height ($y = 6$ on y-axis) to visually indicate the separation of the data into 3 clusters. The resulting figure (Figure 2.5) helps to observe the clusters and make decisions about how to partition the data based on the problem's requirements.

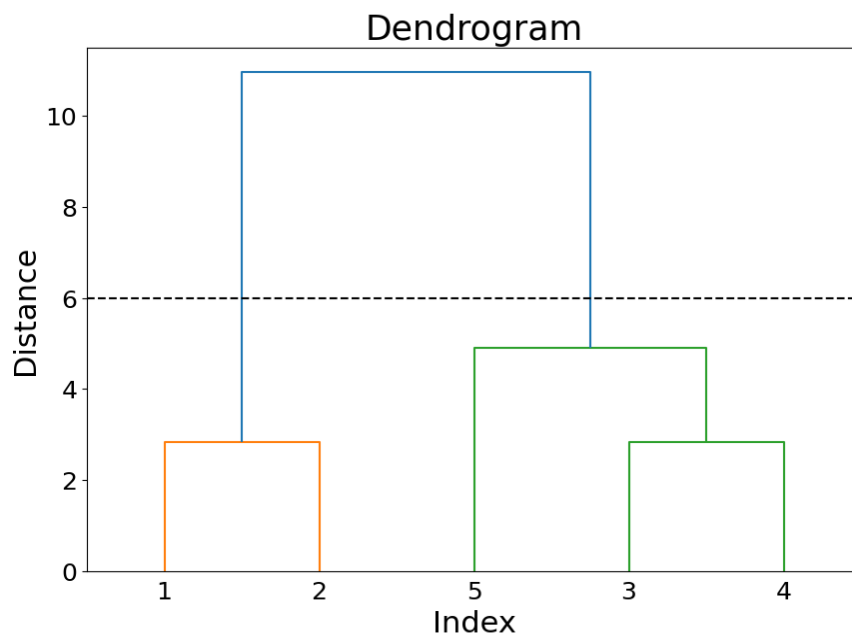


Figure 2.5. Dendrogram illustrating the agglomerative hierarchical clustering process for five observations. The y-axis represents the linkage distance at which clusters are combined. Observations 1 and 2 form the first cluster, observations 3 and 4 merge next, and observation 5 subsequently joins this cluster. Applying a cut-off distance of approximately 6 (dashed line) yields two final clusters: $\{1,2\}$ and $\{3,4,5\}$

The hierarchical clustering depends on how close two groups are to one another. There are several methods for measuring the distance between clusters, and these methods determine the clustering rules. These approaches are referred to as linkage techniques. Below are some commonly used linkage techniques:

Single Linkage: Measures the shortest distance between the nearest points of two clusters.

Complete Linkage: Measures the greatest separation between the farthest points in two clusters. This method tends to create more compact and evenly shaped clusters compared to single linkage and is widely used.

Average Linkage: Calculates the average distance between all pairs of points in two clusters. The total distance is summed and then divided by the number of pairwise comparisons. This method is also widely adopted in clustering tasks.

Centroid Linkage: Calculates the distance between the centroids (geometric centres) of two clusters.

Any of the above methods can be selected based on the specific requirements of the clustering problem.

2.2.3.7.3 Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [61]

DBSCAN is a clustering algorithm that can find clusters of arbitrary shape and identify noise or outliers. The approach groups data points into classes based on their proximity to other points in the data. Two parameters control DBSCAN clustering:

eps: Defines the neighbourhood radius. Two points are considered neighbours if the distance between them is less than or equal to *eps*.

minPts: Determines the minimum number of points needed to form a cluster.

The parameters *eps* and *minPts* in DBSCAN are specific to the algorithm and define its clustering behaviour, though their exact values may vary across different software implementations (such as Scikit-learn). However, choosing these parameters is crucial and dataset-specific, as there is no universal value. Based on these criteria, points are categorised as follows:

Core Point: A point is considered a core point if it has at least *minPts* neighbours within its *eps* radius.

Border Point: A point is classified as a border point if it is reachable from a core point but has fewer than *minPts* neighbours within its own *eps* radius.

Outlier: A point is classified as noise, or an outlier, if it is neither a core point nor density-reachable from any core point. These noise points are typically grouped into a separate cluster,

as they can disrupt the clustering process. For a better understanding, refer to Figure 2.6 for DBSCAN visualisations.

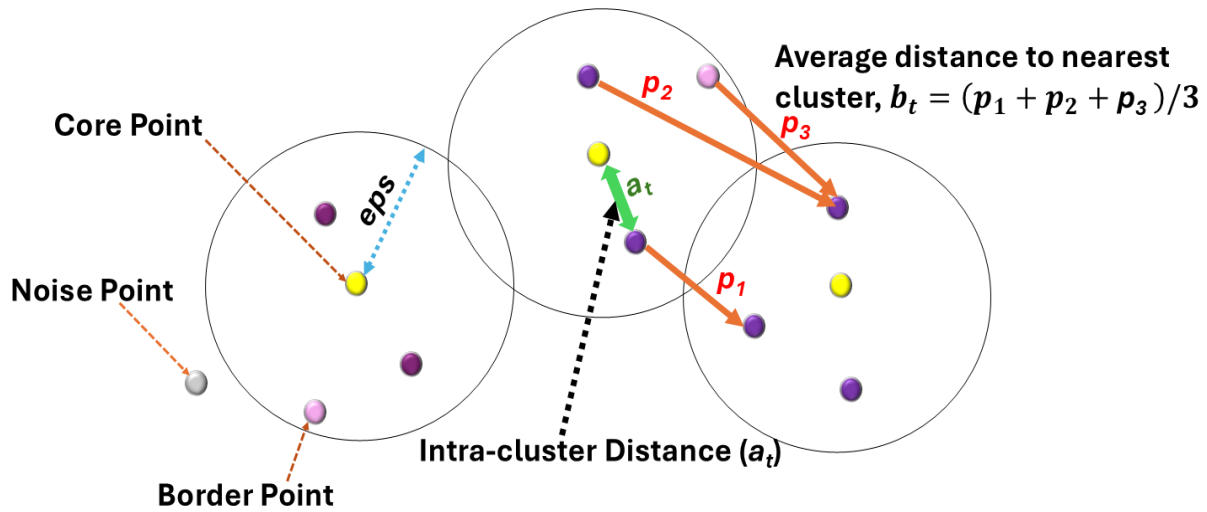


Figure 2.6. Conceptual illustration of DBSCAN clustering and cluster validity distance measures. The eps-radius defines the local neighbourhood used to classify points as core, border, or noise. For a representative data point, the intra-cluster distance (a_t) denotes the mean distance to all other points within the same cluster (cluster cohesion). The arrows labelled p_1 , p_2 , and p_3 illustrate example distances from the selected point to points in the nearest neighbouring cluster; in practice, distances to all points in that neighbouring cluster are considered when computing the inter-cluster distance (b_t), defined as their mean. These measures are commonly used to evaluate cluster cohesion and separation (i.e., silhouette analysis; described in **Section 2.2.3.7.1**).

2.2.3.7.4 Comparative Discussion of Clustering Approaches

K-means, hierarchical clustering and DBSCAN differ fundamentally in their clustering principles and suitability for different data structures. K-means is a centroid-based partitioning method that assumes clusters are compact and approximately spherical in Euclidean space. It requires prior specification of the number of clusters and performs best when clusters are well separated, and the dataset has been properly standardised [60].

Hierarchical clustering constructs a multi-level representation of cluster relationships using a dendrogram and does not require the number of clusters to be fixed in advance. It is particularly useful for exploratory analysis where the underlying structure of the data is of interest.

However, it may be sensitive to the chosen linkage criterion and can become computationally intensive for large datasets.

In contrast, DBSCAN is a density-based method that identifies clusters of arbitrary shape and explicitly detects noise or outliers. It does not require pre-specification of cluster number, but its performance depends on appropriate selection of the ϵ and minPts parameters.

DBSCAN is especially suitable when the dataset contains irregular cluster geometries or significant noise [61].

Overall, K-means is most appropriate for compact, well-separated clusters; hierarchical clustering is advantageous for structural interpretation; and DBSCAN is particularly effective for density-based grouping and outlier detection.

2.2.3.7.5 Gaussian Mixture Model (GMM) [62]

Gaussian Mixture Model (GMM) is a probabilistic clustering algorithm that defines a probability for each point in several clusters (soft clustering), thus appropriate if the distributions overlap. GMM describes data as a mixture of Gaussian distributions with distributions of random means, variances, and weights. GMM is based on the Expectation-Maximisation (EM) algorithm, which follows the following steps:

Step 1: Initialisation – Select the number of clusters K , and initialise the parameters: mean (μ_k), covariance (Σ_k), and mixing coefficient (π_k). These initial values are often determined using k-means clustering as a starting point.

Step 2: E-Step – Determine each data point's responsibilities (γ_{ik}), or the likelihood that it belongs to a cluster.

Step 3: M-Step – Calculate the new μ_k , Σ_k , and π_k using the responsibilities identified in the Step 2.

Step 4: Repeat – E/M steps until a stopping criterion is satisfied.

2.2.3.7.6 Mean Shift Clustering [62]

The density-based clustering algorithm Mean Shift eliminates the need to predetermine the number of clusters. It finds clusters by shifting data points in the direction of denser regions.

Step 1: Bandwidth Selection – Determine the size of the neighbourhood by choosing the radius (bandwidth). A smaller bandwidth results in more clusters, while a larger bandwidth produces fewer clusters.

Step 2: Initialisation – Assign a kernel to each data point.

Step 3: Iterative Shift – Iteratively move each kernel to the mean position of the points within its neighbourhood until convergence is reached.

Step 4: Cluster Assignment – Group points whose kernels converge to the same position.

Step 5: (Optional) – Merge peaks which are spatially near to one another to remove duplicate clusters.

Mean shift is most useful for finding clusters with irregular shapes without requiring prior knowledge about their number.

2.2.3.8 Optimisation Algorithms

2.2.3.8.1 Bayesian Optimisation Algorithm [63]

The goal of optimisation is to find the combination of values for an objective function that minimizes the loss, thereby producing the best possible outcome according to the defined objective. We can determine the minimum loss for straightforward low-dimensional functions by testing various input values to see which one results in the smallest loss. Two methods commonly used for this purpose are:

Grid Search: Creating a grid of input values and testing them all.

Random Search: Selecting values at random.

These basic approaches might work as long as the evaluations of the objective function ("evals") are affordable. Bayesian prediction models, such as Gaussian processes, are a natural choice for optimisation problems because they also provide the variance of the acquisition function. In materials science, Gaussian processes have been widely applied to tackle various structure optimisation and design challenges. The procedures for Bayesian Optimisation are as follows:

Step 1: Defining the Function to Optimise – A black-box function to optimise is defined as $f(\mathbf{x}_u)$. It is not necessary to understand the function's form, as $\mathbf{x}_u$ represents the value that globally optimises $f(\mathbf{x}_u)$. Depending on the problem, $f(\mathbf{x}_u)$ also referred to as the objective function, target function, or loss function. Generally, only the inputs ($\mathbf{x}_u$) and outputs ($f(\mathbf{x}_u)$) are known.

Step 2: Getting Started – Create an instance of a Bayesian optimisation object, describing the function $f(\mathbf{x}_u)$ to be optimised and the parameter bounds (*pbounds*). It is necessary to specify the minimum and maximum values that can be probed for each parameter to ensure this limited optimisation strategy functions effectively. Without significant fine-tuning, the Bayesian optimisation object will function well.

The primary parameters for optimisation are:

n_iter: The number of Bayesian optimisation steps to perform. Conducting more steps increases the likelihood of finding a good maximum.

init_points: The number of random exploration steps to undertake. Increasing exploration helps identify new data points.

The property optimiser provides access to the optimal set of parameters and the maximum goal value.

Step 3: Changing Bounds – If the specified parameter bounds are found to be insufficient during the optimisation process, they can be adjusted using the `set_bounds` method.

Step 4: Acquisition Function – The acquisition function $J(\mathbf{x}_u)$ is defined as:

$$J = \text{mean} + \kappa * \text{Standard deviation} \quad (12)$$

Here, the Gaussian process model outputs the mean and standard deviation ($\mathbf{x}_u, y_u$). The optimiser hyperparameter balances exploration and exploitation during the search for $\mathbf{x}_u$.

For small values of $J(\mathbf{x}_u)$, the mean dominates, focusing on searching around newly discovered local minima.

For large values of $J(\mathbf{x}_u)$, the standard deviation is emphasised, directing the search toward unexplored regions of $\mathbf{x}_u$. The acquisition function determines the next sample of $\mathbf{x}_u$ to be tested.

After completing the optimisation steps, the output consists of the optimal input parameters $\mathbf{x}_u^*$ that maximise or minimise the objective function $f(\mathbf{x}_u)$, along with the corresponding function value $f(\mathbf{x}_u^*)$. The Gaussian Process model also provides predictions, and uncertainty estimates for untested points.

2.2.3.8.2 Artificial Neural Network (ANN) [47], [64]

Machine learning, as a cohesive set of data-driven methodologies, significantly accelerates computational discoveries. Artificial intelligence (AI) has been described as the “fourth paradigm of science”. A neural network (NN) is a computational learning system that uses a network of functions to interpret and transform input data from one form into the desired output form. The concept of ANN is inspired by human biology, particularly the way neurons in the brain work together to process sensory information.

In general terms, NNs are a collection of algorithms designed to identify patterns, correlations, and information from data using a process that mimics the functioning of the human brain. The structure of a neural network consists of the following layers:

Input Layer: Nodes in the input layer receive information from the outside world, which the model uses to learn and make predictions. The input nodes pass data to the next layer, the hidden layer.

Hidden Layer: This layer contains neurons where all calculations on the input data occur. A neural network may have one or more hidden layers. The simplest network has just one hidden layer.

Output Layer: This layer produces the model's final output or predictions, based on the calculations from the hidden layers. The output layer can have one or more nodes, depending on the problem. For binary classification, the output layer typically has one node, whereas for multi-class classification, it may have multiple nodes.

In the first phase, the input layer passes data to the hidden layer, along with certain weights. Inputs x_1 , x_2 , x_3 , and so on, are transmitted, as shown in Figure 2.7. Neurons in each hidden layer are interconnected to receive all inputs. Once the inputs are received, the hidden layer performs calculations in two steps:

Step 1: All inputs are multiplied by their respective weights. The weight of each variable represents its gradient or coefficient and indicates its relative importance. A bias variable is added after the weights are assigned. The bias helps achieve the best model fit.

$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + b_z \quad (13)$$

Where w_1 , w_2 , w_3 are the weights assigned to inputs x_1 , x_2 , x_3 , and b_z is the bias.

Step 2: The linear equation is then passed through an activation function, which introduces nonlinearity to the model. Activation functions are crucial for enabling the model to learn complex patterns. Each hidden layer processes data in this way, and the final layer provides the ultimate output or prediction of the model after passing through all hidden layers.

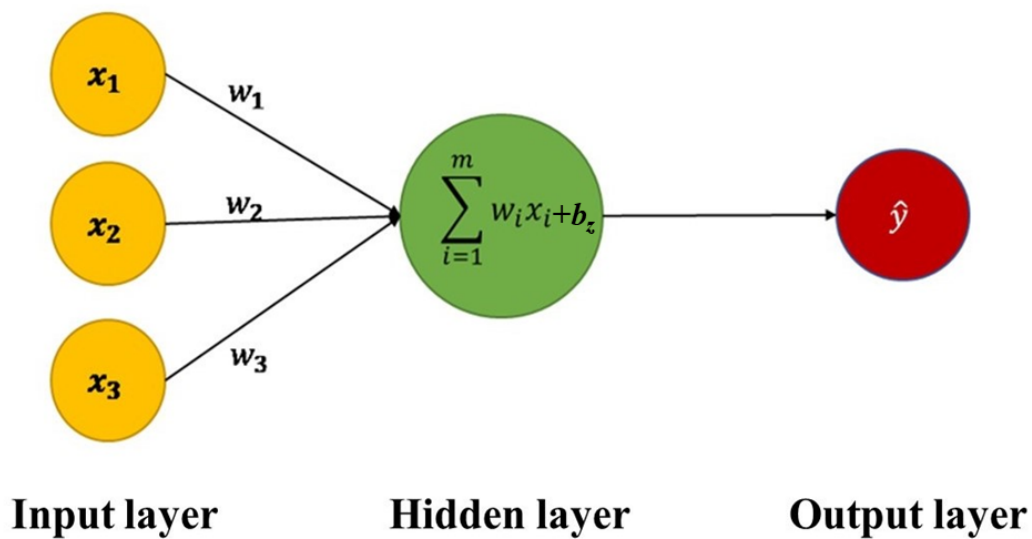


Figure 2.7. Artificial neural network

2.2.3.8.3 TOPSIS (Technique for Order Preference by Similarity to the Ideal Solution)

[65], [66]

TOPSIS is a multi-criteria decision-making (MCDM) method used to evaluate, prioritise, or select among multiple alternatives when decisions involve several conflicting criteria. MCDM techniques help balance trade-offs and provide a structured framework for decision-making. TOPSIS is based on the principle that the best alternative should have the shortest distance from the ideal best solution and the farthest distance from the ideal worst solution. The steps followed in the TOPSIS approach are as follows:

Step 1: Construction of the Decision Matrix – The alternatives and criteria are listed in tabular form. Each row represents an alternative, while each column represents a criterion.

$$E = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & \dots & e_{mn} \end{bmatrix}$$

Where: m : Number of alternatives, n : Number of criteria, e_{ij} : Performance value of the i -th alternative under the j -th criterion.

In the TOPSIS decision matrix, the alternatives (i.e., Alloys) are evaluated based on multiple criteria. In this research, the criteria include yield strength, fatigue strength, modulus of toughness, corrosion resistance, anodising response, and weldability, depending on the alloy series. These criteria represent key performance measures of the alternatives, allowing the decision-making process to balance trade-offs and select the most suitable option.

Step 2: Normalised Decision Matrix – The decision matrix is normalised to make the criteria comparable. Each element is computed as:

$$p'_{ij} = \frac{p_{ij}}{\sqrt{\sum_{k=1}^m p_{kj}^2}} \quad (14)$$

where p'_{ij} is the property value of the i -th alloy for the j -th property.

The normalised matrix becomes:

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mn} \end{bmatrix}$$

Step 3: Determine the Ideal-Best Solution, T^* , and Ideal-Worst Solution, T^- –

Ideal-Best Solution (T^*): The best value for each criterion (maximum for benefit criteria, minimum for cost criteria).

Ideal-Worst solution (T^-): The worst value for each criterion (minimum for benefit criteria, maximum for cost criteria).

$$T^* = \{v^+_1, v^+_2, \dots, v^+_n\}, v^+_j = \max(v_{ij}) \text{ for benefit, } \min(v_{ij}) \text{ for cost}$$

$$T^- = \{v^-_1, v^-_2, \dots, v^-_n\}, v^-_j = \min(v_{ij}) \text{ for benefit, } \max(v_{ij}) \text{ for cost}$$

Step 4: Distance from Ideal Solutions – Calculate the distance of each alternative from the ideal-best and ideal-worst solutions using Euclidean distance.

Distance from the Ideal-Best Solution (D_i^*):

$$D_i^* = \sqrt{\sum_{j=1}^n (p'_{ij} - T_j^*)^2} \quad (15)$$

Distance from the Ideal-Worst Solution (D_i^-)

$$D_i^- = \sqrt{\sum_{j=1}^n (p'_{ij} - T_j^-)^2} \quad (16)$$

Step 5: Compute the Relative Closeness to the Ideal Solutions – The relative closeness of each alternative to the ideal solutions is computed as:

$$C^i = \frac{D_i^-}{D_i^+ + D_i^-} \text{ where } 0 \leq C^i \leq 1 \quad (17)$$

Step 6: Ranking the Alternatives – The alternatives are ranked in order of their relative closeness (C^i) values. The alternative with the highest C^i is the best choice.

2.2.3.8.4 AHP (Analytic Hierarchy Process) [67]

The AHP method simplifies a complex decision by breaking it down into a hierarchy and assigning weights based on pairwise comparison of criteria and alternatives.

The following AHP steps will help to outline and give meaning to the problem with hierarchy: Break down the goal, criteria, sub-criteria, and alternatives.

Step 1: Conduct Pairwise Comparisons – Compare the importance of each pair of criteria with respect to its superordinate based on a comparison scale, for instance: 1 = equally important, 9 = extremely important.

Step 2: Construct the Pairwise Comparison Matrix – Build the pairwise comparison matrix, A using the relative importance measures obtained from the pairwise comparisons.

Step 3: Calculate the Priority Vector – Normalise the pairwise comparison matrix, A and compute the average of each row in the normalised matrix to obtain the priority weights.

Step 4: Check Consistency – Compute the Consistency Index (CI):

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (18)$$

Where λ_{max} is the maximum eigen value and n is the matrix size.

Step 5: Compute the Consistency Ratio (CR) – Where RI is the random consistency index. A CR value below 0.1 indicates acceptable consistency.

$$CR = \frac{CI}{RI} \quad (19)$$

Step 6: Synthesise Results – Combine criteria and alternative weights to rank the alternatives.

2.2.3.8.5 VIKOR (VIšekriterijumsko KOmpromisno Rangiranje) [67]

VIKOR is a method that focuses on the ranking and selection of alternatives according to their proximity to the ideal solution, based on the principle of compromise solutions in multi-criteria problems. Steps in the VIKOR method are as follows:

Step 1: Construct the Decision Matrix – Define the alternatives and criteria to be evaluated and organise them into a decision matrix, A .

Step 2: Normalise the Decision Matrix – Normalise the criteria values to enable effective comparison across different scales.

Step 3: Determine Ideal-Best and Negative-Ideal Solutions – Ideal-Best Solution (f^+): Maximum value for benefit criteria and minimum value for cost criteria.
Negative-Ideal Solution (f^-): Minimum value for benefit criteria and maximum value for cost criteria.

Step 4: Calculate Utility and Regret Measures –

Utility (s_i): Measures the total distance to the ideal solution:

$$s_i = \sum_{j=1}^n w_j \frac{f^+ - f_{ij}}{f^+ - f^-} \quad (20)$$

Regret (R_i): Measures the maximum individual distance to the ideal solution:

$$R_i = \max_j w_j \frac{f^+ - f_{ij}}{f^+ - f^-} \quad (21)$$

Step 5: Compute the VIKOR Index L_i –

$$L_i = v_r \cdot \frac{s_i - s^*}{s^- - s^*} + (1 - v_r) \frac{R_i - R^*}{R^- - R^*} \quad (22)$$

Where, v_r is the weight of utility (normally 0.5).

Step 6: Rank Alternatives – Rank alternatives according to S_i , R_i , and L_i , and identify the compromise solutions that are closest to the ideal solution.

2.2.3.9 Explainable AI technique: LIME Algorithm (Local Interpretable Model-agnostic Explanations) [68]

LIME is a technique to explain the predictions of complex, black-box machine learning models. It is designed to offer interpretable explanations for individual predictions, providing insights into why a model made a particular decision for a given data point. LIME works by approximating the behaviour of a complex model locally around the specific instance we want to explain. The process of LIME can be summarised in the following steps:

Step 1: Select the Instance to be Explained: Identify the specific data point, denoted as $\mathbf{x}_{\text{lme},0}$, for which the prediction from the black-box model $f(\mathbf{x}_{\text{lme}})$ requires explanation.

Step 2: Generate Perturbed Data – Perturbed instances are created by introducing small modifications into the original instance $\mathbf{x}_{\text{lme},0}$, such that samples $(\mathbf{x}_{\text{lme},0}, \mathbf{x}_{\text{lme},1}, \dots, \mathbf{x}_{\text{lme},n})$. The changes are creating new instances which are almost like $\mathbf{x}_{\text{lme},0}$ yet not exactly similar. It is normally performed through feature space sampling, where a perturbation of the original data point is done in small steps.

Step 3: Get Predictions on Perturbed Instances – To get the predictions, query the black-box model one at a time for each perturbed instance.

Step 4: Weight the Perturbed Instances – Assign a weight, g_i , to each perturbed instance depending upon its proximity to $\mathbf{x}_{\text{lme},0}$.

The weight is computed as follows:

$$g_i = \exp\left(-\frac{\text{dist}(\mathbf{x}_{\text{lme},0}, \mathbf{x}_{\text{lme},i})^2}{2\sigma^2}\right) \quad (23)$$

where $\text{dist}(\mathbf{x}_{\text{lme},0}, \mathbf{x}_{\text{lme},i})$ is the distance between $\mathbf{x}_{\text{lme},0}$ and $\mathbf{x}_{\text{lme},i}$, and σ is a scaling parameter.

Step 5: Fit Surrogate Model – Fit an interpretable surrogate model $f(\mathbf{x}_{\text{lme}})$ (e.g., linear

regression) to the weighted perturbed data. This surrogate model serves as a local approximation of the black-box model.

Step 6: Generate the Explanation – Use the coefficients or structure of the surrogate model to explain the black-box model's prediction. For example, in linear regression, the coefficients represent the importance of each feature for the prediction [68].

2.3 Implementation of Machine Learning (ML) and Artificial Intelligence (AI) in Materials Science

In recent years, interest in machine learning has grown significantly, driven by the availability of massive datasets, advancements in algorithms, and exponential increases in computing power. Today, materials science makes effective use of machine learning algorithms [69], [70], [71]. Historically, experiments have played a crucial role in identifying and characterising new materials. However, due to their high resource and equipment requirements, experimental research must be conducted over a lengthy period of time for a very small number of materials. As a result, key discoveries relied primarily on human intuition and experience. The prediction of new stable materials, the calculation of multiple material properties, and the acceleration of first-principal calculations are just a few examples of the numerous machine learning applications in materials research that have already been proven effective. Other fields have already been transformed by machine learning algorithms [69], [72]. The following are several applications of machine learning algorithms in materials science.

Property Prediction

A feedforward neural network with backpropagation was applied to forecast UTS and guide the microstructural design of Nb-Si alloys [47], [73]. The alloys were produced through heat treatment and directional solidification, and five key features were collected as inputs for the ANN. The model accurately predicted UTS values ranging from 343 to 1063 MPa, demonstrating strong generalisation capabilities. Sensitivity analysis further validated the results. By continuously backpropagating to minimise the loss function, the BP ANN enhances prediction accuracy, one of its key strengths [74]. Using this optimised ANN model, a novel method was proposed for designing the microstructure of Nb-Si alloys to achieve targeted UTS values. For example, the microstructure of a sample with an initial UTS of 343 MPa was modified using silicide design maps generated by the model, successfully reaching the desired UTS of 600 MPa [73].

In a similar study, the mechanical properties of A357 alloy were predicted using an artificial neural network (ANN) employing a backpropagation (BP) algorithm, with emphasis on how heat treatment influences mechanical behaviour [75]. The BP model enhanced prediction accuracy by continuously reducing errors, effectively reflecting the effects of heat treatment. This method demonstrated high reliability in forecasting the alloy's mechanical properties. Additionally, by examining iso-grams that combined ultimate tensile strength and elongation, the study offered detailed insights into the impact of aging parameters on these mechanical characteristics [75].

In a study, Pathan et al. [76] employed machine learning and image-based data analysis to predict the macroscopic stiffness and yield strength of unidirectional fibre composites when loaded across the fibres. Rather than performing computationally expensive simulations, they analysed images of 1,800 different microstructures using a two-point correlation function and subsequently reduced the data through principal component analysis. These features were then used to train a gradient-boosted regression model based on finite element simulation results, with model performance evaluated using a 10-fold cross-validation strategy. The model predicted stiffness and yield strength with an error of approximately 5% and revealed that the fibre volume fraction had the greatest influence on the material's performance. This approach provides a fast and cost-effective means of estimating material properties directly from microstructural images.

In another study on 17-4PH steels, the relationship between aging duration, temperature, and hardness was initially modelled using an artificial neural network (ANN). The final ANN model achieved a high R^2 value of 0.97, indicating excellent predictive performance for hardness values. The optimal aging time and temperature for achieving higher hardness were then determined using a genetic algorithm (GA), with the ANN model serving as the GA fitness function. According to the GA-ANN simulation, aging for 129 minutes at 464°C would result in a hardness of 45 HRC. The model was subsequently validated through heat treatment experiments using these predicted parameters. The experimental results produced a maximum hardness of 44 HRC, differing by only 1.8% from the model prediction, indicating a small model error [77].

Using a machine learning design system (MLDS), property-oriented compositional design was achieved for high-performance, copper alloys. Based on a database containing hundreds of samples, two types of backpropagation neural network (BP NN) models were developed to

learn the relationships between copper alloy compositions and their material properties, such as mechanical strength and electrical conductivity [78]. One model predicted alloy properties from their compositions (composition-to-performance, C2P), while the other predicted alloy compositions based on desired properties (performance-to-composition, P2C). The P2C model was employed to screen the composition space for specific property requirements, and the C2P model was subsequently used to recommend the optimal composition within a predetermined confidential range. The MLDS-designed high-performance copper alloys exhibited ultimate tensile strengths ranging from 600 to 950 MPa and electrical conductivity of 50.0% IACS. This approach enables the rapid and accurate design of multi-component, high-performance copper alloys [78].

In a conventional analysis for predicting creep behaviour of nickel-based superalloys, an artificial neural network (ANN) algorithm was utilised [79]. The study employed alloy composition, heat treatment, grain size, temperature, and applied stress as input parameters, with creep rupture life as the predicted output. The ANN architecture consisted of 25 input nodes, seven hidden layers, and a single output layer, chosen as the optimal dense layer configuration due to its minimal model error. The model demonstrated strong predictive capability for creep rupture life, achieving an R^2 value of 0.93 between predicted and actual values. Additionally, time-dependent deformation of nickel-based superalloys 276 and 617 was predicted using the ANN, with alloying content, applied stress, and operating temperature as inputs. The analysis considered temperatures ranging from 750–850°C and loadings of 10% and 25% of the yield stress. The model produced creep strain and creep strain rate as outputs. A notable limitation was the small size of the dataset; nevertheless, the technique successfully predicted creep behaviour with a high degree of accuracy [79].

A convolutional neural network (CNN) model was used in a study on crystalline perovskites to directly capture material properties from the atomic connections within the crystal [80]. CNNs are a type of deep learning model particularly effective at identifying patterns in structured data, such as images or spatially organised information, by using convolutional layers to automatically extract features. This technique enables a representation of crystalline materials that is both universal and easy to understand. The method proved highly effective, as training on 104 data points resulted in extremely accurate predictions of eight different crystal properties, as calculated by DFT, across various structure types and compositions. Additionally, this

method can be beneficial since it was possible to predict the contributions of atmospheric features to the properties of perovskites [80].

Evans et al. [81] investigated siliceous zeolites, predicting the logarithmic values of their bulk and shear moduli using gradient boosting regressors. Gradient boosting is an ensemble learning method that constructs a series of weak prediction models, usually decision trees, where each successive model aims to correct the errors of its predecessor, resulting in highly accurate predictions. The model achieved an error of 0.102 GPa for the bulk moduli and 0.0847 GPa for the shear moduli. Despite the training set consisting of only 121 zeolites, this approach was shown to be effective in finding these mechanical properties.

In one study, several machine learning (ML) and artificial neural network (ANN) models were developed to estimate the thermal resistance (R) at the interface between graphene and hexagonal boron nitride [82]. The models required only temperature, coupling strength, and in-plane tensile strains to predict the (R) values. The ML approaches tested included decision trees (DT), random forests (RF), polynomial regression, and linear regression. The ANN architectures evaluated were 3-10-1, 3-10-10-1, 3-20-1, and 3-20-20-1 layers. The linear regression model produced a high mean square error (MSE), limiting its accuracy in predicting (R). Polynomial regressions of second order and higher performed worse than linear regression but outperformed DT, RF, and the smaller ANN architectures. Overall, the most accurate prediction of (R) was obtained using the 3-20-20-1 layer ANN [82].

In a study, extremely randomised trees and deep neural networks (DNNs) were employed to generate data for constructing a hierarchical model aimed at predicting ternary parameters such as bulk and shear moduli, based on machine learning models trained on binary data. For ternary perovskite alloys, the extremely randomised tree algorithm produced mean absolute errors (MAEs) of 0.56 GPa for bulk modulus and 1.49 GPa for shear modulus, using composition and temperature as input features and elastic properties as targets. The results demonstrated that this approach is effective for predicting the elastic properties of ternary alloys [83].

Xiong et al. [84] examined the mechanical properties, such as fatigue strength, tensile strength, hardness, and fracture strength, of 360 steel samples obtained from the Japan National Institute of Materials Science database. They employed various machine learning algorithms and identified random forest regression as the most effective for predicting steel performance. They found that temperature and compositions such as Cr, Mo, and C were the most influential

factors affecting these properties. In addition, symbolic regression produced clear mathematical expressions relating these features to the mechanical properties, improving interpretability. The random forest model demonstrated high validation accuracy, with $R^2 > 0.9550$ and $RMSE < 3.25\%$.

In a study on Ni-based superalloys, two adaptive numerical modelling techniques were employed to estimate the fatigue strength [85]. The capabilities of a Bayesian neural network and a neuro-fuzzy network to automatically adapt their complexity to the dataset were compared. In both cases, the models successfully predicted strength values despite limitations in the data. However, the neuro-fuzzy modelling approach was argued to offer distinct advantages over conventional neural networks. Its mathematical complexity could be constrained to account for limited data availability, and the linguistic fuzzy algorithm allowed model evaluation without extensive interrogation, even when using randomly designed datasets. Using the additive neuro-fuzzy network structure, unnecessary inputs could be removed, enabling the creation of simple sub-networks that captured global output trends. These features were critical for the final confirmation and verification of information derived from numerical data. Furthermore, because transparent linguistic rules were provided, neuro-fuzzy networks may, in some cases, require less data to achieve a stable solution and can be more easily assessed within the framework of current physical understanding [85].

Additionally, there are programs like ElemNet that use elements as input to predict material properties. ElemNet is a model built on a deep neural network (DNN). It generates quick and accurate predictions by automatically extracting the physical and chemical relationships and similarities between elements. Chemception is a model based on a convolutional neural network (CNN) that predicts toxicity, activity, and solvation properties by converting raw compound data into 2D images, similar to the approach used by ElemNet [86].

Physical Property and Treatment Condition Problems

Kusaba et al. [87] tackled the challenge of predicting stable crystal structures from a given chemical composition. Traditional first-principles methods are computationally expensive, especially for complex systems. Their approach uses metric learning to identify templates with similar known structures from a database and applies element substitution to generate candidate structures. The method achieves around 96.4% accuracy in determining structural similarity and only requires local relaxation calculations, avoiding full ab initio computations. While it

relies on available templates, about 3% of structures in the Materials Project database lack them, it can predict 50–65% of all crystal systems, including unique structures without direct templates. As databases grow, its applicability is expected to expand, offering an efficient alternative to conventional crystal structure prediction.

Pilozzi et al. developed a neural network model to design photonic devices [88]. The target properties were edge states at a specific frequency t , defined by a set of structural parameters. Both forward and backpropagation ANN models were employed in tandem to find the solution. Initially, the backpropagation neural network was used to determine the structural parameters required for an edge state with frequency t . Initially, a backpropagation neural network was employed to identify the structural parameters needed for an edge state at a specific frequency, t . To simplify the analysis, some categorical features were omitted from the network; nevertheless, a separate neural network was effectively trained for each categorical variable. The direct neural network then uses the structural parameters provided by the backpropagation (BP) network as input to generate a new frequency, referred to as $\text{dir } t$. The structure parameters from the BP neural network were approved if the difference between the two frequencies was less than a predetermined threshold $\text{ind}t$. Structures that were not relevant were removed from the BP neural network using this self-driven method to predict the photonic device frequency [88].

As a data processing application in materials science, Raccuglia et al. trained a machine learning model on failure data from unsuccessful experiments [89]. They combined experimental data from hydrothermal synthesis procedures that failed or were only moderately effective to construct a machine learning model to predict the crystallisation of template vanadium selenite crystals. This algorithm successfully predicted the conditions in which new organic and inorganic compounds were formed and passed the conventional manual analysis, with an accuracy of 89% [89].

A study described a general machine learning approach employing artificial neural networks (ANNs) to predict the properties of inorganic materials, including band gap energy and glass-forming ability for both amorphous and crystalline forms. Neural networks have also been applied to control process parameters in ceramic casting by predicting product quality based on raw materials, casting conditions, and line settings. By adjusting these parameters, the approach

enables the production of superior castings, offering an effective machine learning solution to reduce waste in the casting process [90].

To evaluate ultrasonic testing results for determining the thickness of an adhesive fillet in aluminium T-peel joints, Bork and Challis [91] employed a neural network to classify the fillets into three categories. As no straightforward analytical method existed for analysing the Lamb waves used in the evaluation, the neural network approach was adopted. The network achieved a 95% success rate in classifying fillet sizes from an independent test dataset. Furthermore, researchers have utilised neural networks to predict material properties such as the austenite start and completion temperatures in steel, based on chemical composition and heating rate, with results showing strong agreement with transformation theory. Neural networks have also been applied to investigate the influence of loading angles on the measured strength of fibre-reinforced metal matrix composites (MMCs) [92].

Unsupervised Learning and Explainable AI Problems

The LIME approach is frequently used in research to improve the interpretability of black-box algorithms, including supervised and unsupervised learning algorithms. It helps explain why certain results were produced. The method helps in measuring trust in prediction and model selection, improving untrustworthy classifiers, and identifying why a model is not trusted.

Park et al. [93] successfully developed an AI-driven deep neural network (DNN) model to design high-strength 7xxx series aluminium alloys. Despite the complexity of multiple processing steps, the model accurately predicted mechanical properties based on seven alloy compositions and five fabrication processes, even with a small dataset. The model also predicted optimal fabrication conditions, which were then experimentally tested to produce alloys with the anticipated properties. This approach significantly reduced the time and cost involved in alloy design. To increase the interpretability of the DNN model, they employed the LIME algorithm. LIME attributed the positive or negative effect of each variable to the mechanical properties and output the metallurgical reasoning behind predictions. LIME, for example, showed how the precipitating effect of Zr as Al_3Zr contributed positively to YS, UTS, and EI. Although LIME was effective at identifying the contributions of individual variables, it fell short in capturing the global interactions between specific elements and fabrication processes.

Despite this, the explanations of LIME were useful to the alloy designers, as they offered real-time feedback and guided redesign.

Using non-destructive testing data, Golowanov et al. [94] developed a recursive algorithm to cluster semiconductor devices into homogeneous groups by maximising the silhouette score of the K-means clustering algorithm. They introduced the concept of batch homogeneity, which is defined based on the silhouette score criterion. Their research addressed the challenge of creating uniform production batches of integrated circuits for the space industry. The proposed method successfully separated a mixed batch of semiconductor devices into homogeneous groups with 100% accuracy and precisely determined the number of these groups. Its effectiveness was confirmed through computer analysis, which also demonstrated a perfect success rate in clustering the devices.

Similarly, Kazakovtsev et al. [95] developed clustering methods to divide heterogeneous batches of EEE semiconductor devices into homogeneous groups based on electrical testing data. Traditional methods such as K-means can be effective, but they require predefining the number of clusters and are highly sensitive to initial seeding, which makes their results unstable in practice. Since the number of clusters is generally unknown for EEE production batches, alternative approaches that do not require predefined cluster counts are needed.

To address these issues, the authors introduced a genetic algorithm combined with a greedy agglomerative heuristic, capable of solving series of K-means or p-median clustering problems and producing more stable results. Although post-processing based on a chosen criterion is still needed to determine the final number of clusters, the method showed high accuracy and practical success when applied to real semiconductor device data from a specialised testing facility for space applications [95].

In an insightful study on superconductivity, scanning tunnelling microscopy data were analysed to extract information on electronic interactions within the spin-density wave regime using K-means clustering [59], PCA [58], and Bayesian linear unmixing. This approach made a significant contribution to advancing the understanding of unconventional superconductivity [96].

Yeung et al. [97] demonstrated that convolutional neural networks (CNNs) trained to predict the electromagnetic responses of metal–dielectric–metal metamaterials, including complex

freeform designs, can be interpreted using SHAP (SHapley Additive exPlanations). This approach provides deeper insight into the underlying physics of nanophotonic structures. By applying SHAP, they identified the significance of specific spatial regions within nanophotonic structures in relation to the presence or absence of absorption peaks. Their research underscores the potential of machine learning techniques to advance optics and photonics design and to facilitate discoveries in fundamental physics. The contribution of each feature to the model's predictions was illustrated through relative SHAP values, with feature importance ranked by summing SHAP weights across all instances [97].

Additionally, studies have shown that principal component analysis (PCA) is an effective tool for correlating clusters across various datasets [98], [99], [100], [101]. Rossiter [102] employed principal component regression on infrared spectra of concrete samples to predict their material properties. Godin et al. [103] conducted tensile tests on pure resin and polyester/glass fibre unidirectional composites, identifying mechanisms such as matrix cracking and interfacial decohesion using K-means clustering. Although accurately distinguishing signals related to debonding from those associated with fibre failure proved challenging, the results were nevertheless considered promising.

This overview demonstrates that in many materials' science studies, including alloy research, PCA-K-means clustering and the LIME XAI technique have become prominent methods for analysing machine learning outcomes [104].

Optimisation Problems

Regression analysis and Bayesian optimisation are two machine learning design algorithms that can accelerate the discovery of new materials by integrating compositional design, property prediction, and machine learning modelling. A data-driven approach was employed to identify BaTiO₃-based piezoelectrics with a high piezoelectric coefficient, d_{33} . In this method, a surrogate optimisation model predicted d_{33} with an associated uncertainty, followed by a design step to select the next best compound for synthesis. Using training data from several previous experiments, they evaluated various combinations of model and design selection methods, including regression, Bayesian optimisation, and effective global optimisation, and identified the top two performers to guide future trials aimed at designing new piezoelectrics with the desired d_{33} . Five iterations of their adaptive design approach were conducted, synthesising four novel compounds in each iteration based on the two different design selection criteria.

Compared to the best compound in the training data, BCT-0.5BZT, which had a d_{33} of 610 pC/N, the best novel compound discovered, $(\text{Ba}_{0.85}\text{Ca}_{0.15})(\text{Ti}_{0.91}\text{Zr}_{0.09})\text{O}_3$, exhibited a d_{33} value lower than that of the training dataset. The notably high d_{33} value of BCT-0.5BZT was challenging to replicate with any substitute model, emphasising the necessity of combining physics-based methods with purely data-driven strategies, as applied in this study. Nonetheless, their model successfully fitted the majority of the available d_{33} data [105].

In the study of multiple complex alloys, mechanical and electrical properties were enhanced by optimising alloy compositions [106]. To identify the most important alloy components, several feature selection methods were employed, including correlation screening, recursive elimination, and exhaustive screening. Using the features selected through correlation analysis, a machine learning model was developed. Iterative Bayesian optimisation was then applied to design the alloy compositions. Through feature screening, five key alloy factors influencing hardness (HV) and six key factors affecting electrical conductivity (EC) in precipitation-strengthened copper alloys were identified. Both the “HV – key alloy factors” model and the “EC – key alloy factors” model achieved prediction errors below 7% and 9%, respectively. By combining Bayesian optimisation with iterative optimisation tests, novel copper alloys were successfully designed. One such alloy, Cu-1.3Ni-1.4Co-0.56Si-0.03Mg, exhibited a measured ultimate tensile strength (UTS) of 858 MPa and an electrical conductivity (EC) of 47.6% IACS, demonstrating an excellent balance of mechanical and electrical properties. This approach simultaneously improved both properties, surpassing the performance of previously reported precipitation-strengthened copper alloys. Similar improvements in performance in alloy design have been reported by other researchers using these methods [107].

To maximise the band gap of $\text{A}_{10}(\text{BO}_4)_6\text{X}_2\text{A}$ compounds, Lookman et al. [108] employed support vector regressors (SVRs) combined with effective global optimisation. This approach compared the performance of two feature sets: one based on variations in Pauling electronegativity and the other on Shannon ionic radii. Interestingly, the design based on the ionic radii performed better, identifying the ideal material after 22 material iterations (13 materials in the initial training set, 9 selected by the design algorithm), compared to 30 iterations for the electronegativities, despite having a significantly higher error in the machine learning model (0.54 eV compared to 0.19 eV for the electronegativities). The three elemental features (P, V, and As) that were considered for the B-site were P, V, and As. P supplied noticeably greater band gaps than the other elements and had a distinct ionic radius, while P and As had roughly the

same electronegativity. Using this knowledge, the algorithm rejected all compositions lacking phosphorus (P) at the B-site. This example illustrates that the algorithm with the highest predictive accuracy does not always yield the best results for an optimal design. While the ideal composition was found after a single iteration when combining both predictors and input variables, the model's mean absolute error (0.21 eV) remained somewhat higher than that of the model based solely on electronegativity.

Ling et al. [109] tackled a high-dimensional materials design problem (in terms of input space) using the random-forest framework FUELS, which incorporates a bias component to account for noise and missing degrees of freedom, thereby addressing uncertainty. FUELS demonstrated excellent performance when tested on four datasets, magnetocaloric, thermoelectric, superconductors, and thermoelectric, with descriptor counts of 54, 54, 56, and 22, respectively. This performance was nearly an order of magnitude better than that of the Bayesian framework COMBO, and random sampling. Various metrics were used to compare different modelling algorithms and selection procedures, with optimisation algorithms, measured by the number of iterations needed to discover the ideal material, being the most commonly employed statistical method in materials research. Although this metric has some value, opportunity cost (the distance between the current best and the overall best) or the number of experiments required until the current best material is designed within a given feature space (for example, 1%) is preferable. This strategy is also more frequently employed in many studies, as mentioned earlier.

In several studies, Multi-Criteria Decision-Making (MCDM) methods have been applied to material selection. For instance, Khorshidi and Hassani [110] compared the TOPSIS and PSI methods for selecting Al/SiC powder metallurgy composites and found that both identified the Al-5% SiC composite, with a particle size of 16 μm , a relative density of 90%, and a milling time of 12 hours, as the best option. While TOPSIS used the Analytic Hierarchy Process (AHP) to assign weights, the simpler PSI method did not require weight allocation. They concluded that the PSI method offers a more straightforward, efficient, and easily understandable approach compared to the more complex TOPSIS, highlighting its practical benefits in material selection [110].

Singaravel et al. [111] conducted research focused on optimising turning process parameters to minimise microhardness and surface roughness while maximising the material removal rate (MRR). The key input variables included the type of coated tools, cutting speed, feed rate, and depth of cut, while the response variables were microhardness, surface roughness, and MRR.

The optimum parameters were identified as a cutting speed of 179 m/min, a feed rate of 0.26 mm/rev, and a depth of cut of 1.8 mm. Their study employed AHP to determine weight factors for the parameters, introducing a robust methodology for process optimisation. The approach recognised that precise control over machining parameters can significantly enhance manufacturing quality and productivity [111].

Yang et al. [112] used the TOPSIS method alongside various normalisation techniques to select the best material. They introduced performance ratings based on entropy and the variation coefficient to compare the effectiveness of different normalisation methods. By combining results from multiple normalisation approaches, they developed a final ranking system for selecting the optimum tribological coating material. The authors concluded that the normalisation method with the highest performance score should be preferred and emphasised the importance of considering practical factors such as productivity, cost, and safety in material selection. Furthermore, they highlighted that both the choice of normalisation method and attribute weighting significantly affect outcomes, underscoring the need for careful determination of these factors to ensure accurate material selection.

Narrowing the Solution Space and Addressing Recycling Problems

Some initiatives aimed to reduce the solution space of materials; however, by designing new alloys to replace lower-performing standard ones, they effectively expanded the solution space once again. In one study, machine learning was employed to accelerate the discovery of metallic glasses by narrowing down the search space. A predictive ML model, trained on data from known metallic glasses, was used to rapidly filter the vast field of potential materials down to a few promising candidates. Over 20,000 new metallic glass compositions were screened, leading to the discovery of three useful new alloys. This process involved iterative feedback, where model predictions were experimentally tested, and the results were fed back into the dataset for retraining, continuously improving the model's ability to identify favourable compositions [113], [114], [115].

The SALEMA project focused on creating new aluminium alloys with improved performance and sustainability, particularly for automotive applications. This was achieved by minimising the use of critical raw materials such as silicon and magnesium and by integrating scrap-metal recycling to promote a sustainable circular economy. Although the novel alloys were designed

for specific applications, most notably electric vehicles, the introduction of new grades inevitably increased the diversity of alloy compositions available. All the increased numbers of alloy grades introduce complexities into the manufacturing and recycling processes since each novel alloy requires special processing, handling, and potential recycling methods. Thus, while the SALEMA project sufficiently addressed the issues of sustainability and material efficiency, it also diversified the industry in terms of alloy grades [115].

One proposed solution was a “uni-alloy” concept, an aluminium composition specifically designed with broader tolerance to residual elements, making it suitable for multiple applications and simplifying recycling streams [116], [117].

Artificial intelligence and machine learning are also increasingly applied to scrap sorting, alloy classification, and process control, overcoming many of the limitations of traditional methods. Campanella et al. [118] combined LIBS with artificial neural networks (ANN) to classify wrought aluminium scrap alloys into three groups (5xxx+6xxx, 7xxx, and others), providing a rapid and reliable automated sorting method. Similarly, Díaz-Romero et al. [119] used deep learning-based computer imaging to distinguish cast from wrought aluminium scrap, achieving high classification accuracy and highlighting the potential of vision-based systems.

Beyond aluminium, a review [120] surveyed advances in AI and robotics for municipal waste sorting, outlining practical applications, challenges, and lessons learned from large-scale implementations, and emphasizing their role in enhancing recycling efficiency in the circular economy. More recently, Auer et al. [121] developed an AI-based metal-scrap sorting system specifically for tool steels and high-speed steels. Their approach demonstrated higher accuracy and faster identification compared to conventional methods, highlighting its potential to reduce labour demand and improve operational efficiency in scrap-sorting processes. However, the system was less robust when applied to real industrial scrap conditions, where greater variations in composition and surface quality reduced its reliability.

Comparable efforts have been reported in aluminium scrap sorting [122], [123] and bin sorting [124], [125] and recycling [126]. Although these works were able to sort aluminium scrap, the models do not differentiate between aluminium alloy grades, detect mixed-alloy compositions. While these approaches improve sorting efficiency and demonstrate the potential of AI- and ML-driven systems for smarter separation and resource management, they do not enable directed recycling of real industrial scrap.

2.4 ML/AI Implementation in Aluminium Alloy Development

Machine learning (ML) is emerging as a promising approach to forecast the properties of aluminium alloys while reducing the need for time-consuming testing. Yet, difficulties in digitising features from different manufacturing processes have limited the development of ML models for wrought aluminium alloys. Consequently, prior research has focused on specific alloy groups, restricting the applicability of such models to a broader range of wrought alloys.

To tackle the complexity arising from manufacturing processes in wrought aluminium alloys, Hu et al. [127] introduced a novel feature engineering method, termed procedure-oriented decomposition (POD), to improve the prediction framework. Their model incorporated both chemical compositions and manufacturing procedures as features. Using a support vector regressor (SVR), they established a correlation mapping between these features and the mechanical properties of wrought aluminium alloys. Their prediction framework showed the ability to build novel alloys with a higher accuracy. The algorithm was developed to establish a quantitative mapping between chemical composition and manufacturing processes, including processing parameters and methods, used as input features, and mechanical properties, specifically yield strength, ultimate tensile strength, and elongation, referred to as the targeted properties, for various wrought aluminium alloys. Through feature engineering, the complex feature set was effectively transformed into robust components possessing the desired characteristics, resulting in faster computation and improved prediction accuracy.

In a study Li et al. [128] employed a hybrid approach combining machine learning and experimental validation to study the mechanical properties of 7xxx series aluminium alloys. Three neural network models, backpropagation (BP), BP optimised with a genetic algorithm (BP-GA), and radial basis function (RBF), were developed to predict how alloy composition and processing parameters influence yield strength, tensile strength, and elongation. The BP-GA model achieved the highest prediction accuracy ($R = 0.95$, AARE = 4.28%, RMSE = 0.087) and was employed for multi-objective optimisation of mechanical properties. Experimental validation using SEM, EBSD, HRTEM, and tensile testing confirmed the model's predictions, with precipitation strengthening observed after ageing treatment. This study demonstrates that machine learning, combined with targeted experiments, can effectively guide alloy design and property optimisation, providing a reliable framework for developing high-performance aluminium alloys.

An ANN model (with 259 samples) was utilised by Dey et al. [129] to forecast the mechanical characteristics of age-hardenable wrought Al alloys. A new age-hardenable aluminium alloy composition was developed using composition and tempering conditions as input parameters. Unfortunately, the experimental mechanical characteristics of this novel alloy greatly deviated from the prediction, suggesting that the trained ML models' generalisation (i.e., their capacity to make predictions about new samples) was insufficient. This error was probably caused by the uneven data points in the distribution range and the insufficient sample size. Additionally, this ANN model ignored the manufacturing procedure (such as rolling, extrusion, etc.) and production shape (e.g., plate, bar, etc.) and only considered numerical processing factors, such as heat-treatment temperature and time. Although it is challenging to include these non-numerical (i.e., categorical) inputs in ML models, those have a considerable impact on the alloy characteristics. Because only a few processing features were involved, the ML models previously discussed can only be employed for a limited range or specific types of Al alloys.

Altinkok and Koker [130] successfully applied a BP-ANN model to predict the bending strength and hardness of aluminium matrix composites (AMCs). Various SiC particle sizes were used as input parameters for training the neural network, while bending strength, hardening behaviour, and porosity of the resulting AMCs served as output targets. The network was trained using the prepared learning set, and its accuracy was subsequently validated with test data. The results demonstrated that the model could reliably predict bending strength, hardness, and porosity for any specified SiC particle size range in the fabricated AMCs.

Tang et al. [131] employed support vector regression (SVR) to predict the strength of the Al-Cu-Mg-Ag alloy. Under identical training conditions, their SVR model demonstrated higher prediction accuracy compared to a backpropagation artificial neural network (BP-ANN) when estimating tensile strength and hardness of Al-Cu alloys, as well as providing composition design recommendations to achieve target values.

To explore the potential of machine learning methods such as random forest regression, M5P tree regression, and artificial neural networks (ANN) for predicting the ultimate tensile strength (UTS) of AA6082 friction stir welded (FSW) joints, three regression models were developed [132]. The outcomes of these models were compared to identify the optimal approach. The study's results supported the use of these techniques for evaluating FSW joint performance. Additionally, a one-factor ANOVA verified the significance of all models in predicting UTS

values. For the dataset used, random forest regression outperformed both M5P tree and ANN models. Sensitivity analysis revealed that rotational speed was the most influential parameter.

In their study, Li et al. [133] examined 7xxx series aluminium alloys using a machine learning-based model aimed at optimising both composition and processing parameters. By implementing an efficient processing approach, they developed an optimised alloy containing yttrium that was compositionally lean while achieving a high ultimate tensile strength of 952 MPa and 6.3% elongation. Beyond the commonly observed irregular-shaped particles, they found that the $\text{Al}_8\text{Cu}_4\text{Y}$ phase in wrought 7xxx T6 alloys also formed a nanoscale network along sub-grain boundaries. Their findings demonstrated that machine learning can be effectively used to identify 7xxx series alloys with excellent mechanical properties.

Devi et al. [134] studied three machine learning methods, Linear Regression (LR), K-Nearest Neighbors (KNN) [135], and Artificial Neural Networks (ANN), for predicting the mechanical properties of aluminium alloys. The training and testing datasets were sourced using the KNN method, which demonstrated superior accuracy in predicting tensile strength and hardness values. In comparison to the other two algorithms, ANN produced more precise findings when predicting yield strength. Linear regression produced good results, but it was only applicable to linear relationships and was outlier sensitive. With high-dimensional data, the KNN algorithm did not perform well due to its sensitivity to noise, missing values, and outliers.

Durodola [70] reviewed the utilisation of machine learning across multiple areas, including material design, processing, characterisation, fabrication, and environmental impact assessment. He emphasised that no single algorithm is suitable for every problem. Although the application of deep learning for microstructure characterisation remains limited, there is considerable potential for automating quality control during material processing.

Fang et al. [136] developed a novel machine learning model using the eXtreme Gradient Boosting (XGBoost) algorithm to assess the web crippling behaviour of perforated roll-formed aluminium alloy (RFA) unlipped channels, both fastened and unfastened, under interior two-flange stress. An elastoplastic finite element (FE) model generated 1080 data points to train the XGBoost model, which was then validated against 30 published experimental results. Comparison with numerical failure loads indicated that the XGBoost model achieved a prediction accuracy of 94%, outperforming both Random Forest and linear regression approaches. The performance of existing design standards, including those from the American Iron and Steel

Institute (AISI), Australian/New Zealand Standards (AS/NZS 1664.1; AS/NZS 4600:2018), and Eurocode, was evaluated using web crippling strengths obtained from the XGBoost model, experiments, and finite element analysis (CEN 2007). The results demonstrated that the strength of perforated RFA unlipped channels exceeded predictions from earlier design criteria. Based on a parametric study, new equations for web crippling strength and reduction factors were proposed, and their accuracy was confirmed through reliability analysis, which verified that they met the required reliability standards.

One of the first "user interactive" ML models for Al-alloy design was reported by Mangos et al. [32]; it predicted properties as output, including tensile yield strength, ultimate tensile strength, and elongation. Their work also showed that a piece of software with a graphical user interface (GUI) that can perform alloy design tasks and is simple to use. A user might estimate the alloy composition and thermomechanical processing necessary to meet a set of property targets by using a straightforward, yet efficient optimiser used in this software.

2.5 Conclusions

2.5.1 Aluminium and Aluminium Alloys, and Recyclability Demands

This literature review provides a broad overview of aluminium and its alloys, highlighting why aluminium is one of the most widely used metals today, with demand continuing to grow. Aluminium's combination of light weight, high strength, and other advantageous properties makes it highly versatile. Compared to steel, aluminium offers superior resistance to rust, making it particularly well-suited for marine and aerospace applications. When combined with various alloying elements and subjected to different tempering conditions, aluminium's properties are further enhanced, allowing it to meet a wider range of application demands. At present there are two main classes of aluminium alloys, i.e., wrought and cast aluminium alloys, depending on the type of alloying elements and tempering conditions applied. Each class consists of eight series of aluminium alloys, with each series containing approximately one hundred alloys that vary in the composition of major elements such as copper, magnesium, zinc, manganese, and silicon, along with impurities, often differing only slightly in their composition ranges. Due to the very narrow composition ranges and closely related properties of aluminium alloys, having so many alloys within each series is unnecessary, as it complicates alloy selection and recycling.

Also, due to the large consumption of aluminium, amounting to millions of tonnes of primary aluminium annually, there is a pressing need to increase the use of aluminium from scrap. This is because producing primary aluminium requires significantly more energy compared to recycling, making aluminium scrap a preferable and more cost-effective option both energetically, environmentally, and economically.

Therefore, there is an urgent need to resolve this problem by narrowing down the search space of aluminium alloys that will lead to a lesser number of alloy grades for various applications and will cover a wider range of supply demands.

2.5.2 Machine Learning/Artificial Intelligence

Machine learning (ML) and artificial intelligence (AI) are rapidly growing fields within computer science and mathematics, fundamentally rooted in programming and statistics for predictive analysis. As highlighted in the literature review, AI and ML are closely related, with ML considered a subset of AI. Artificial intelligence enables a computer to mimic human thinking and perform tasks independently, while machine learning is the technique through which the computer system develops this intelligence.

ML and AI have various capabilities, such as predictive analysis, recommendation systems, speech recognition, image and video processing, and better and faster decision-making. These capabilities make them more advanced at solving complex real-world problems with less time consumption. Mostly, ML and AI are implemented using the Python programming language, due to its extensive libraries such as NumPy, Scikit-learn, Pandas, Keras, and Matplotlib, which simplify many complex tasks. Machine learning typically follows a step-by-step process: data collection → data preparation → model selection → model training → prediction. After collecting data relevant to the task, the most critical step is selecting the appropriate model or algorithm. ML algorithms are broadly divided into two categories: supervised learning and unsupervised learning. Under these categories, there are various algorithms such as regression analysis (to establish relationships between dependent and independent variables), classification analysis (to categorise data into classes), clustering analysis (to group data based on feature similarity), and optimisation algorithms used for predictive analysis and iterative learning. The need for these algorithms depends on the task in hand. They have the ability to make complex tasks easier and can recognise the patterns which the human eye cannot reach. ML and AI have significantly impacted materials science, particularly in composition and property

optimisation, accelerating the discovery of new alloys. These challenges can often be addressed within just a few iterations of machine learning algorithms, as discussed in **Section 2.3: Implementation of ML/AI in Material Science**. Numerous studies have demonstrated the strong potential of ML and AI in this field. Consequently, it is now possible to design and optimise alloys efficiently through a few logical lines of code, greatly speeding up the alloy discovery process.

CHAPTER 3: METHODOLOGY OF STUDY

3.1 Preliminary Analysis for Model Selection

A detailed examination of several techniques applied during the preliminary analysis phase was conducted to identify the most suitable approach for alloy selection and recycling, as illustrated in Figure 3.1. This initial analysis involved range of algorithms, including several clustering algorithms such as K-means clustering, hierarchical clustering, Gaussian Mixture Models (GMM), mean shift clustering, and DBSCAN clustering (**Section 2.2.3.7**). Additionally, multi-criteria decision-making (MCDM) methods such as TOPSIS, VANCOR, and AHP were investigated, alongside more complex algorithms like Artificial Neural Networks (ANN) and Bayesian optimisation (**Section 2.2.3.8**). Each method was then evaluated for its effectiveness in categorising alloys into meaningful groups and facilitating optimal alloy selection and recycling.

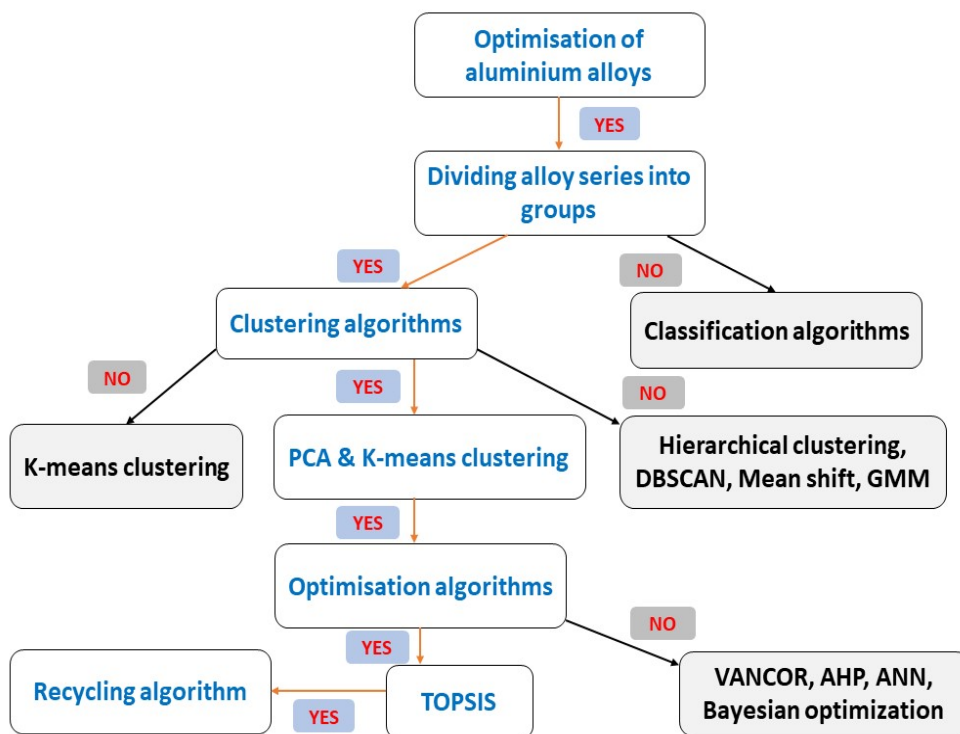


Figure 3.1. Preliminary analysis of algorithms for methodology selection

Clustering Algorithms for Alloy Classification

In this project, clustering algorithms were used instead of classification algorithms because the dataset did not contain predefined labels or target categories. The primary objective was to identify patterns within the dataset of various alloys and to determine natural groupings based on their compositions and mechanical properties. Since classification models require predefined target labels, clustering provides a more appropriate approach, allowing the identification of inherent data structures without reliance on pre-existing categories. Through clustering, groups of alloys with similar compositions and properties were identified, which can guide material design decisions and inform future experimental studies.

Dataset Description and pre-processing

For this analysis, a curated dataset of 41 wrought 6xxx series aluminium alloys was compiled to evaluate the performance of various clustering algorithms. The complete raw dataset containing all compositional, mechanical, and qualitative variables used in this study is provided in **Appendix 11.3.1.1**. The alloys were sourced from the Encyclopaedia of Aluminium and Its Alloys [137], the ASM Speciality Handbook [1], the MakeItFrom website [138], and various literature sources.

To ensure consistency in mechanical property comparison and to reflect common industrial practice, all alloys were considered in the T6 extrusion temper condition.

The dataset comprised both quantitative and qualitative variables describing chemical composition, mechanical behaviour, corrosion resistance, anodising response, and weldability performance in total 13 features. The numerical composition ranges were: $0.35 \leq \text{Si} \leq 1.35 \%$, $0.075 \leq \text{Fe} \leq 0.50 \%$, $0.10 \leq \text{Cu} \leq 0.95 \%$, $0 \leq \text{Mn} \leq 0.85 \%$, $0.35 \leq \text{Mg} \leq 1.40 \%$

For the T6 temper condition, the mechanical properties were within the following intervals: $\leq \text{Yield Strength (YS)} \leq 430 \text{ MPa}$, $205 \leq \text{Ultimate Tensile Strength (UTS)} \leq 483 \text{ MPa}$, $3 \leq \text{Elongation} \leq 15 \%$, $13.14 \leq \text{Modulus of Toughness} \leq 49.5 \text{ MPa}\cdot\%$, $62 \leq \text{Fatigue Strength} \leq 180 \text{ MPa}$ (high-cycle fatigue, $\sim 5 \times 10^8$ cycles).

Qualitative performance indicators for corrosion resistance, weldability, and anodising response were also included.

To enable robust machine learning implementation, the dataset was standardised as follows: the principal alloying elements (Mg and Si) were averaged within their specified composition ranges to represent typical reported values; strengthening elements (Cu and Mn) were assigned their maximum specified values to reflect their upper contribution to mechanical

performance; and impurity elements (Fe) were taken as half of the maximum allowable limit to approximate realistic impurity control levels.

Categorical properties were converted into numerical form (Poor = 0, Fair = 1, Good = 2, Excellent = 3) to facilitate algorithmic processing.

The standardised dataset was then used to apply and compare multiple clustering algorithms in order to group the 6xxx-series alloys based on their elemental composition and T6 condition properties.

K-means Clustering: K-means clustering was implemented on the 6xxx-series dataset as described in **Section 2.2.3.7.1**. As shown in Figure 3.2(a), clustering was performed using all compositional and mechanical property features simultaneously. However, for visualisation purposes, only first two columns of the standardised feature dataset were used as the plot axes. It is emphasised that the clustering itself was conducted in the full 13-dimensional feature space (compositions and properties) and not limited to the two plotted variables.

The K-means algorithm produced an overall silhouette score of 0.31. This value was slightly higher than the score obtained using the PCA + K-means approach (0.20), as presented in Figure 3.2(b) and summarised in Table 3.1.

Although K-means achieved the highest silhouette score among the evaluated algorithms, a value of 0.31 indicates only moderate cluster separation rather than clearly distinct groupings. The cluster boundaries remained partially overlapping in the high-dimensional feature space, suggesting that the alloys do not naturally form strongly separated clusters.

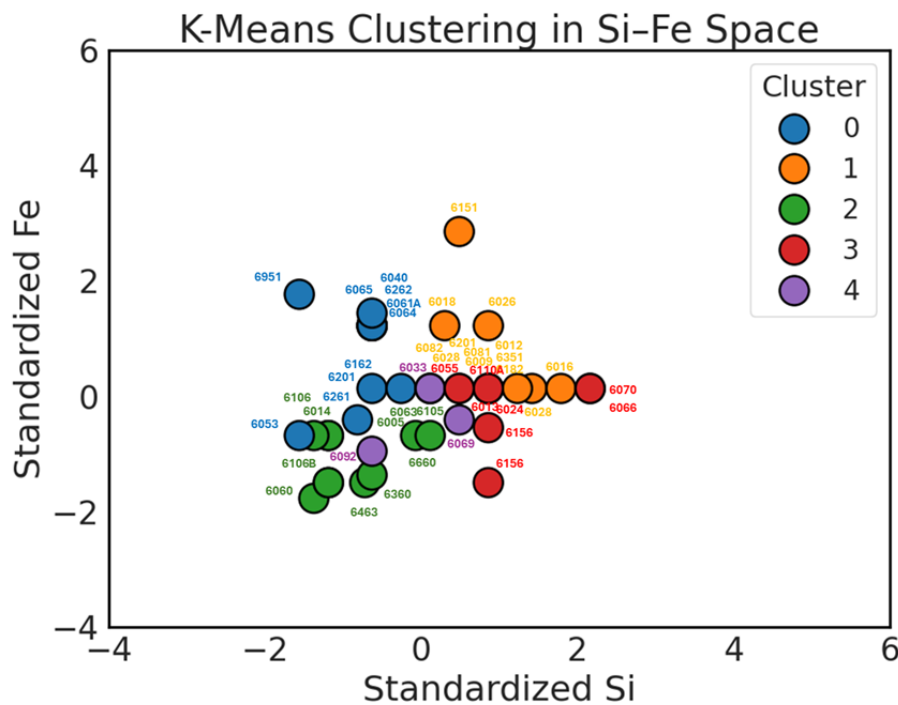
K-means Clustering Combined with PCA: In the PCA + K-means approach, Principal Component Analysis (PCA) was first applied to the standardised multi-feature dataset to reduce dimensionality while preserving the majority of the data variance. In the PCA + K-means approach, Principal Component Analysis (PCA) was first applied to the standardised multi-feature dataset to reduce dimensionality while preserving the majority of the data variance. PCA transforms the original correlated variables into orthogonal principal components that maximise explained variance, while lower-variance components may be discarded during dimensionality reduction. K-means clustering was subsequently performed in the transformed principal component space (as described in **Sections 2.2.3.6** and **2.2.3.7.1**).

This approach differs from standard K-means, which operates directly on the original multi-dimensional feature space without dimensionality reduction. Dimensionality reduction by

PCA can influence cluster geometry and silhouette values: it may enhance apparent separation by removing noise and redundant information or reduce separation by compressing feature variability. PCA modifies cluster geometry by redistributing and compressing variance along principal directions, thereby influencing inter- and intra-cluster distances and the resulting silhouette values. Consequently, the lower score for PCA + K-means does not necessarily indicate inferior clustering performance.

As illustrated in Figure 3.2(a) and (b), visualisation of clustering results in the original high-dimensional space is inherently challenging. K-means alone did not produce visually distinct cluster boundaries when projected for plotting. In contrast, after PCA transformation, five visually separable clusters were observed, despite the lower silhouette score (0.20).

Although the numerical silhouette score for PCA + K-means was slightly lower, this approach yielded more interpretable and visually distinct clusters. Consequently, algorithm selection was not based solely on quantitative metrics. In addition to silhouette scores, domain expertise was applied to manually verify the alloys within each cluster. Detailed spreadsheet-based analyses were conducted to examine alloy compositions and mechanical properties within each group, ensuring that the resulting grouping were metallurgically meaningful and practically relevant see **Appendix 11.3.1.1** for details.



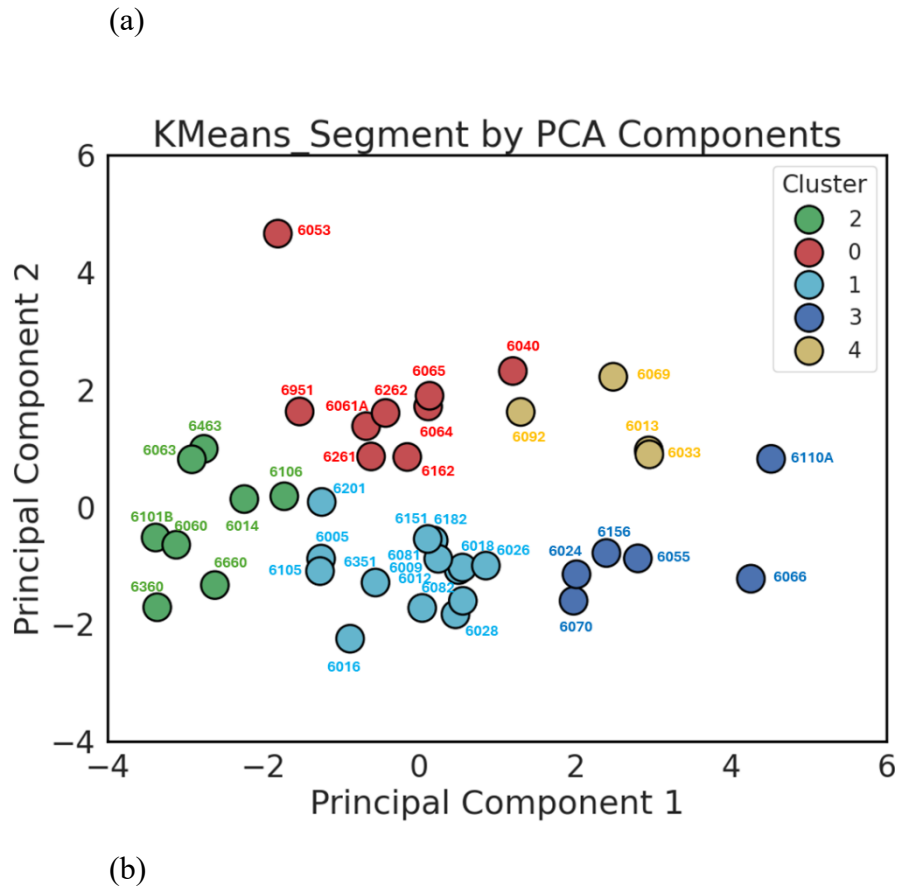


Figure 3.2. Clustering results for 6xxx series alloys based on chemical composition (Si, Fe, Cu, Mn, Mg) and mechanical and qualitative properties at T6 tempering condition (yield strength, UTS, elongation, and modulus of toughness and corrosion, weldability and anodising response). (a) K-means clustering applied to the standardised multi-dimensional dataset; the axes represent scaled feature dimensions, (b) PCA + K-means clustering, where Principal Component 1 and Principal Component 2 are linear combinations of the same composition and mechanical and qualitative variables. The plots therefore represent clustering based on all variables simultaneously

Table 3.1 summarises the overall silhouette scores for all evaluated clustering algorithms, illustrating their relative cluster separation performance.

Table 3.1. Silhouette Scores for all Clustering Algorithm comparison

Algorithm	Silhouette Score (Overall)
K-means	0.31
PCA + K-means	0.20
Hierarchical	0.22

GMM	0.19
Mean Shift	0.25

Note: Silhouette scores (as described in **Section 2.2.3.7.1**) were calculated using all alloy properties simultaneously (yield strength, fatigue strength, modulus of toughness, corrosion resistance, weldability, and anodising response), reflecting multi-dimensional cluster separation. Algorithm selection also incorporated domain knowledge and manual verification of alloy groupings.

The silhouette coefficient ranges from -1 to $+1$, where values close to $+1$ indicate well-separated and compact clusters, values around 0 indicate overlapping clusters, and negative values suggest potential misclassification. In general, silhouette scores above 0.5 represent strong cluster separation, whereas values between 0.25 and 0.5 indicate weak to moderate separation. In the present research, all evaluated algorithms produced silhouette scores below 0.35 , indicating that the alloy dataset does not naturally form strongly distinct clusters. Therefore, the reported values (e.g., 0.31 for K-means and 0.20 for PCA + K-means) reflect moderate multi-dimensional separation rather than clearly defined cluster boundaries. Consequently, clustering decisions were not based solely on silhouette criteria; domain expertise and metallurgical similarity were also considered to ensure that the identified groups were physically meaningful and practically interpretable.

Hierarchical Clustering: The hierarchical clustering showed a Silhouette Score of 0.22 as shown in Table 3.1, which means this clustering algorithm showed moderate separation among the clusters. The property and composition ranges in clusters were not clear, as was found in combined PCA and K-means clustering. Therefore, hierarchical clustering could not beat combined PCA and K-means clustering, as shown in Figure 3.3 (See **Section 2.2.3.7.2** for details).

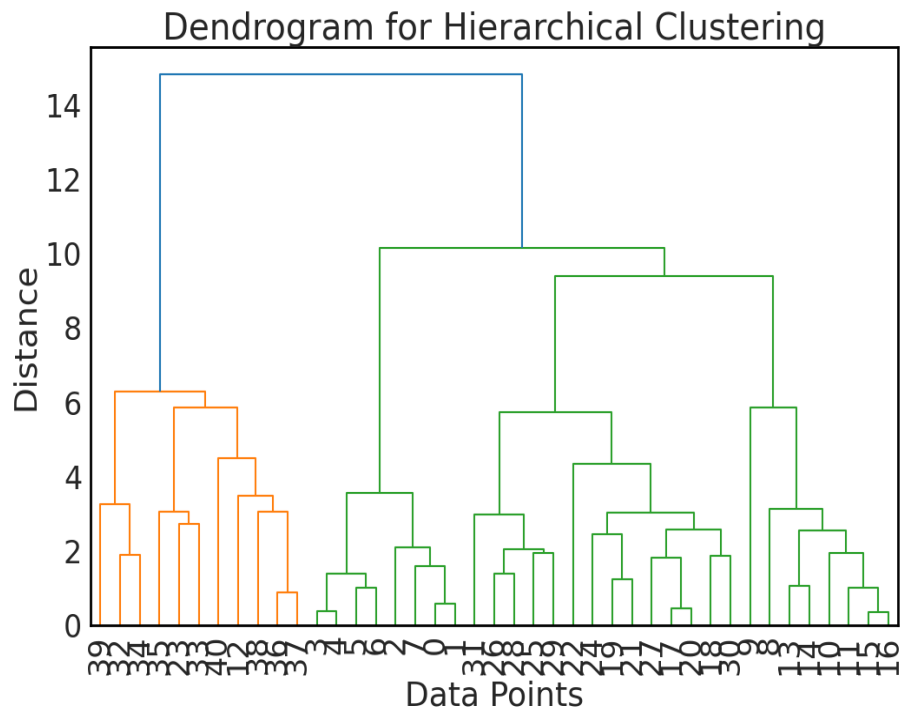


Figure 3.3. Hierarchical clustering: Dendrogram diagram

DBSCAN: DBSCAN (see **Section 2.2.3.7.3**) was unable to form informative clusters due to the sparsity of the dataset (Figure 3.4), making it impossible to evaluate its clustering performance. In contrast, K-means clustering combined with PCA consistently produced stable and visually distinct clusters for the alloy dataset, making it the most suitable method for this analysis.

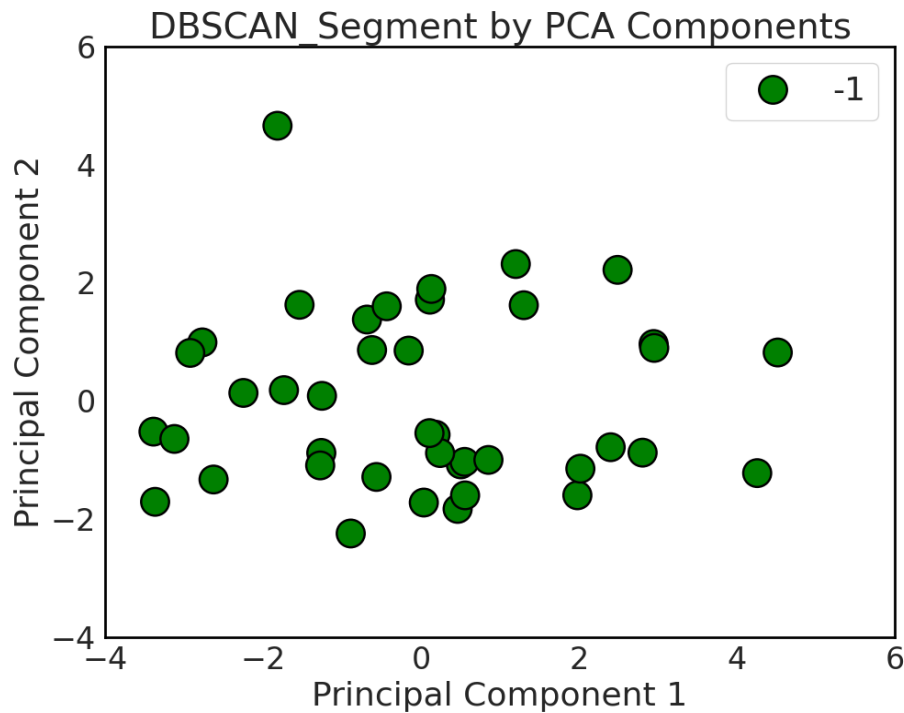
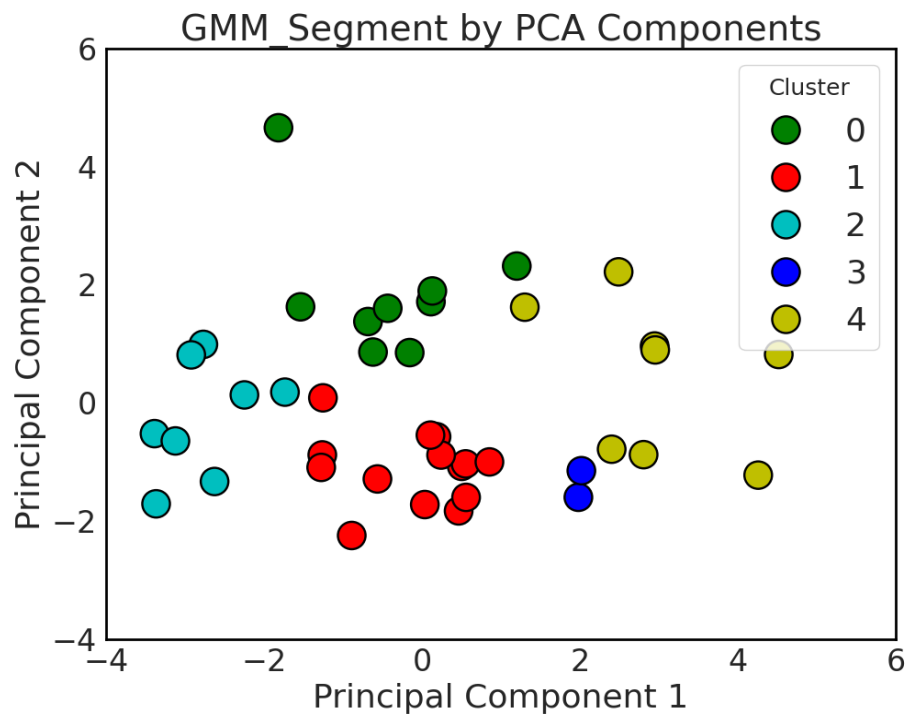
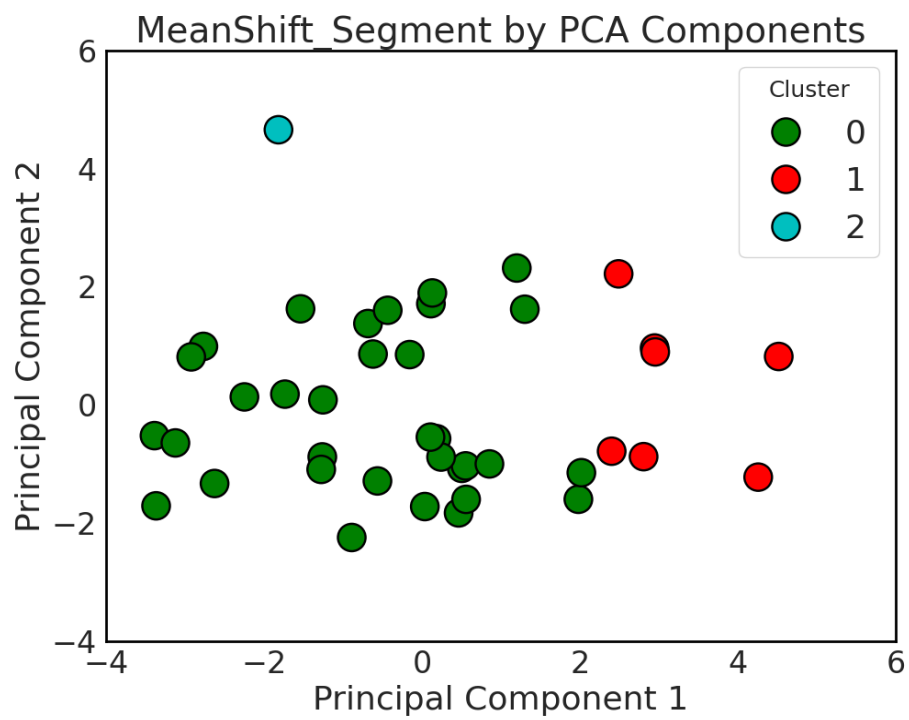


Figure 3.4. DBSCAN clustering

Other algorithms, such as Gaussian Mixture Models (GMM), (see **Section 2.2.3.7.5**) (Figure 3.5(a)) and Mean Shift Clustering, (see **Section 2.2.3.7.6**) (Figure 3.5(b)), yielded silhouette scores of 0.19 and 0.25, respectively, which are comparable to those obtained with PCA combined with K-means clustering as shown in Table 3.1. However, these methods would have required additional computational resources without offering significant improvements for this relatively small dataset. Additionally, GMM allows a single data point to belong to multiple clusters, which may complicate interpretation of the data.



(a)



(b)

Figure 3.5. Clustering of data using (a) Gaussian Mixture Models (GMM), and (b) Mean Shift algorithm

Optimisation Algorithms

After successfully obtaining five distinct clusters from PCA and K-means clustering, we proceeded to analyse the algorithm by selecting a narrower range of aluminium alloys within Cluster 0. To evaluate the performance of the MCDM algorithms in identifying the optimum alloy under tighter boundaries, we selected alloys Alloy A:6261, Alloy B:6162, Alloy C:6262 and Alloy D:6065 for the present analysis. The following multi-criteria decision-making algorithms, were then applied to select the optimal alloy exhibiting the most favourable combination of properties within cluster 0 identified through PCA and K-means clustering:

TOPSIS: For our dataset, TOPSIS performed effectively. It successfully identified the alloys closest to the ideal solution i.e., those with the best combination of properties. The method produced the highest closeness scores for the selection of optimal alloys, as illustrated in the comparison chart for Cluster 0 (Figure 3.6) (see **Section 2.2.3.8.3**).

VIKOR: VIKOR yielded useful results but was less consistent in ranking than TOPSIS (see **Section 2.2.3.8.5**).

AHP: The AHP algorithm proved useful; however, compared to the TOPSIS method, it was less precise in identifying the optimal alloys within the clusters, likely due to its lower consistency and resulting lower scores (see **Section 2.2.3.8.4**).

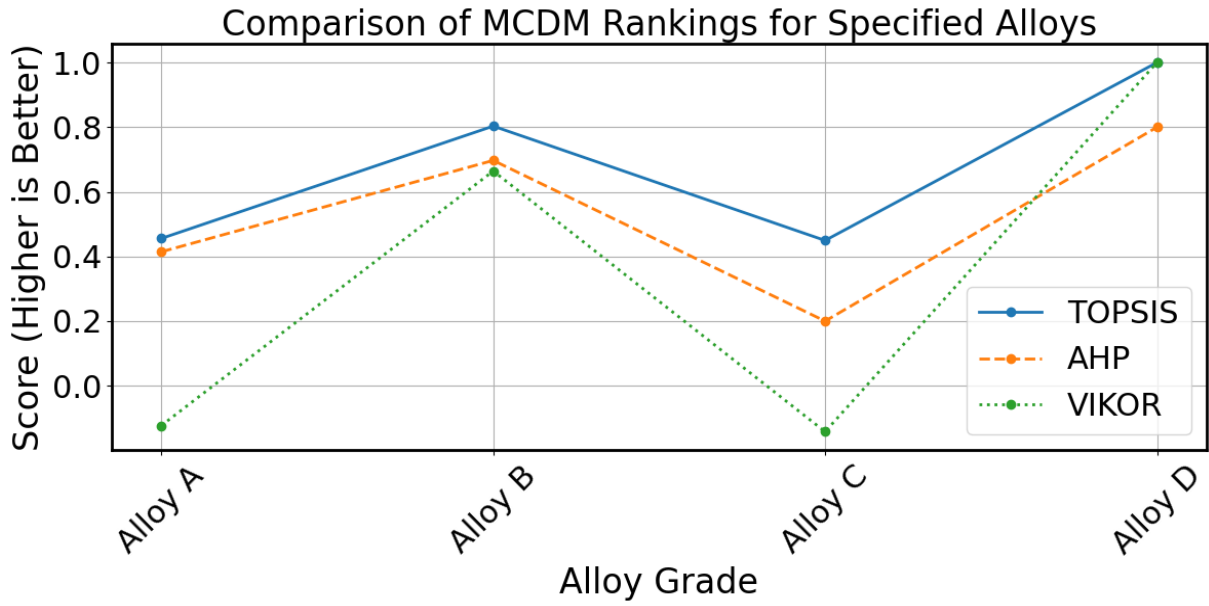


Figure 3.6. MCDMA algorithms comparison

Bayesian Optimisation

Bayesian optimisation was applied to Cluster 2 comprised commercially established 6xxx series aluminium alloys, including 6101, 6106, 6660, 6063, 6360, and 6060. These alloys exhibit comparable compositional ranges and mechanical properties at the T6 tempering condition, making them suitable for focused optimisation within an industrially relevant design space. The optimisation dataset for Cluster 2 was constrained within the observed compositional bounds: Si (0.40–0.60 wt.%), Fe (0.075–0.175 wt.%), Mn (0.05–0.125 wt.%), and Mg (0.35–0.675 wt.%). The corresponding yield strength values ranged from 160 to 220 MPa. These bounds were used to define the parameter search space for Bayesian optimisation. Bayesian optimisation was applied to a multiple linear regression model linking alloy composition (Si, Fe, Mn, Mg) to yield strength within Cluster 2 of the 6xxx series dataset. The regression model was fitted using ordinary least squares (OLS), a statistical method used in linear regression to estimate unknown parameters by determining the best-fitting linear relationship between input variables and the response variable [54]. The resulting regression equation was:

$$YS = 113.03 - 6.23(Si) - 138.37(Fe) + 160.16(Mn) + 144.93(Mg) \quad (24)$$

where 113.03 is the intercept and coefficients represent the fitted weighting factors used in the target (yield strength) prediction during optimisation. The model indicates that Mn and Mg contribute positively to predicted yield strength within the studied range, whereas Fe exhibits a strong negative effect and Si shows a minor negative influence.

The objective function (“target” in Figures 3.7 and 3.8) corresponds to the predicted yield strength calculated from the regression equation. The optimisation search was constrained within the compositional bounds observed in the dataset (Si: 0.40–0.60 wt.%, Fe: 0.075–0.175 wt.%, Mn: 0.05–0.125 wt.%, Mg: 0.35–0.675 wt.%).

This implementation follows exactly the Bayesian optimisation workflow outlined in **Section 2.2.3.8.1**, including surrogate modelling with a Gaussian Process and acquisition-based sampling strategy. At each iteration, a Gaussian Process surrogate model was updated using previously evaluated compositions and their predicted yield strength values. The acquisition function (Equation 12) was then employed to balance exploration of uncertain compositional regions and exploitation of high-performing regions within the defined bounds. The next candidate composition was selected by maximising the acquisition function and subsequently evaluated using the OLS regression model, which served as the objective function. The predicted yield strength obtained was appended to the optimisation history, and the surrogate model was updated prior to the next iteration. This sequential update process generated the optimisation trajectory shown in Figures 3.7 and 3.8.

As shown in Figure 3.7, rapid convergence occurred within the first seven iterations, with predicted yield strength increasing from 162.65 MPa in the initial iteration to values exceeding 217 MPa. The maximum predicted yield strength achieved was 218.01 MPa. The optimisation trajectory demonstrates stable convergence toward an optimum rather than erratic behaviour, indicating a well-defined objective surface within the selected compositional space. The key conclusion from Figure 3.7 is that Bayesian optimisation efficiently identified high-performing compositions within the established compositional limits of standard 6xxx series alloys. Bayesian optimisation is widely used to explore new compositional spaces and can potentially identify novel combinations leading to enhanced mechanical performance and superior properties.

However, it is important to note that the primary objective of the present research was not to discover entirely new alloy systems, but rather to refine existing alloy formulations and identify optimal compositions within established industrial boundaries. While more aggressive exploration of novel compositional combinations could potentially yield additional insights,

such an approach would introduce unnecessary complexity and extend beyond the central goal of optimising known standard alloys.

iter	target	Si	Fe	Mn	Mg
1	162.64770	0.4834044	0.1470324	0.0500085	0.4482580
2	175.94591	0.4293511	0.0842338	0.0639695	0.4623072
3	188.25332	0.4793534	0.1288816	0.0814395	0.5726963
4	178.39693	0.4408904	0.1628117	0.0520540	0.5679019
5	161.65923	0.4834609	0.1308689	0.0605290	0.4143829
6	205.87987	0.4601817	0.075	0.1193838	0.6001200
7	217.06601	0.5509996	0.075	0.125	0.675
8	218.00669	0.4	0.075	0.125	0.675
9	205.49693	0.4799454	0.075	0.05	0.675
10	202.92386	0.6	0.175	0.125	0.675
11	217.55398	0.4726703	0.075	0.125	0.675
12	162.00192	0.4182180	0.1240501	0.0983584	0.3656295
13	216.76075	0.6	0.075	0.125	0.675
14	216.76075	0.6	0.075	0.125	0.675
15	216.76075	0.6	0.075	0.125	0.675
16	216.76075	0.6	0.075	0.125	0.675
17	216.76075	0.6	0.075	0.125	0.675
18	217.85431	0.4244613	0.075	0.125	0.675
19	217.85431	0.4244613	0.075	0.125	0.675
20	217.85431	0.4244613	0.075	0.125	0.675

Figure 3.7. Convergence of Bayesian optimisation showing predicted yield strength (MPa) versus iteration number for Cluster 2 (6xxx series) alloys. The objective function corresponds to yield strength predicted by the OLS-based regression model within the defined compositional bounds.

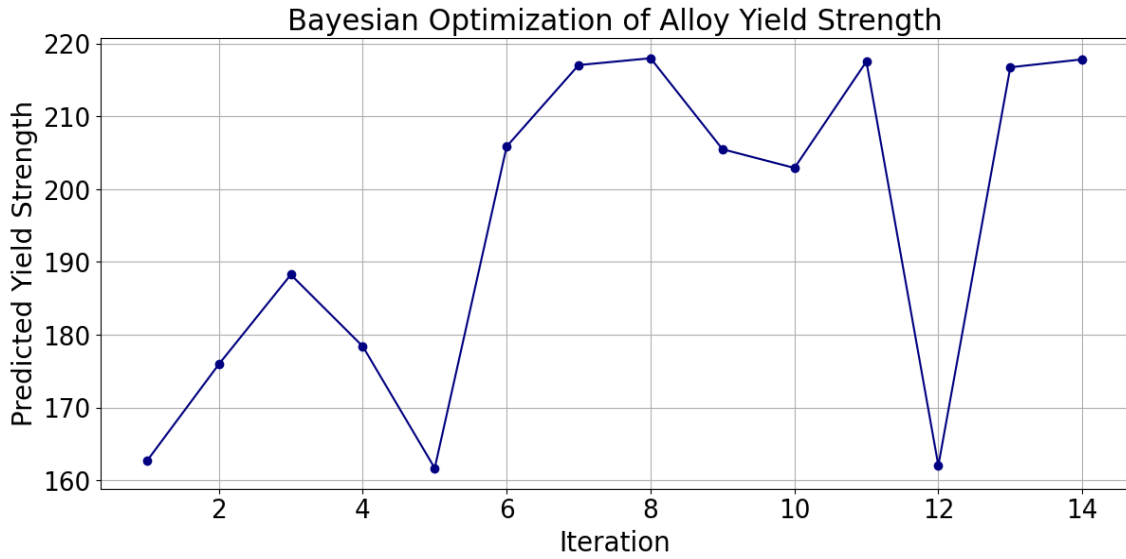


Figure 3.8. Iterative Bayesian optimisation trajectory showing predicted yield strength (MPa) versus iteration number for Cluster 2, illustrating sequential model updates and convergence toward the optimal composition.

Artificial Neural Network (ANN)

A small artificial neural network (ANN) was applied to five clusters derived from the alloy dataset using principal component analysis (PCA) followed by K-means clustering. The clustering procedure grouped alloys based on their compositional variables (Si, Fe, Cu, Mn, Mg) and mechanical properties. The dataset used for ANN training was obtained by extracting the data corresponding to each individual cluster. For each cluster, the compositional variables served as input features, while the yield strength, modulus of toughness, and fatigue strength were used as target outputs.

Due to the limited number of alloys per cluster, a separate ANN model was trained for each cluster using all available samples within that cluster. No additional synthetic data were generated. Prior to training, input features were standardised using z-score normalisation (StandardScaler), and the target variables were scaled to the range [0,1] using min-max normalisation to ensure numerical stability and balanced multi-objective optimisation.

The network architecture was deliberately kept small to reduce overfitting risk, consisting of a single hidden layer with six neurons and ReLU activation, followed by a linear output layer predicting the three target properties simultaneously. Predicted targets were normalised and combined into a composite score to identify the highest-performing optimum alloy in each

cluster across all three properties. This multi-objective approach enabled balanced optimisation, avoiding trade-offs associated with maximising a single property.

Figure 3.9 shows the training loss per epoch for each cluster, illustrating the convergence behaviour of the ANN during training. In this context, the **training loss** corresponds to the mean squared error (MSE) between the predicted and actual target values across all training samples. The MSE is calculated as the average of the squared differences between predicted and measured values for yield strength, modulus of toughness, and fatigue strength [139]. An **epoch** represents one complete pass of the entire training dataset through the neural network, during which the model weights are updated using backpropagation and the Adam optimiser to minimise the loss function [47].

Overall predictive performance was evaluated using the coefficient of determination (R^2), which provides a scale-independent measure of accuracy across all target properties as described in **Section 2.2.3.4**. On average, the ANN achieved R^2 values of 0.89 for yield strength, 0.60 for modulus of toughness, and 0.79 for fatigue strength. These results indicate that the model captured meaningful relationships between alloy composition and mechanical properties despite the limited dataset size.

Although the small dataset limited the robustness of predictions and increased the risk of overfitting, the ANN demonstrated the ability to extract relevant compositional–property patterns, providing a viable framework for exploratory multi-property alloy optimisation that could be further strengthened in future studies using larger datasets (see **Section 2.2.3.8.2** for methodological details).

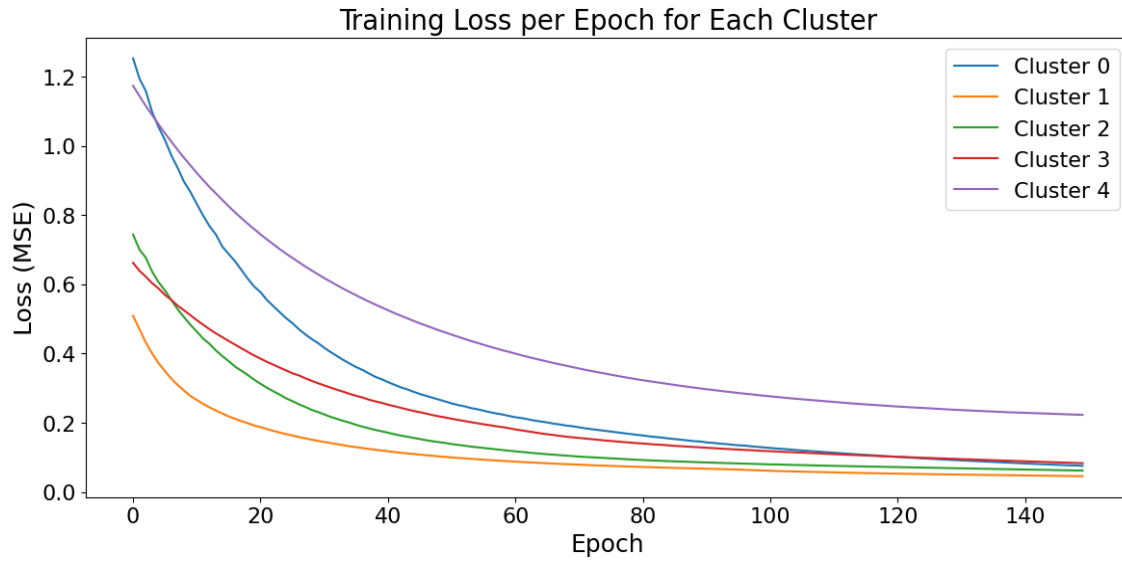


Figure 3.9. Training loss (mean squared error, MSE) versus epoch number for the artificial neural network (ANN) models trained separately on each PCA–K-means-derived alloy cluster. The curves show the progressive reduction in MSE during backpropagation using the Adam optimiser, indicating stable convergence behaviour across all five clusters despite the limited dataset size.

Final Methodology Selection: After a comprehensive evaluation of all approaches, the selected methodology combined principal component analysis (PCA) for dimensionality reduction with K-means clustering to group alloys into distinct clusters. Subsequently, the TOPSIS optimisation algorithm was applied within these clusters to identify the best-performing optimum alloys, while a designed recycling algorithm facilitated alloy mixing within the clusters to again generate the optimum alloys resulted by optimisation algorithm. The complete processed dataset generated from all standardisation steps, PCA transformations, clustering assignments, optimisation scores, calculated features, and model outputs is provided in **Appendix 11.3.1.2**. A detailed step-by-step reproducibility workflow outlining data preparation, dimensionality reduction, clustering implementation, optimisation procedures, and model training is documented in **Appendix 11.3.1.3**, enabling full replication of all analyses, figures, and tables presented in this chapter.

3.2 Design Framework

A systematic design framework for the optimisation of the aluminium alloy series is illustrated in Figure 3.10. It consists of the following key stages: Data Collection & Pre-processing → Combined PCA and K-means Clustering → Sub-clustering → Optimisation Algorithm →

Alloy Mixing: Recycling Algorithm. Additional steps, including Explainable AI (LIME), literature-based metallurgical reasoning, Thermo-Calc single-point equilibrium analysis for phase formation, and validation of the selected optimum alloys using unseen data, were also conducted to support the findings of the design framework. Initially, dataset collection was done by collecting chemical compositions, mechanical properties, weldability and corrosion properties, and thermal and electrical properties at various tempering and hardening conditions. Following data collection, clustering was performed. Once distinct clusters with similar property ranges, supported by LIME analysis and metallurgical reasoning, the dataset was further sub-clustered to create more compact groups of compositions and properties. An optimisation algorithm was then employed to identify the alloys with the best combination of properties within each sub-cluster. Next, the recycling algorithm was applied to perform alloy mixing within sub-clusters, yielding optimum alloy compositions and properties as determined by the optimisation algorithm. This framework enabled the design and selection of the most promising aluminium alloys across multiple series, resulting in alloys that not only met but exceeded the desired performance criteria.

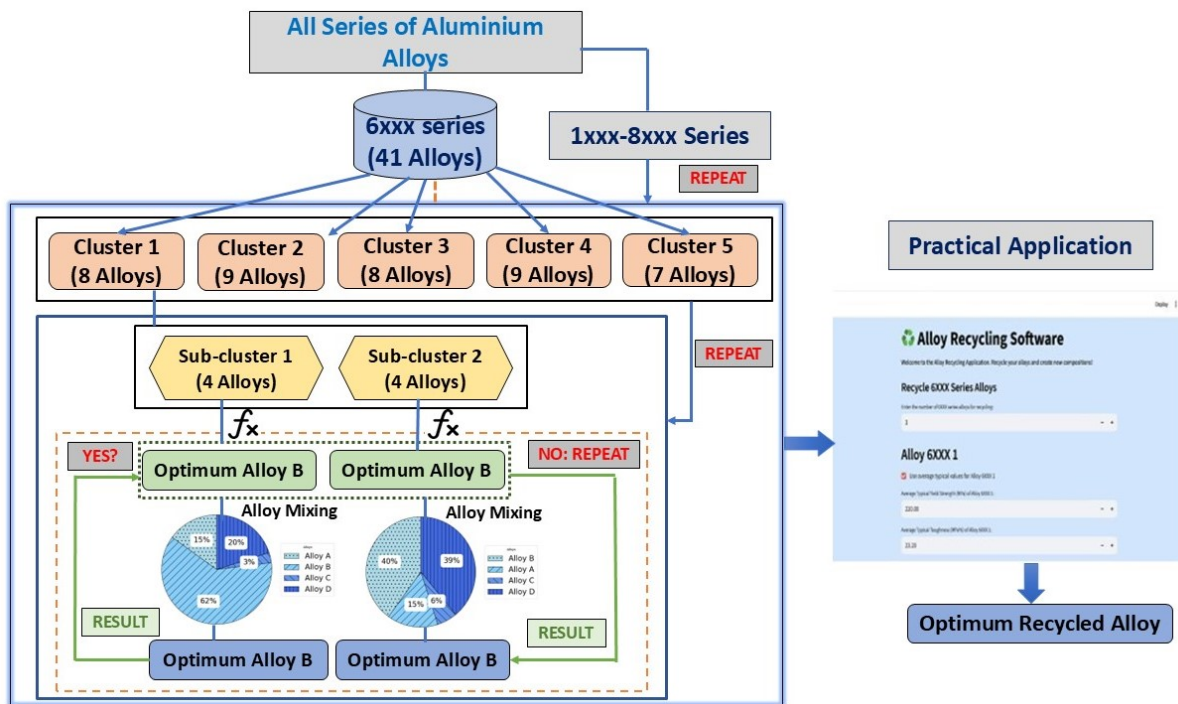


Figure 3.10. Design framework

3.3 Data Collection and Pre-processing

3.3.1 Data Collection

In this study, the focus was limited to wrought aluminium alloys. For the two main types of wrought aluminium alloy series, non-age-hardenable and age-hardenable, data were systematically collected from peer-reviewed journal articles, established textbooks, and other reputable literature sources. For non-age-hardenable alloys, emphasis was placed on the 1xxx, 3xxx, 5xxx, and selected 8xxx series alloys. The properties considered under standard hardening conditions, such as H14, H32, as defined in **Section 2.1.5**, encompassed a broad spectrum of mechanical, electrical, thermal, corrosion, and weldability characteristics.

Data for age-hardenable alloy series, specifically the 2xxx, 6xxx, and 7xxx series, were collected at various stages of artificial ageing, including T5, T6, and T7 temper conditions, as defined in **Section 2.1.5**. The collected data were organised into a structured Comma-Separated Values (CSV) format [140], which was subsequently analysed and modelled.

All the ML deployments were done inside the Anaconda Jupyter Notebook environment and Google Colaboratory [141], [142]. Data visualisation and graph plotting were done using the Matplotlib and Seaborn libraries. This ensured that the workflow of data exploration and model development was smooth and efficient.

3.3.2 Data Pre-processing

Following data collection, extensive preprocessing was carried out to prepare the dataset for machine learning analysis. Key steps included data cleaning, feature engineering, data transformation, and standardisation.

Data Cleaning

Data cleaning was necessary to render the dataset complete and credible. For both age-hardenable and non-age-hardenable aluminium alloys, missing values were handled using appropriate imputation methods. For numerical features such as YS (yield strength), UTS (ultimate tensile strength), elongation (EI), fatigue strength, electrical conductivity, and thermal conductivity, missing values were imputed using the Random Forest Regression algorithm (see **Section 2.2.3.4**). For the categorical variables such as corrosion resistance, weldability, and anodising response, the Random Forest Classification model was applied to fill in missing data (see

Section 2.2.3.5). With this cleaning process, both numerical and categorical variables were effectively handled, allowing machine learning algorithms to be applied smoothly.

Feature Engineering: A Combining Function

To enhance the model's performance and reduce dataset dimensionality, a combining function was applied to calculate the modulus of toughness. The modulus of toughness is defined as the total area under the stress–strain curve and represents the amount of energy per unit volume that a material can absorb before fracture. It reflects the combined effect of strength and ductility and is therefore a useful indicator of a material's ability to withstand plastic deformation without failure [143].

For ductile materials, the modulus of toughness can be approximated mathematically. In this study, it was calculated using the expression: $\text{Modulus of Toughness} = 0.5 \times (\text{UTS} + \text{YS}) \times \text{EI}$ [143], where UTS is the ultimate tensile strength, YS is the yield strength, and EI is the elongation to fracture. This formulation follows the approximation proposed in the ASM Handbook. By combining these mechanical properties into a single engineered feature, the dimensionality of the dataset was reduced while retaining the underlying physical information related to both strength and ductility. This improved model efficiency without loss of interpretability.

Data Transformation and Integration

In contrast to previous research that largely employed numeric attributes, our data set contained both numeric and categorical attributes.

The important categorical features were ratings of weldability, corrosion resistance, and anodising response, which were converted to numerical values. After transformation, these categorical features were combined with the numerical features of chemical composition and other properties (mechanical, thermal, and electrical properties depending on the alloy series) using a joining function. This integration created a comprehensive dataset for modelling, ensuring that all relevant features were included in the machine learning process.

Data Standardisation

Following dataset transformation, data standardisation was performed to bring all features onto a comparable scale, addressing differences in their magnitudes. This process included normalising features to ensure that no single attribute disproportionately influenced the model due to

its larger values. Standardisation was particularly important for algorithms sensitive to input scale, as it improved model performance and ensured balanced contributions from each feature, thereby enhancing the effectiveness of the machine learning models [144].

3.4 Combined Approach of PCA and K-means Clustering

K-means clustering and principal component analysis (PCA) were employed to group aluminium alloy datasets into clusters containing alloys with similar compositions and property Ranges. The hybrid PCA and K-means clustering approach was implemented through the following steps [145]:

Step 1: Perform PCA – PCA was applied to the standardised data to reduce its dimensionality by identifying the principal components.

Step 2: Select Principal Components – The optimal number of principal components to retain was determined from the PCA variance plot.

Step 3: Data Transformation – Once the principal components were chosen, the initial data were mapped onto a lower dimensionality space.

Step 4: Choose the Number of Clusters – The elbow method was used to determine the optimal number of clusters.

Step 5: K-means Clustering – K-means clustering was performed to place each data point to its closest cluster centroid.

Step 6: Validation of Results – Cluster quality was assessed using the silhouette score method.

Step 7: Interpret Results – Clustering results were analysed in terms of cluster characteristics, including compositions and properties that differentiate the clusters, as well as any other relevant dataset information.

Step 8: Visualise Results – The clustering outcomes were visualised to facilitate interpretation and evaluation of the alloy groupings.

3.5 Sub-clustering of Clusters

Sub-clustering analysis was performed using a combined PCA and K-means clustering approach. Following the primary clustering, each cluster was further subdivided to achieve more refined groupings. This additional step allowed for the identification of optimum alloys within these sub-clusters, enhancing the efficiency and precision of the subsequent optimisation algorithm.

We also employed a mathematical model together with PCA and K-means clustering for further consistent cluster partition.

The model determines the number of sub-clusters based on the size of each primary cluster. Larger clusters are subdivided into more sub-clusters to achieve efficient data structuring, while smaller clusters are preserved to avoid over fragmentation.

Mathematical Model:

Let C_k be the cluster containing N_k data points. The number of sub-clusters, S , is determined using a piecewise function as follows:

$$S_k(N_k) = \{ \begin{array}{l} 1 \text{ if } N_k \leq 3, \\ 2 \text{ if } 4 \leq N_k \leq 10, \\ 3 \text{ if } 11 \leq N_k \leq 20, \\ 4 \text{ if } 21 \leq N_k \leq 30, \\ 5 \text{ if } N_k > 30, \\ \end{array} \}$$

Explanation:

- Small Clusters ($N_k \leq 3$): If there are fewer than or equal to 3 alloys, no meaningful sub-clustering can be performed. The function assigns a single sub-cluster to prevent over-segmentation.
- Medium-Sized Clusters ($4 \leq N_k \leq 10$): Clusters with 4–10 alloys can be split into 2 sub-clusters.
- Large Clusters ($11 \leq N_k \leq 20$): Clusters with 11–20 alloys are split into 3 sub-clusters.
- Large Clusters ($21 \leq N_k \leq 30$): For clusters containing 21–30 alloys, a higher number of sub-clusters (4) was assigned.

- **Very Large Clusters ($31 \leq N_k \leq 40$):** Clusters containing 31–40 alloys are assigned 5 sub-clusters to ensure effective data separation. For clusters exceeding this range, additional sub-cluster allocations can be defined as needed to maintain granularity and efficient data structuring.

This finer granularity helped in selecting an optimum alloy within a tight range of similar alloys. This also reduced the variability in the range of properties.

3.6 Optimisation Algorithm

Following sub-clustering, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), a multi-criteria decision making (MCDM) algorithm, was applied. TOPSIS optimisation algorithm enabled systematic decision-making by simultaneously evaluating multiple property criteria, thereby identifying the alloys with the most optimal combination of properties within each sub-cluster. The procedure of TOPSIS involves the following steps (see **Section 2.2.3.8.3**) [66]:

Step 1: Convert Categorical Variables to Numerical Values – Qualitative ratings for corrosion resistance, anodising response, and weldability were converted into numerical values to ensure consistency in calculations.

Step 2: Construct Properties Matrix – A matrix was created comprising the properties of alloys within each sub-cluster.

Step 3: Normalise the Properties Matrix – All alloy properties within the sub-cluster were normalised using Equation (14).

Step 4: Determine the Ideal-Best Solution, T^+ , and Ideal-Worst Solution, T^-

Ideal-Best Solution (T^+): The highest values of all alloy characteristics (for benefit criteria, maximum; for cost criteria, minimum).

Ideal-Worst solution (T^-): The lowest values of all alloy characteristics (for benefit criteria, minimum; for cost criteria, maximum).

Step 5: Calculate the distance of each alloy from the ideal-best and ideal-worst solutions using Equation (15) and Equation (16).

Step 6: Closeness Coefficient, i.e., TOPSIS Score, C_i – The higher the value of C_i as calculated using Equation (17), the higher the performance of the alloy in terms of all properties.

Step 7: Ranking of All Alloys Based on TOPSIS Score (C_i) – All alloys were ranked according to their TOPSIS scores (C_i) to identify the optimum alloy.

Step 8: Optimum Alloy Prediction – The optimum alloy is identified by the highest TOPSIS score.

This approach ensured the selection of an optimum alloy within each sub-cluster, characterised by a superior combination of properties that encompassed the broader property range of the alloys in that sub-cluster, thereby effectively narrowing the overall solution space.

3.7 Alloy Mixing in Clusters: Recycling Algorithm

After applying the optimisation algorithm and identifying the optimum alloys, defined as those closest to the ideal best solution i.e., the solution consists of the best combination of properties, the subsequent step involved determining the appropriate mixing ratios of alloys within each sub-cluster to achieve the desired optimum properties and compositions. This was accomplished through a designed recycling algorithm, which followed the steps outlined below.

Step 1: Selection of the Optimum Alloy – The optimum alloy was identified using the closeness score obtained from the TOPSIS optimisation algorithm. Specifically, the alloy with the maximum closeness score was selected, representing the properties nearest to the ideal best solution.

Step 2: At Least 50% Mixing Ratio Assigned to the Optimum Alloy – Once the optimum alloy was identified, a minimum of 50% of the total mixing ratio was allocated to it. This ensured that the final alloy design retained the superior characteristics of the optimum alloy while maintaining flexibility to incorporate the benefits of other alloys within the sub-cluster.

Step 3: Distribution of Remaining Mixing Ratios – The remaining 50% of the mixing ratio was proportionally distributed among the other alloys within the sub-cluster, based on their closeness scores. Alloys with higher closeness scores were allocated a larger proportion of the mixture, reflecting their greater similarity to the ideal best solution. The mixing ratio for each alloy i was determined using the formula:

$$\text{Mixing ratio of Alloy } i = \frac{(\text{Closeness score of alloy } i)}{\sum(\text{Closeness score of all alloys})} \times 0.5 \quad (25)$$

In this manner, alloys with superior characteristics contribute proportionally more to the final mixture.

Step 4: Normalisation of Mixing Ratios – The calculated mixing ratios will be normalised to ensure their total equals 100% (or 1). This prevents any single alloy from having too high or too low a percentage in the final mixture. The normalised mixing ratio for each alloy is given by:

$$\text{Normalised Mixing Ratio of Alloy } i = \frac{\text{Mixing Ratio of Alloy } i}{\sum(\text{Mixing Ratios of all alloys})} \quad (26)$$

Step 5: Calculate the Composition and Properties of the Resulting Mixed Alloy – After determining the mixing ratios, the composition of the resulting mixed alloy was calculated. This was achieved by multiplying each alloy's mixing ratio by its respective composition values and summing the contributions across all alloys in the mixture. The resulting composition for each alloy can be expressed as:

$$\text{Resulting composition} = \sum(\text{Mixing Ratio of Alloy } i \times \text{composition of Alloy } i) \quad (27)$$

The same procedure was applied to each composition and property, including Si, Fe, Cu, Mn, Mg, yield strength, modulus of toughness, fatigue strength, corrosion resistance, weldability, anodising response, and other relevant attributes depending on the alloy series.

Step 6: Evaluate Final Composition and Properties of the Mixed Alloy – The composition and properties of the mixed alloy were assessed by comparing the resulting values with those of the original standard alloys and the optimum alloy determined via optimisation algorithm. This evaluation ensured that the mixed alloy's composition closely approximates the optimal values, thereby maximising the desired material characteristics.

Step 7: Validation and Optimisation – Finally, the mixed alloy composition was validated either through experimental methods (if available) or by assessing performance against other

material property standards, such as an unseen dataset. In some cases, further refinement of the mixing ratios may be necessary to meet specific engineering or industrial requirements.

3.8 LIME Algorithm

To understand the rationale behind clustering and its metallurgical significance, it is important to analyse how alloy compositions influence the properties within each cluster. The combined PCA and K-means clustering approach was applied to visualise distinct clusters of alloys with similar compositions and properties. Subsequently, the LIME algorithm, as described in **Section 2.2.3.9**, was employed to interpret the influence of compositions on properties within each cluster. LIME focuses on training local surrogate models to explain the predictions for individual clusters generated by the PCA and K-means clustering algorithm. By processing input data samples, i.e., alloy compositions, LIME ranks the positive and negative impacts of each composition on the properties. This approach enables a clearer understanding of the clustering results and provides insights into how variations in composition drive property differences within the “black box” clustering model [68].

CHAPTER 4: RESULTS AND DISCUSSION: HYBRIDISING ALL ALUMINIUM ALLOY SERIES: VALIDATING CLUSTERING ALGORITHMS

In this chapter, we discussed the hybridisation of various aluminium alloy series collectively, prior to performing individual alloy series optimisations. This provided a broad understanding of how different alloy series can be combined, resulting in clusters of alloys with similar ranges of properties and compositions, which will serve as a basis for mixed series recycling. The process begins by collecting the dataset on various age-hardenable and non-age-hardenable alloy series, then performing the combined approach of PCA and K-means clustering for their categorisation into various alloy clusters possessing similar ranges of properties and composition's.

4.1 Data Collection

4.1.1 2xxx, 6xxx, 7xxx Age-Hardenable Alloy Series

The dataset of age-hardenable aluminium alloy series includes 211 alloys, namely, 89 alloys of 2xxx, 44 alloys of 6xxx, and 73 alloys of the 7xxx series with and without Cu and 5 alloys of 8xxx series, collected at the commonly available T6 tempering condition from the Encyclopaedia of Aluminium and Its Alloys [137], the ASM Speciality Handbook [1], the MakeItFrom website [138], and various literature sources. The complete raw dataset for the age-hardenable alloys is provided in **Appendix 11.3.3.1**, while the fully processed and standardised dataset used for clustering analysis is presented in **Appendix 11.3.3.2**. A detailed stepwise reproducibility workflow describing data preparation, preprocessing, PCA implementation, clustering, and validation procedures is documented in **Appendix 11.3.2.3**. This dataset comprises a wide range of chemical compositions, mechanical properties, and a variety of corrosion, weldability, electrical, and thermal properties.

It should be noted that the 44 reported 6xxx alloys correspond to the initially compiled dataset and include closely related sub-grades (e.g., 6262 and 6262A). In this section, all available grades were retained to provide a comprehensive overview of the 6xxx age-hardenable alloy series. However, during the data cleaning and preprocessing stage (**described in Section 5.1**), closely related sub-grades which are not widely used in practical applications, were consolidated to avoid duplication. Alloys with incomplete property data, particularly missing

anodising performance information, were also excluded to ensure consistency in the modelling variables. Consequently, although 44 alloys are discussed in this section for descriptive completeness, the final dataset used for modelling and optimisation in **Chapter 5** comprised 41 distinct 6xxx alloy grades.

The chemical compositions of these alloys are as follows: $0.05 \leq \text{Si} \leq 1.35 \%$, $0.025 \leq \text{Fe} \leq 0.75 \%$, $0 \leq \text{Cu} \leq 6.3 \%$, $0 \leq \text{Mn} \leq 1.3 \%$, $0 \leq \text{Mg} \leq 4.05 \%$, and $0 \leq \text{Zn} \leq 11.5 \%$, and trace elements such as $0 \leq \text{Cr} \leq 0.175 \%$, $0 \leq \text{Ti} \leq 0.15 \%$, $0 \leq \text{Ag} \leq 0.35 \%$, $0 \leq \text{Li} \leq 3.8 \%$, $0 \leq \text{Ni} \leq 2 \%$, and $0 \leq \text{Zr} \leq 0.15 \%$.

The mechanical properties span $121 \leq \text{yield strength} \leq 730 \text{ MPa}$, $180 \leq \text{fatigue strength} \leq 210 \text{ MPa}$, measured under high-cycle fatigue conditions ($\sim 5 \times 10^8$ cycles, corresponding to high fatigue life), and $11.9 \leq \text{modulus of toughness} \leq 13.14 \text{ MPa } \%$. Moreover, weldability ranges from poor to excellent, corrosion resistance from fair to excellent, and thermal conductivity between 20% IACS and 60% IACS, while electrical conductivity ranges from 120 W/m·K to 220 W/m·K.

These properties highlight the versatility of the 2xxx, 6xxx, 7xxx, and selected 8xxx series alloys, making them particularly suitable for aerospace, automotive, and other high-performance engineering applications.

4.1.2 1xxx, 3xxx, 5xxx and 8xxx Non-Age-Hardenable Alloy Series

The datasets used in this study comprise 152 aluminium alloys from the non-age-hardenable series, specifically 33 alloys of 1xxx, 36 alloys of 3xxx, 76 alloys of 5xxx, and 7 selected alloys of the 8xxx series. These data were collected from the Makeitlearn website [138], ASM handbooks [1], and literature sources. The complete raw dataset for the non-age-hardenable alloys is provided in **Appendix 11.3.2.1**, while the processed and standardised dataset applied in the clustering analysis is presented in **Appendix 11.3.2.2**. The full reproducibility documentation for the preprocessing, PCA transformation, and clustering implementation is included in **Appendix 11.3.2.3**. All datasets for these alloys were collected in the commonly available H14 and H32 hardened state, consisting of chemical composition, mechanical properties, corrosion, technological, electrical and thermal properties. The compositions of the complete non-age-hardenable series dataset are as follows: $0 \leq \text{Si} \leq 1.8 \%$, $0 \leq \text{Fe} \leq 0.6 \%$, $0 \leq \text{Cu} \leq 0.9 \%$, $0 \leq \text{Mn} \leq 1.5 \%$, and $0 \leq \text{Mg} \leq 8 \%$. Trace elements are present within the following ranges: $0 \leq \text{Cr} \leq 0.2 \%$, $0 \leq \text{Zn} \leq 2.8 \%$, $0 \leq \text{Ti} \leq 0.175 \%$, $0 \leq \text{Ga} \leq 0.05 \%$, $0 \leq \text{V} \leq 0.05 \%$, and $0 \leq \text{Zr} \leq 0.15 \%$.

The mechanical properties of the alloys span $23 \leq \text{yield strength} \leq 410$ MPa, $23 \leq \text{fatigue strength} \leq 210$ MPa, measured under high-cycle fatigue conditions ($\sim 5 \times 10^8$ cycles, corresponding to high fatigue life), and $3.23 \leq \text{modulus of toughness} \leq 66.7$ MPa %. These ranges demonstrate the variability in compositions and mechanical performance across the dataset.

Besides mechanical properties, the alloys exhibit a wide range of corrosion and weldability characteristics. The weldability was rated from fair to excellent, while corrosion resistance was rated from good to excellent. Their electrical conductivity varies between 28 % and 66 % IACS, while thermal conductivity ranges from 110 W/m·K to 243 W/m·K.

4.2 Data Cleaning and Pre-processing

4.2.1 Predicting Missing Values in the Dataset

The data cleaning process for filling missing values was carried out for both age-hardenable and non-age-hardenable alloy series using the Random Forest Regressor and Random Forest Classifier, as described in **Sections 2.2.3.4** and **2.2.3.5**, to predict missing values in numerical and categorical columns, respectively. For the numerical features such as YS, UTS, fatigue strength, elongation, modulus of toughness, electrical conductivity, and thermal conductivity, we used Random Forest Regressor. The model was trained on the available data to predict missing numerical values based on the relationships between chemical compositions and their corresponding mechanical and physical properties.

For the categorical features, including corrosion resistance and weldability, which consist of discrete ratings, a Random Forest Classifier was employed. The model was trained on the existing data to predict the missing values in these categorical columns.

4.2.2 Data Pre-processing

Data pre-processing was performed for both age-hardenable and non-age-hardenable aluminum alloy series datasets. The categorical features, including corrosion resistance and weldability, were rated as 'poor', 'fair', 'good', or 'excellent'. To facilitate machine learning analysis, these ratings were converted into numerical values using label encoding, with the following representations: poor = 0, fair = 1, good = 2, and excellent = 3.

After this transformation, we concatenated the numerical features with label-encoded categorical variables so that we have both the numerical and categorical data in a uniform numerical form. Following this, data standardisation was applied using the StandardScaler method as described in Equation 6. Due to the significant variation in feature scales, for example, Fe values ranging from 0.0% to 0.6%, while YS values exceed 100 MPa, the dataset exhibited considerable disparities in range. Standardisation addressed this issue by scaling all features to a common scale with a mean of 0 and a standard deviation of 1, ensuring consistency across variables. This is important for most machine learning algorithms to function correctly, since without standardising, features with large values will end up dominating the learning process.

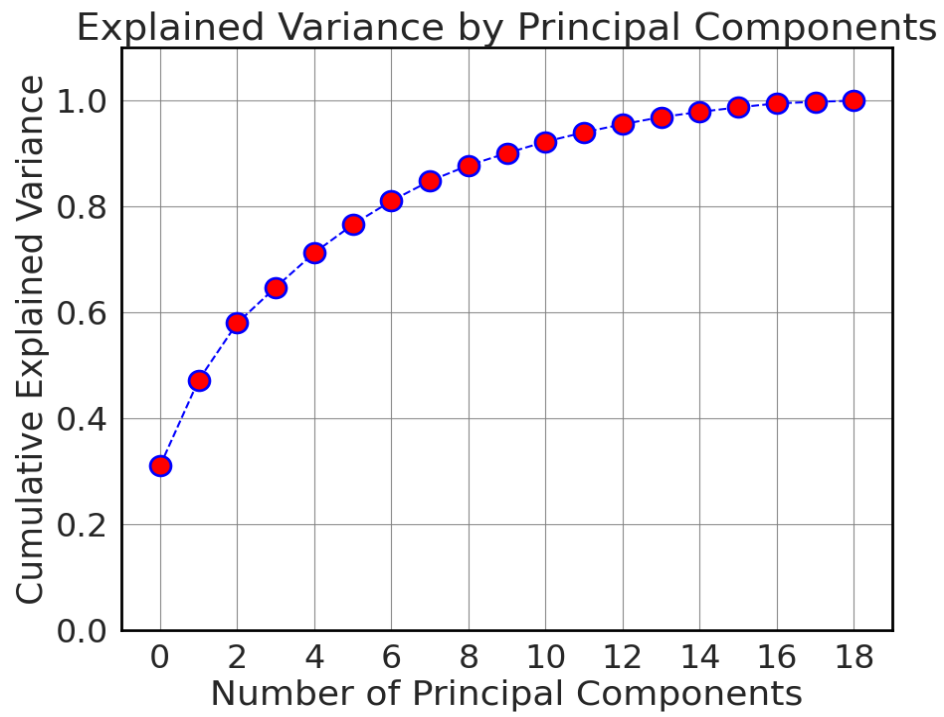
4.3 Combined Approach of PCA and K-means Clustering

4.3.1 Clustering of all Age-Hardenable Alloy Series

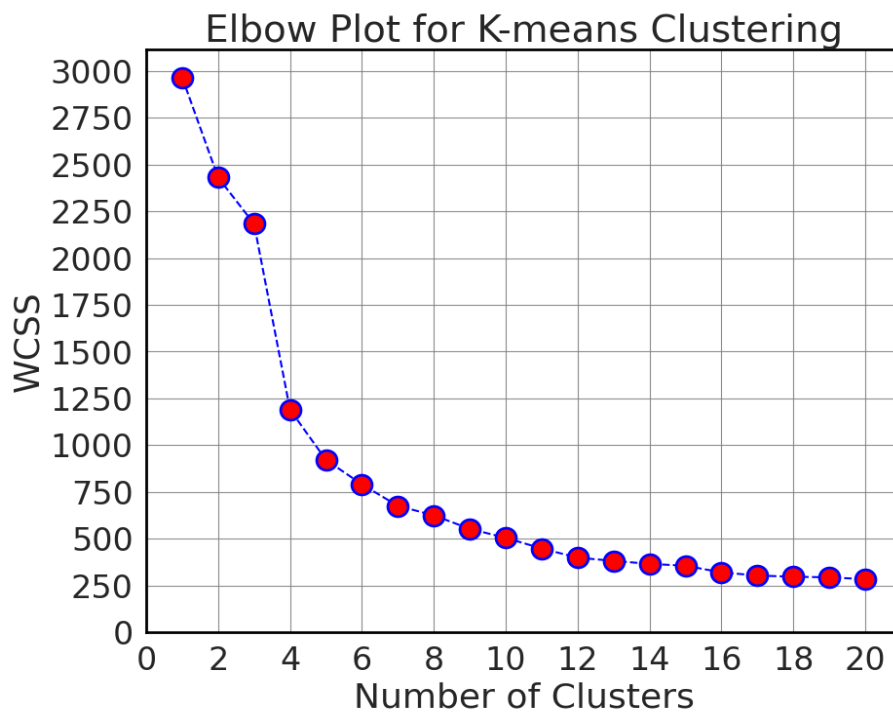
We applied a combined PCA and K-means clustering approach to the standardised age-hardenable alloy series dataset to identify clusters containing similar range of properties and compositions, as described in **Sections 2.2.3.6 (PCA), 2.2.3.7.1 (K-means), and 3.4 (combined PCA & K-means clustering)**.

We began by constructing a PCA variance plot (Figure 4.1(a)), which depicts the cumulative explained variance against the number of principal components. This plot helps to determine the number of principal components that are required to explain a large part of the total variance within the data. Every principal component is a different source of variability in the original dataset. As can be seen in the plot, there is a direct relationship between the number of components and the variance they explain.

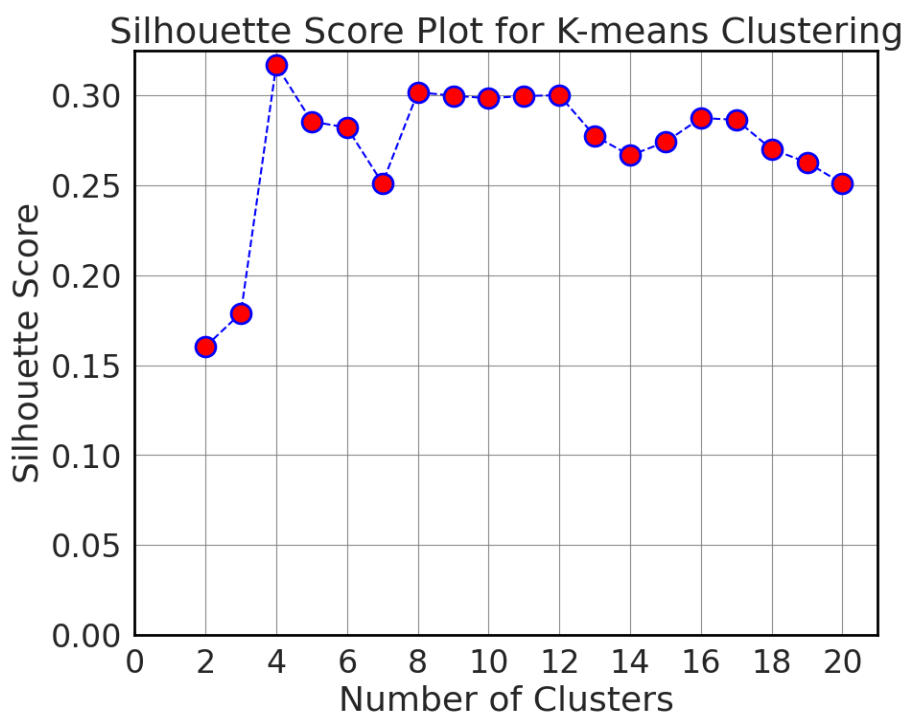
The plot shows that retaining six principal components, accounts for most of the variance in the dataset (**see Section 2.2.3.6**). This indicates that six components are sufficient to summarise the original dataset, enabling effective analysis and visualisation. This approach was necessary because the dataset consisted of 19 features, making direct visualisation in a 2D or 3D space impractical. PCA was therefore employed to extract the most significant components that capture the highest variance in the data.



(a)



(b)



(c)

Figure 4.1. (a) PCA-plot, (b) Elbow plot and (c) Silhouette plot for all age-hardenable alloy series

We then applied the Elbow method to determine the optimal number of clusters in the dataset. The elbow plot (Figure 4.1(b)) shows a “kink” at $K = 5$, indicating that 5 clusters are suitable for the age-hardenable alloys. To further validate this, we calculated the silhouette score, which reached a maximum of 0.31 for 4 clusters (Figure 4.1(c)) (see **Section 2.2.3.7.1**). When comparing $K = 4$ and $K = 5$, it was observed that $K = 5$ produced well-separated clusters, consistent with the elbow method and domain knowledge.

After determining the number of clusters, K-means clustering results were visualised in a two-dimensional space using the first two principal components with maximum variance (Figure 4.2). The clustering results indicate the presence of five distinct clusters, with some overlap observed, particularly between clusters 0 and 1, suggesting some similarities in composition and properties within these two alloy clusters. This analysis shows that the existing age-hardenable alloys can be grouped into five clusters based on their properties and compositions, with greater emphasis placed on properties when conducting the further analysis in **Section 4.4**, which interprets the hybridisation results.

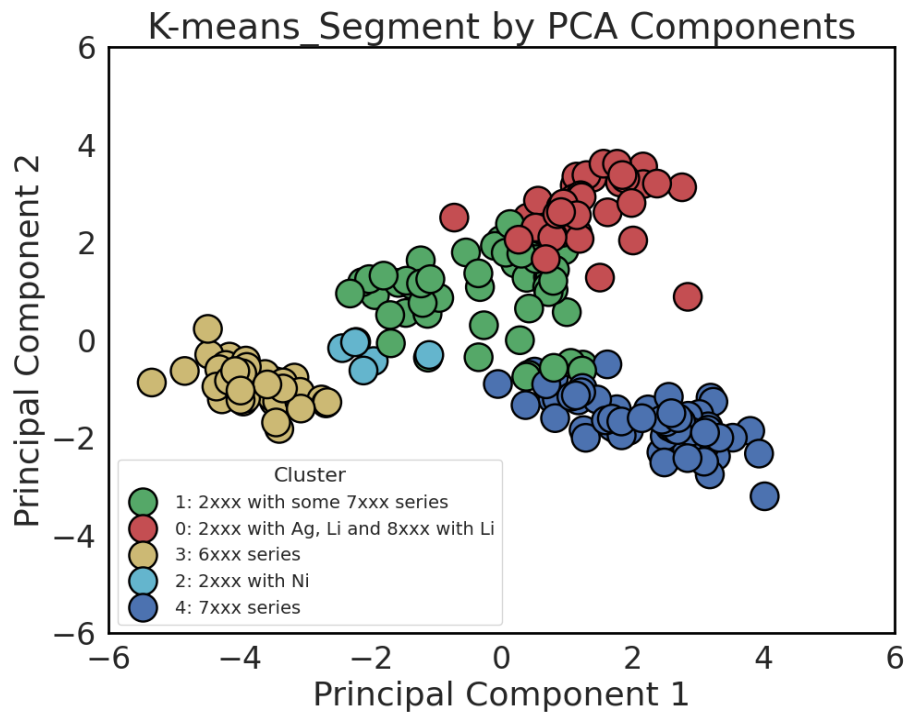
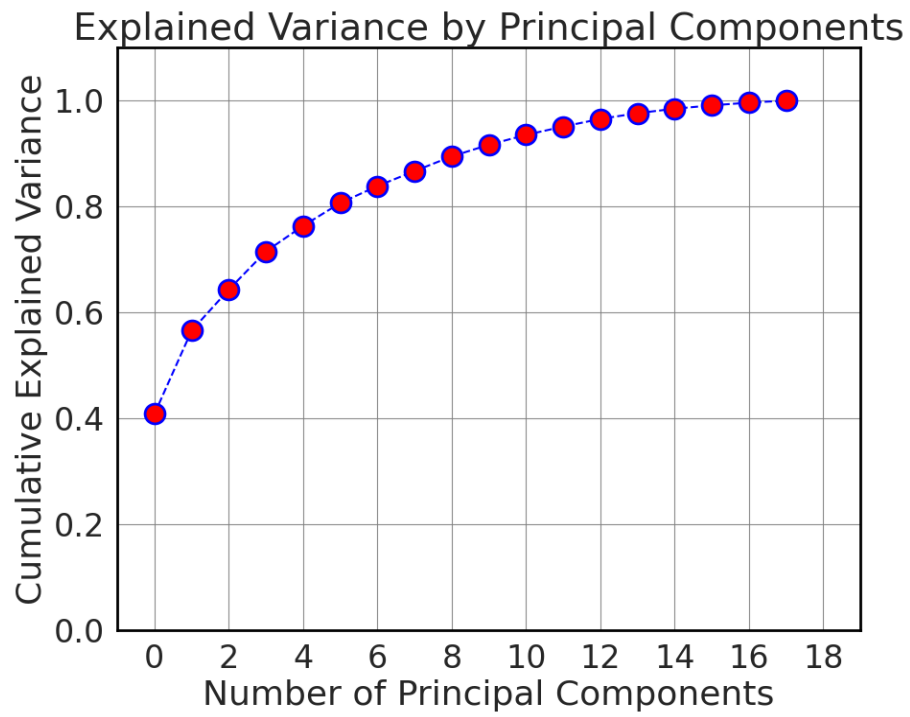


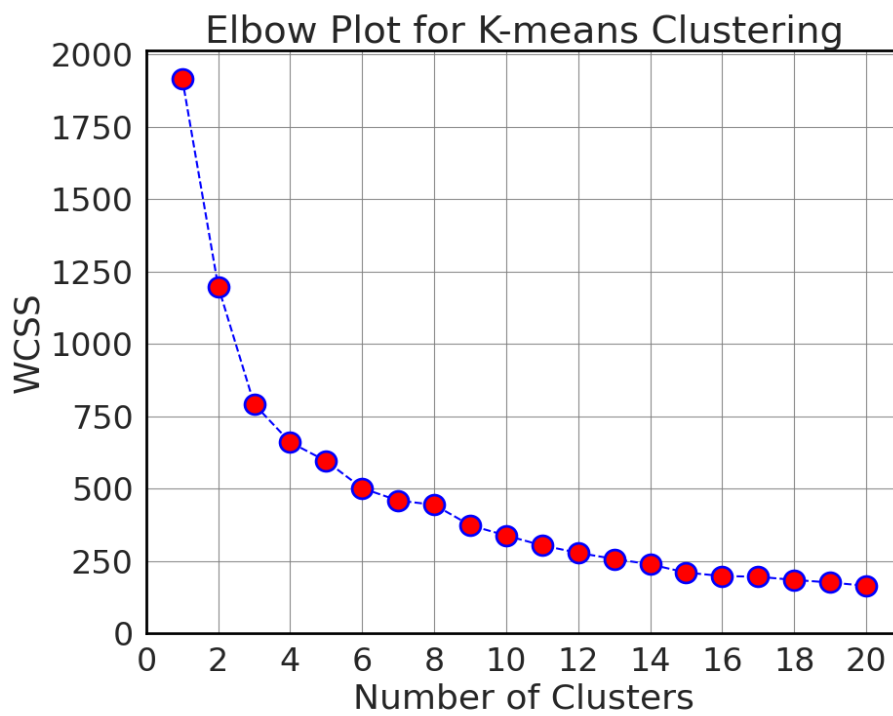
Figure 4.2. Clustering visualisation of age-hardenable alloy series

4.3.2 Clustering of All Non-Age-Hardenable Alloy Series

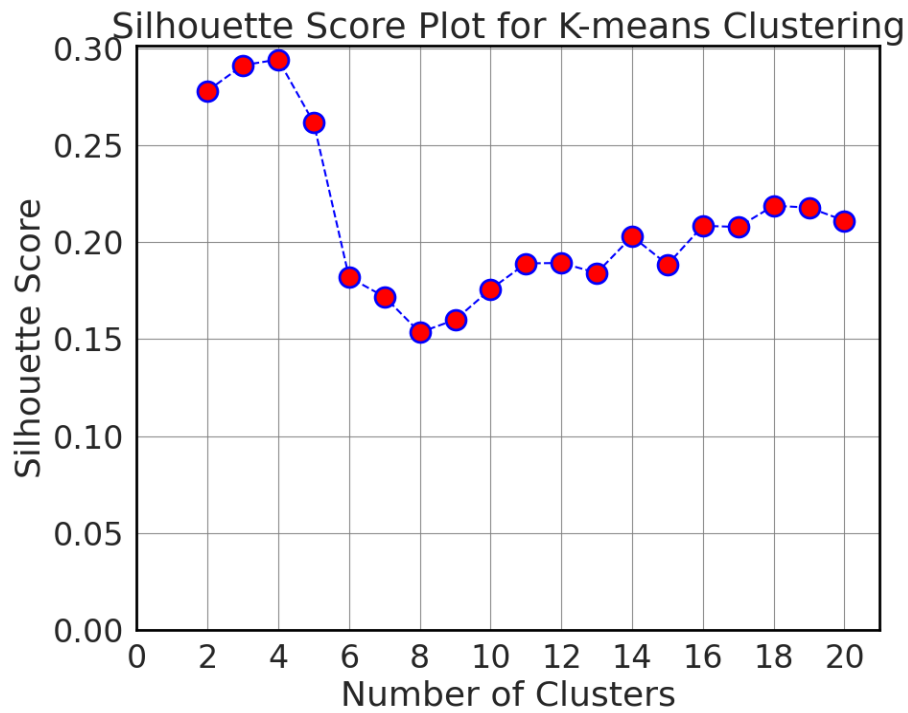
For clustering the non-age-hardenable alloy series, a similar methodology was employed. The PCA variance plot applied to the standardised dataset (Figure 4.3(a)) indicated that three principal components accounted for over 75% of the dataset's variance. Subsequently, the Elbow method (Figure 4.3(b)) was used to determine the optimal number of clusters, revealing a distinct kink at $K=3$ clusters. This finding was further validated by the highest silhouette score of 0.29, which also occurred at three and four number of clusters (Figure 4.3(c)), as described in **Sections 2.2.3.6 (PCA), 2.2.3.7.1 (K-means), and 3.4 (combined PCA & K-means clustering)**.



(a)



(b)



(c)

Figure 4.3. (a) PCA-plot, (b) Elbow plot, and (c) Silhouette plot for all non-age-hardenable alloy series

We visualised the K-means clustering results using the first two principal components, as shown in Figure 4.4. The three clear clusters showed no overlaps, showcasing the groups of alloys with similar compositions and properties grouped tightly. Consequently, the initial 1xxx, 3xxx, 5xxx, and 8xxx series were consolidated into three distinct clusters.

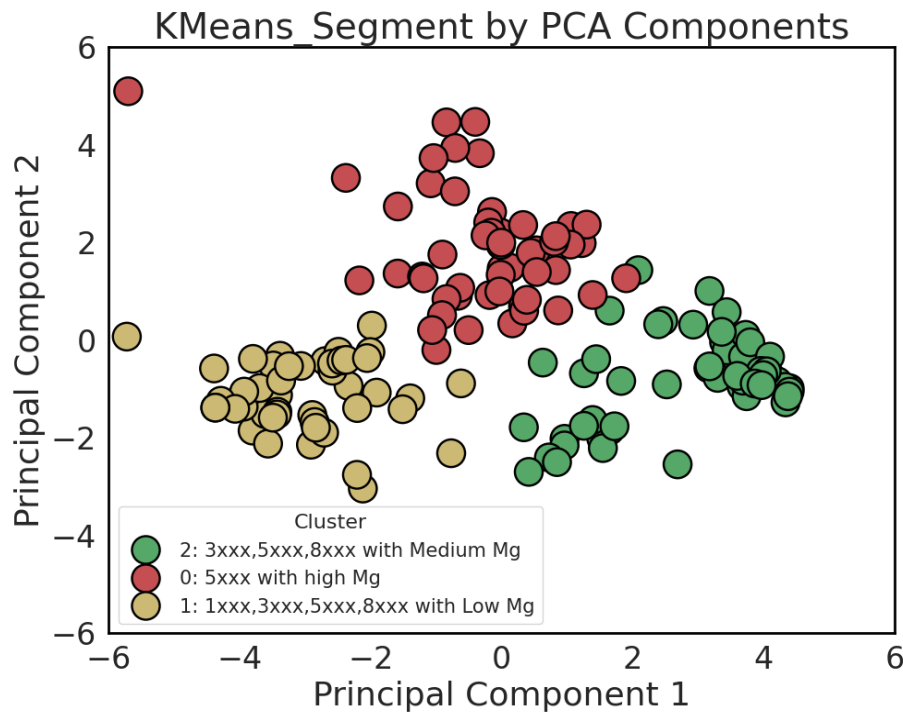


Figure 4.4. Clustering visualisation of non-age-hardenable alloy

4.4 Interpreting Hybridisation Results: New Classification Based on Properties

4.4.1 Age-Hardenable Alloys Series Hybridisation: Interpretation

For the age-hardenable alloys, clustering analysis yielded five distinct clusters (Figure 4.2), clearly differentiated by their compositions and properties. Some overlap was observed between clusters 0 and 1, indicating shared characteristics among these alloys that may warrant further investigation or optimisation in material design. This overlap also suggests that the alloys in these clusters may exhibit similar behaviour for certain applications. Further analysis of the clustering revealed that:

Cluster 0 comprises corrosion-resistant alloys, specifically the 2xxx and 8xxx series containing Ag and Li. These alloys are particularly suitable for applications in critical environments where corrosion resistance is essential.

Cluster 1 includes the 2xxx series and some 7xxx series alloys, which are noted for poor weldability and a reduced corrosion resistance; hence, these are not suitable for applications requiring harsh conditions.

Cluster 2 contains Ni-containing 2xxx series alloys. This cluster is characterised by weldable alloys with medium thermal conductivity, making them ideal for applications requiring good weldability and thermal performance.

Cluster 3 primarily comprises 6xxx series alloys, characterised by medium strength and high electrical and thermal conductivity. These alloys are well suited for applications requiring moderate strength combined with excellent conductivity.

Cluster 4 comprises alloys with high fatigue and yield strengths, predominantly Cu-containing 7xxx series alloys. These alloys are suitable for structural applications demanding both high yield and fatigue strength.

4.4.2 Non-Age-Hardenable Alloys Series Hybridisation: Interpretation

After hybridising the non-age-hardenable alloy series, 1xxx, 3xxx, 5xxx, and selected 8xxx series alloys, three distinct clusters emerged (see Figure 4.4), each representing alloys with similar ranges of properties and compositions.

Further analysis revealed that Cluster 0 primarily comprises 5xxx series alloys with high Mg content. These alloys exhibit high yield strength, fatigue strength, and modulus of toughness, making them well suited for applications demanding both strength and durability.

In contrast, Cluster 1 comprises a mixture of alloys from the 1xxx, 3xxx, 5xxx, and 8xxx series with lower Mg content. These alloys are known for their good thermal and electrical conductivity, making them ideally suited for applications requiring efficient conduction of heat and electricity. Cluster 2 comprises 3xxx, 5xxx, and selected 8xxx series alloys with medium magnesium content, resulting in moderate strength properties.

These clustering results suggest a new classification based on the alloys' properties rather than their series alone, making them suitable for different sets of industrial and engineering applications. This classification based on alloy properties offers a broad overview that aids in selecting alloys according to specific application requirements. Understanding the general properties of each cluster facilitates informed selection of alloys for applications where strength, conductivity, corrosion resistance, or weldability are critical considerations.

4.5 Conclusion: Hybridisation of Age-Hardenable and Non-Age Hardenable Alloy Series

This chapter aimed to investigate the potential for hybridising various aluminium alloy series, including 1xxx, 2xxx, 3xxx, 5xxx, 6xxx, 7xxx, and selected 8xxx series, to evaluate their compatibility for effective inter-series mixing. The data was divided into two groups: age-hardenable alloys, comprising the 2xxx, 6xxx, 7xxx, and some selected 8xxx series alloys, and non-age-hardenable alloys, including the 1xxx, 3xxx, 5xxx, and selected 8xxx series. Following data collection, a combined approach of PCA and K-means clustering was applied to classify the alloys into distinct clusters. The clustering of age-hardenable alloys resulted in five clearly defined groups, differentiated primarily by strength, corrosion resistance, weldability, and thermal and electrical conductivity. For instance, Cluster 0 comprised alloys characterised by excellent corrosion resistance, Cluster 3 encompassed alloys with high thermal and electrical conductivity, and Cluster 4 consisted of high-Cu 7xxx series alloys exhibiting superior strength. These clustering outcomes clearly indicate that alloys can be systematically categorised based on their dominant properties, thereby informing alloy selection and recycling strategies. For the non-age-hardenable alloys, the combined PCA and K-means clustering approach yielded three distinct clusters based on composition and property ranges. Cluster 0 consisted primarily of alloys with high magnesium content, exhibiting elevated yield strength, fatigue strength, and modulus of toughness, making them suitable for applications requiring strength and durability. Cluster 1 comprised alloys with lower magnesium content and high thermal and electrical conductivity, while Cluster 2 included alloys displaying properties within a medium range. Overall, the clustering analysis demonstrated the potential for alloy series recycling by identifying alloys that can be effectively blended while maintaining critical properties such as strength, corrosion resistance, and thermal conductivity. This series hybridisation approach offers a practical strategy for optimising aluminium alloys for diverse industrial applications, enabling performance enhancement without sacrificing essential material characteristics.

CHAPTER 5: RESULTS AND DISCUSSION – OPTIMISATION OF 6xxx SERIES ALUMINIUM ALLOYS [146], [147]

In this chapter, we performed the optimisation of 6xxx series alloys by implementing the methodology discussed in **Chapter 3: Methodology of Study**, with the aim of reducing the solution space of the 6xxx series to a smaller set of optimum alloys that exhibit comparable properties and applications to existing standard alloys.

5.1 Data Collection and Pre-processing

The accuracy of machine learning prediction depends on the availability of a high-quality dataset. We gathered 279 sets of data on 41 wrought 6xxx-series aluminium alloys from the Encyclopaedia of Aluminium and Its Alloys [137], the ASM Speciality Handbook [1], the MakeItFrom website [138], and various literature sources. The data set involved a number of quantitative and qualitative properties at T5, T6, and T7 tempering conditions (as defined in **Section 2.1.5**) and at various forming processes, i.e., extrusion, forging, plate, sheets, and rolling (as defined in **Section 2.1.4**) [148]. Every stage of artificial ageing gives the alloys a unique set of properties. In terms of data collection, we only considered artificial ageing, as this is the main temper to achieve high strength of 6xxx series alloys. Similar methodology can be applied to other products, tempers, and alloying systems. The numerical columns in the dataset represented chemical compositions spanning the following intervals: $0.35 \leq \text{Si} \leq 1.35 \%$, $0.075 \leq \text{Fe} \leq 0.5 \%$, $0.1 \leq \text{Cu} \leq 0.95 \%$, $0 \leq \text{Mn} \leq 0.85 \%$, and $0.35 \leq \text{Mg} \leq 1.4 \%$. The tensile properties depended on the tempering conditions employed. For instance, for T6 tempering, the property ranges were as follows: $160 \leq \text{Yield strength (YS)} \leq 430 \text{ MPa}$, $205 \leq \text{Ultimate tensile strength (UTS)} \leq 483 \text{ MPa}$, $3 \leq \text{Elongation} \leq 15 \%$, $13.14 \leq \text{Modulus of toughness} \leq 49.5 \text{ MPa } \%$, and $62 \leq \text{Fatigue strength} \leq 180 \text{ MPa}$. It should be noted that the chemical composition data were standardised to ensure consistency and to facilitate the clustering algorithm by:

- Averaging the main alloying elements (Mg, Si) to represent typical content.
- Taking the maximum values for alloying elements that enhance properties (Cu, Mn).
- Using half of the maximum allowable value for impurity elements (Fe) to balance their influence. This pre-processing step also helps control impurity levels in recycled alloys.

To facilitate easy processing in ML algorithms, the ratings of categorical columns, i.e., corrosion resistance, weldability, and anodising response, were transformed into numerical values as Poor = 0, Fair = 1, Good = 2, and Excellent = 3. The total solute content (Mg + Si + Cu) ranged from 1 to 2.7 %, indicating that higher solute levels were associated with greater precipitation of strengthening phases. Furthermore, the ratio of Mg to Si varied from 0.34 to 1.71. This ratio was important because it impacted the kinetics of precipitation hardening in the alloy, which in turn affected the mechanical properties [149].

For transparency and reproducibility, the complete raw dataset, processed dataset at each step, and reproducibility notes are provided in **Appendix Sections 11.3.4.1, 11.3.4.2, and 11.3.4.3**, respectively.

5.2 Data Visualisation to Understand the Dataset

Histograms: Distribution of Compositions and Properties in Dataset

To understand the dataset, histogram analysis was performed on the 6xxx series, as shown in Figure 5.1 (see **Section 2.2.3.1** for details). The distributions of chemical compositions and properties revealed several key patterns. Si values were concentrated around 0.6 and 1.0 %, indicating two commonly occurring levels in the dataset. Mg showed a higher frequency between 0.8 and 1.0 %, with a smaller secondary peak around 1.2–1.3 %. Fe displayed a peak near 0.25 % with fewer samples at higher values, while Cu values were primarily concentrated near 0.2 %, indicating limited variation.

Regarding mechanical properties, the modulus of toughness exhibited a roughly symmetric distribution, clustered around 25–35 MPa %, whereas yield strength was slightly skewed toward lower and median values. The Mg:Si ratio was highly skewed.

Overall, these histograms highlight the variation and skewness in the dataset, providing insight into the typical compositions and property ranges of the 6xxx series alloys.

Distribution of Features

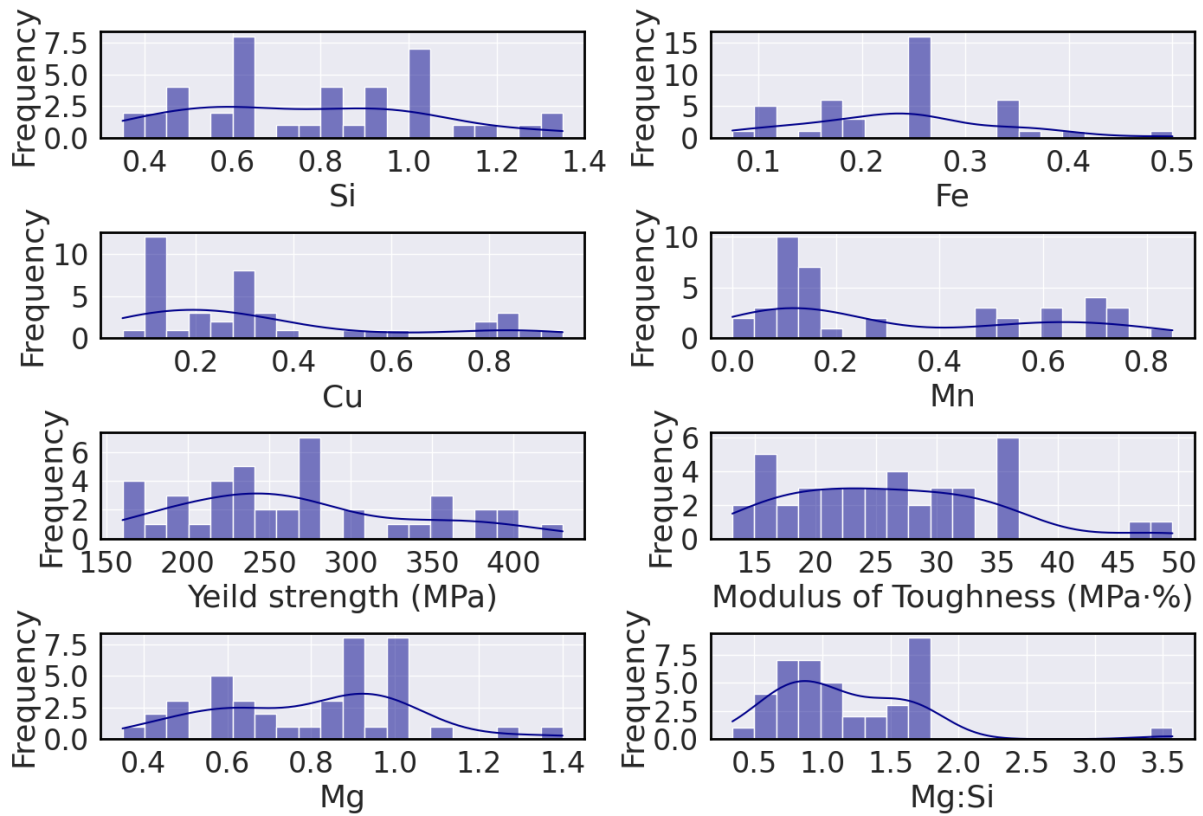


Figure 5.1. Histogram analysis

Box Plot Analysis: Analysing Outliers in the Dataset

A box plot analysis was conducted to identify outliers in the dataset, as shown in Figure 5.2 as described in **Section 2.2.3.2.1**. This enabled the detection of extreme values in both properties and compositions. Closer analysis showed that a majority of these outliers were confined to the areas of alloy compositions such as Cu and Fe, and properties including yield strength, corrosion resistance and anodising.

However, these extreme compositions and properties were considered potentially significant. They often represented ultra-performing alloys or unusual compositions that extended beyond the conventional performance envelope. Instead of being dismissed as anomalies, these extreme values were retained, as they could offer valuable insights into the limits of material performance and highlight rare but critical alloys.

For example, an alloy with an unusual composition might exhibit properties that surpassed all others, such as exceptional corrosion resistance or outstanding strength, making it crucial for

innovative applications. Extremes in properties, like yield strength, could also indicate novel material compositions. Removing these outliers could have eliminated opportunities to discover groundbreaking materials or identify compositions with extraordinary potential. Therefore, these outliers were preserved for further analysis to maintain their valuable contributions and facilitate the exploration of innovative materials and applications.

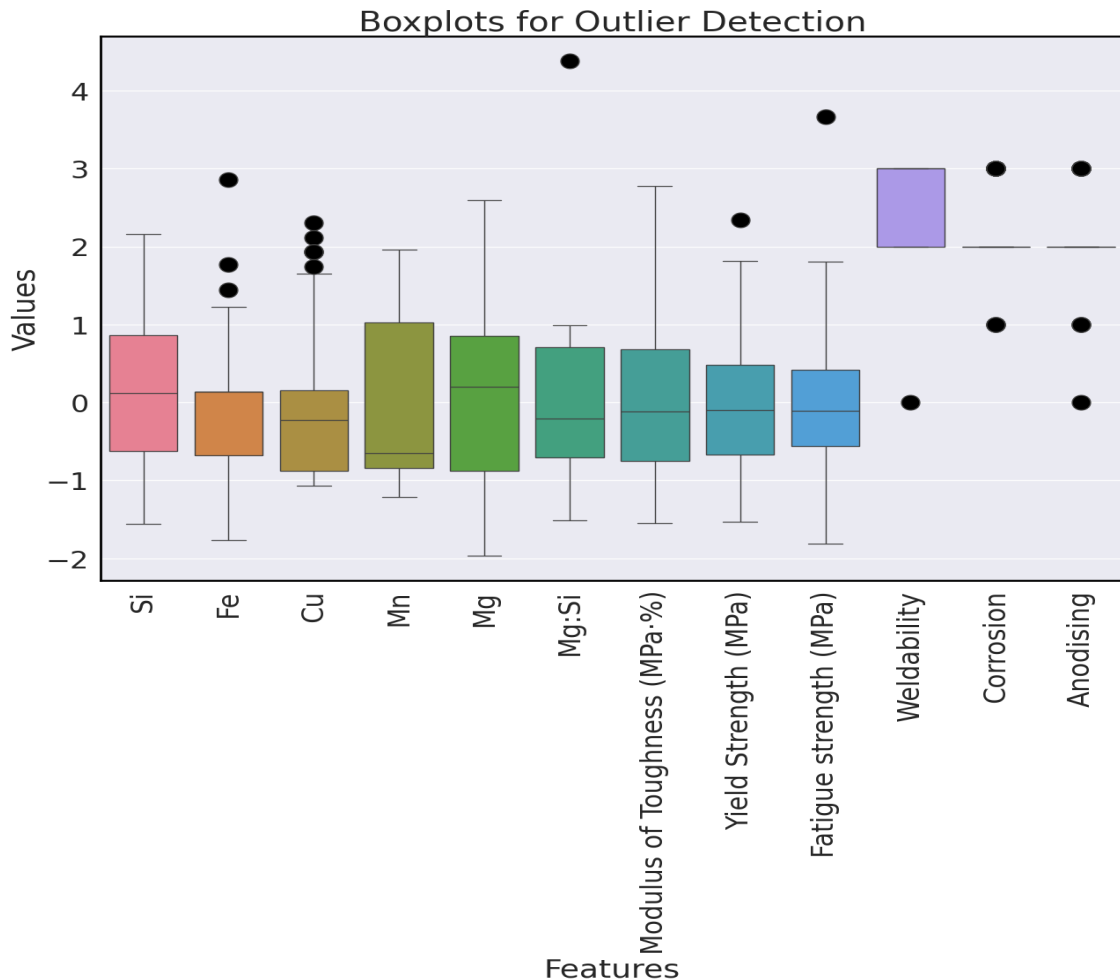


Figure 5.2. Outlier detection in dataset

Pair-plot: Pairwise Relationships Between Composition and Properties

We plotted correlations between selected features using a pair plot (Figure 5.3) as described in **Section 2.2.3.1**, to explore how alloy composition influences properties. The selected features were Si, Mg, Fe, Cu, Mn, yield strength, and modulus of toughness. The pair plot revealed several trends in the relationships between elements and alloy properties:

- Yield strength showed slight positive correlations with Mg, Mn, and Cu, suggesting their strengthening effect.
- Si exhibited little independent influence on yield strength, indicating that its effect may depend on interactions with other elements.
- Fe displayed scattered behaviour with respect to yield strength, reflecting inconsistent influence that may arise from interactions with other elements or microstructural factors.

The histograms along the diagonal provided insights into feature distributions:

- The histogram for Cu was skewed, with most values concentrated in the lower range.
- Yield strength spanned a wide range, reflecting variations in alloy composition and processing conditions.
- Si was slightly skewed, Mn was roughly uniform, and Fe showed an uneven distribution.

Overall, the pair plot highlights the complex interplay between mechanical properties and composition. While Mg and Mn contribute to strengthening, other factors such as microstructure and processing play a critical role in determining the ultimate alloy properties.

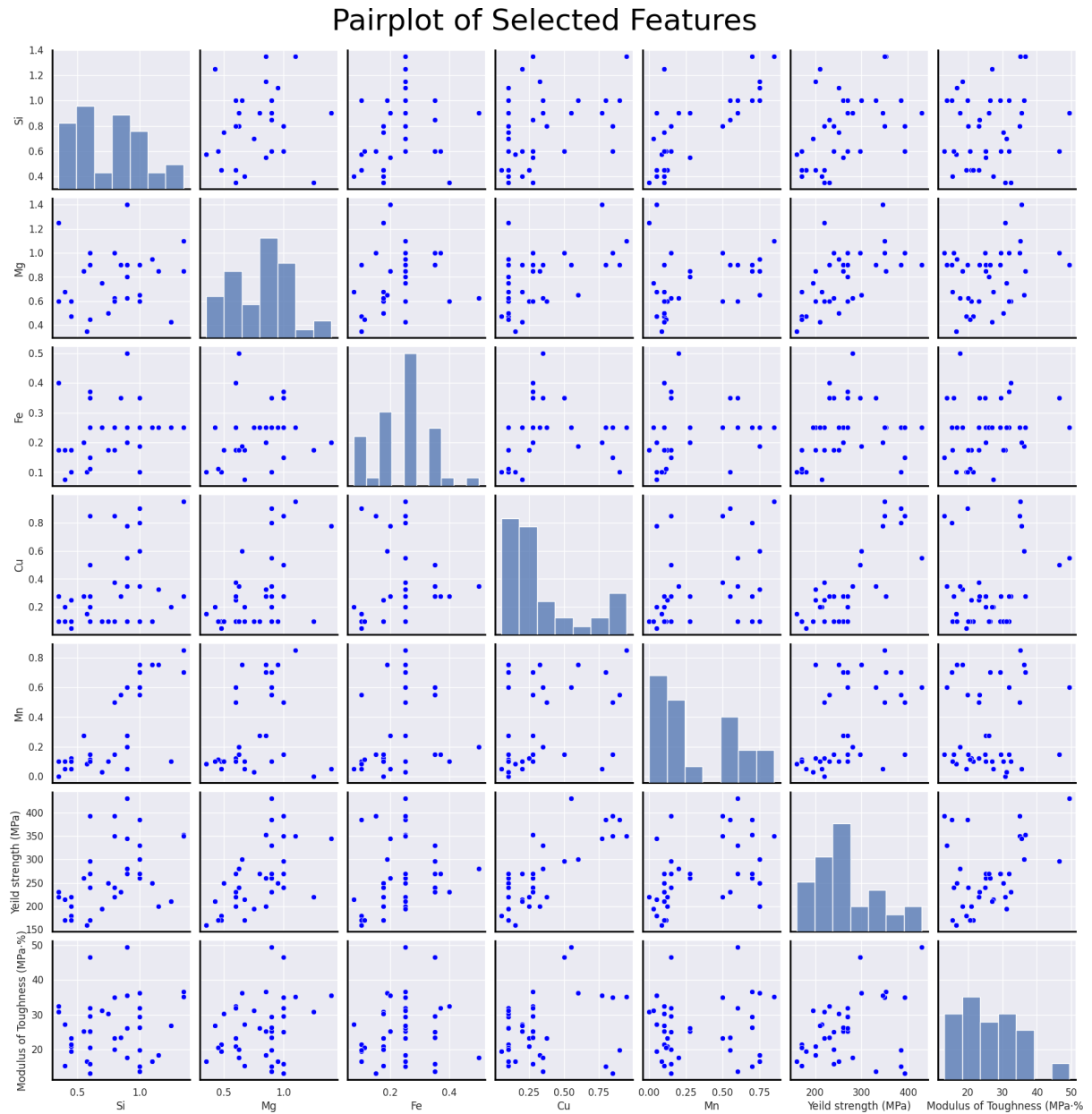


Figure 5.3. Pair-wise analysis of dataset

Correlation Analysis: Relation between Composition and Properties

We conducted Pearson correlation analysis (Figure 5.4) (see **Section 2.2.3.3**), which revealed significant correlations between certain compositions and properties. We observed the highest correlations between the following features:

- Fatigue strength and yield strength: 0.84
- Cu and yield strength: 0.80
- Si and Mn: 0.78

- Cu and Fatigue strength: 0.68

These findings demonstrated strong relationships between the properties and compositions, consistent with established metallurgical principles. For instance, the strong correlation between yield strength and fatigue strength is well documented, as both properties are governed by similar alloy composition and microstructure [150]. Furthermore, the observed correlation between copper content and mechanical properties, such as yield strength and fatigue strength, highlighted copper's contribution to enhanced alloy performance.

Although the correlations involving Si and Mn were deemed irrelevant to the current study, as the increase of one did not significantly affect the other, the focus was placed solely on the influence of composition on properties.

These relationships are examined in greater detail in **Section 5.9**, where the underlying metallurgical rationale for these correlations and their implications for alloy design are discussed.

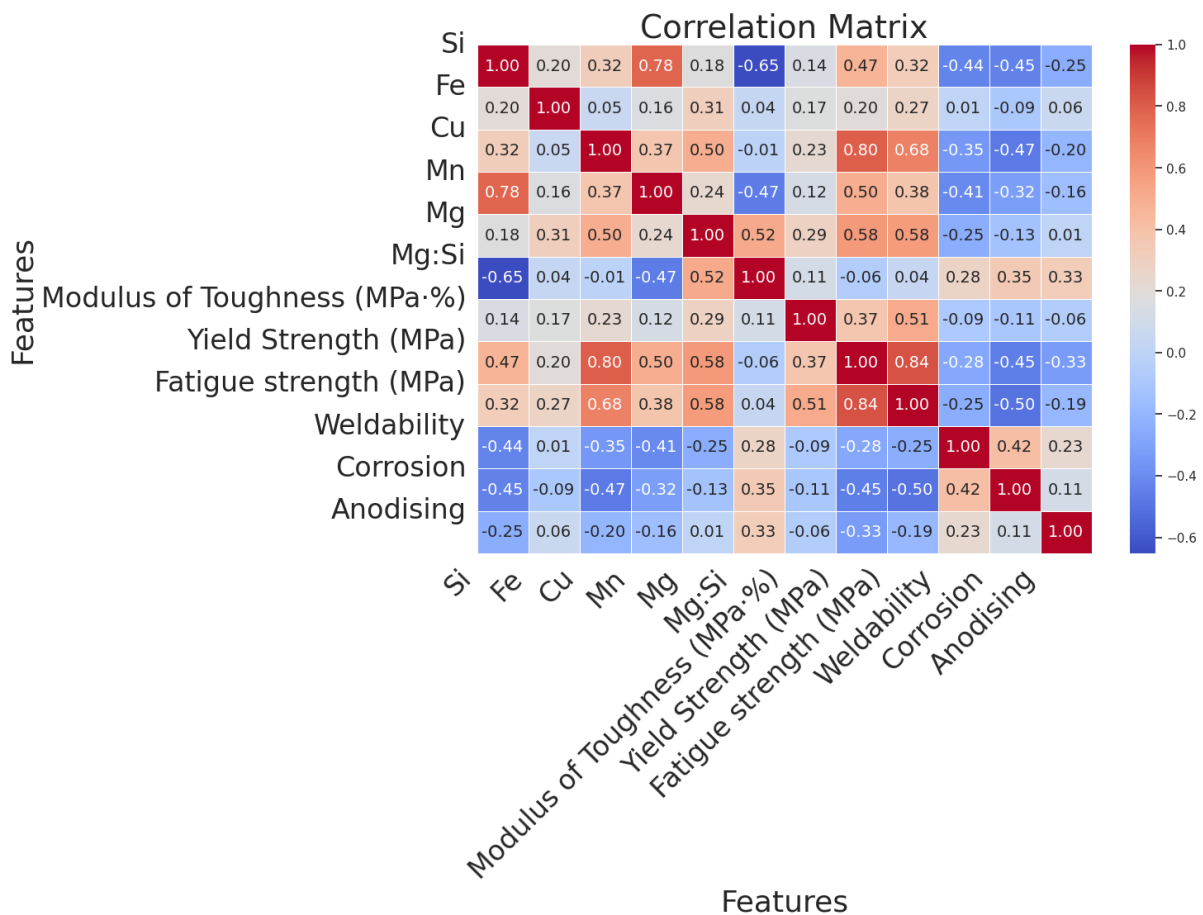
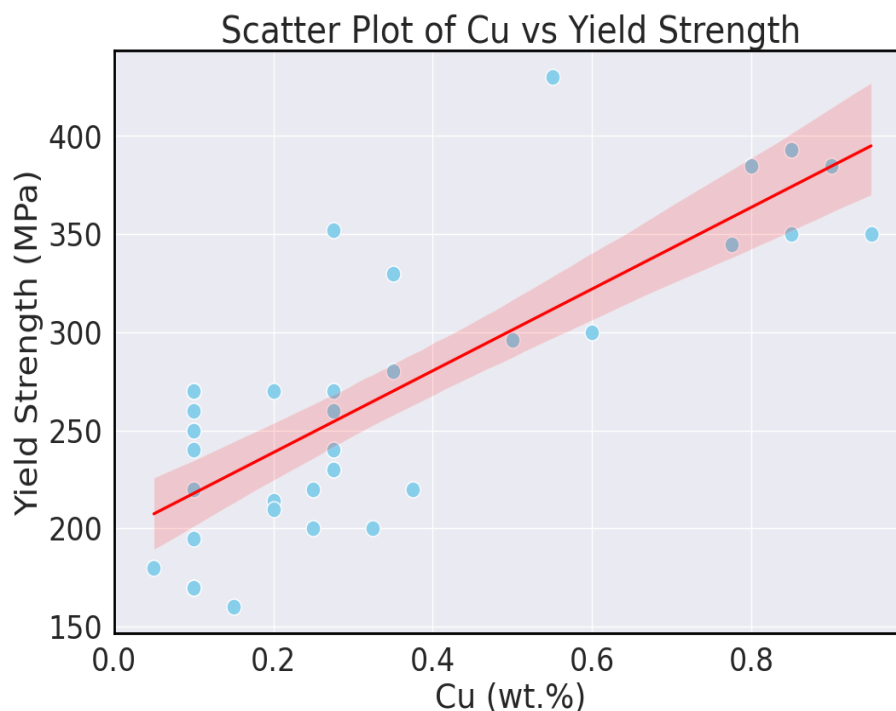


Figure 5.4. Correlation matrix of properties

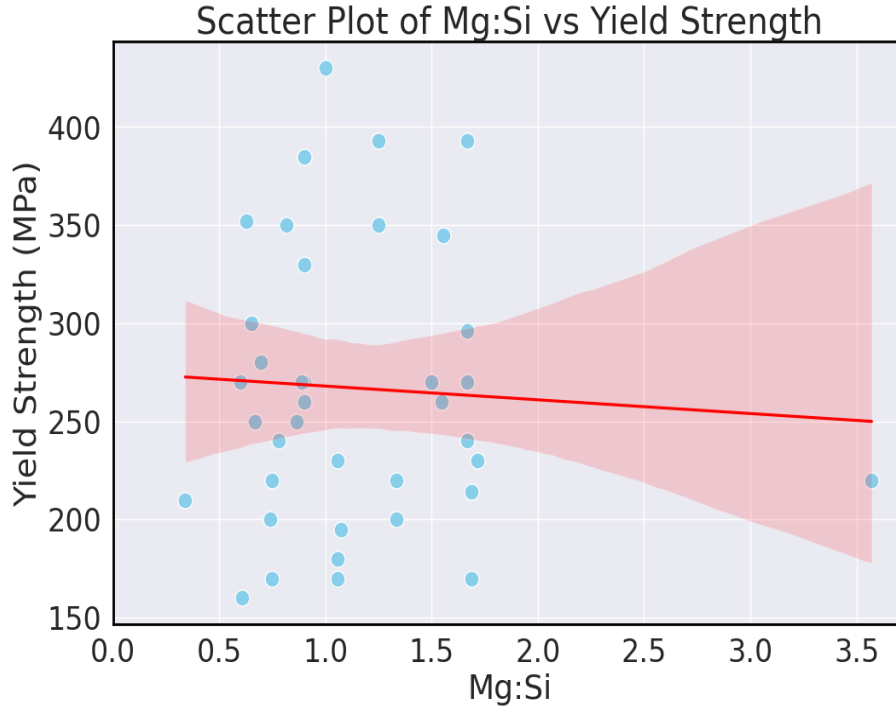
Scatter Plot for Yield Strength Relationship with Strengthening Compositions

The first scatter plot, shown in Figure 5.5(a), (see **Section 2.2.3.1**), illustrated the relationship between Cu content and yield strength. From the plot it was clear that there was a positive correlation between the two variables, meaning that as the Cu increased, the yield strength also increased. This indicated that higher copper content in the alloy yielded better mechanical properties and possibly made the material more resistant to deformation. This was supported by the red regression line, as it moved upwards, further strengthening the conclusion that copper had a positive effect on the material's strength.

By comparison, the second scatter plot, Figure 5.5(b), showed the relationship between yield strength and the Mg:Si ratio. In this graph, the trend did not indicate any apparent relationship between the two variables. Even with some fluctuations in the Mg:Si ratio, the yield strength did not significantly change or slightly decreased, showing a weak negative trend. This suggested that altering the Mg:Si ratio did not significantly affect the yield strength of the material, and other parameters possibly had a larger effect on the strength characteristics of the material. The lack of a clear trend in this graph supported the idea that the Mg:Si ratio did not have as much influence on the mechanical properties of the alloy as copper concentration did. The effects of other alloying elements on mechanical properties, will be examined in detail in **Sections 5.8 and 5.9**.



(a) Yield strength vs Cu



(b) Yield strength vs Mg:Si

Figure 5.5. Scatter plot: (a) Yield strength vs Cu and (b) Yield strength vs Mg:Si

Violin Plot for Analysing Distribution of Yield Strength in Dataset

The violin plot presented in Figure 5.6 (see **Section 2.2.3.1** for methodological details) illustrated the distribution of yield strength across varying levels of Mg:Si, namely low, medium, and high. Alloys with low Mg:Si exhibited the widest distribution, with yield strength ranging from 160 MPa to approximately 430 MPa, indicating substantial variability. Medium Mg:Si alloys displayed a more concentrated distribution, predominantly around 270 MPa, with some data extending toward higher values. Conversely, alloys with high Mg:Si formed a narrow band, reflecting minimal representation at this level. Overall, the plot demonstrated that yield

strength varied most at low Mg:Si levels and was nearly uniform at high levels. The influence of Mg:Si on strength properties is further analysed in **Section 5.9**.

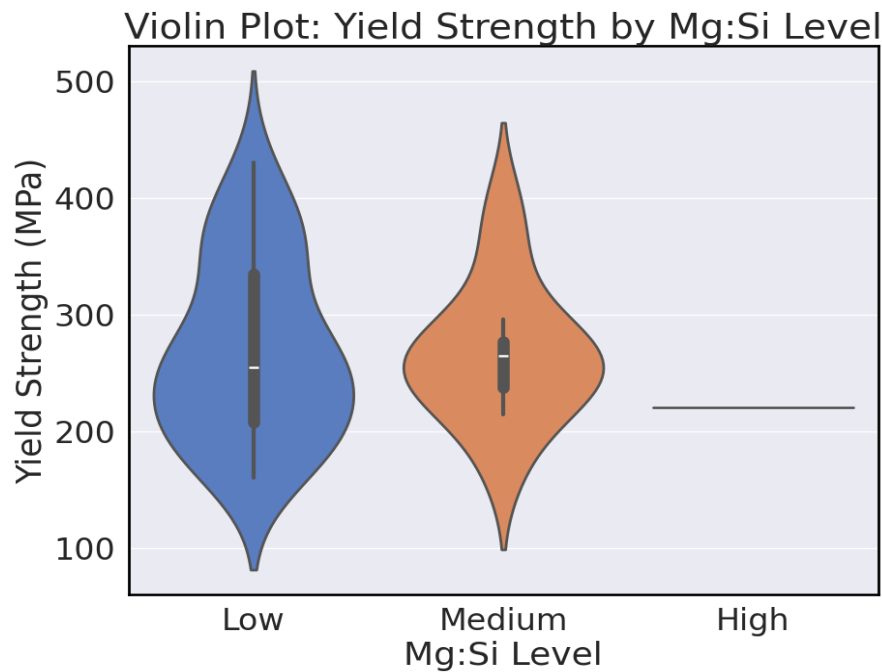
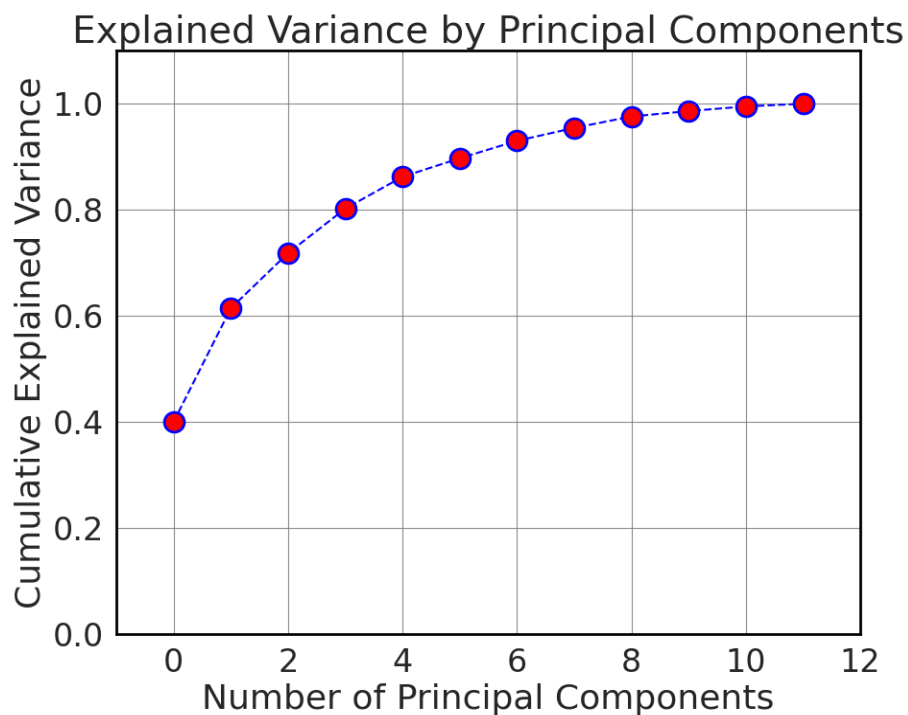


Figure 5.6. Violin plot for yield strength distribution

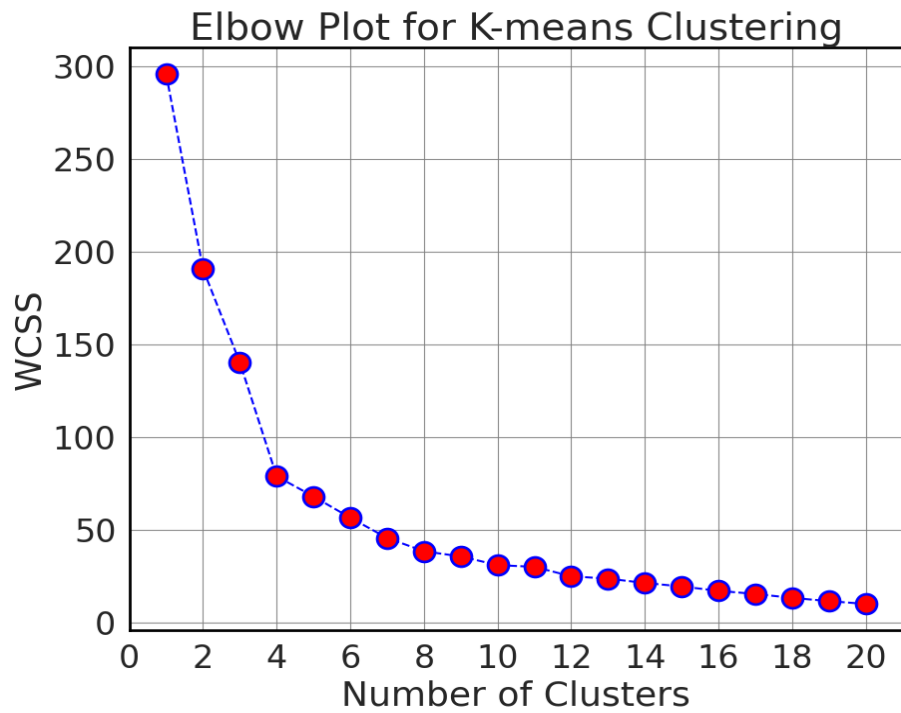
5.3 Combined Approach of PCA and K-means Clustering: Clustering of 6xxx Series Dataset

After analysing the data and gaining insights into the dataset, we applied a combined PCA and K-means clustering approach as described in **Section 3.4 (combined PCA & K-means clustering)**. This enabled us to identify patterns within the 6xxx series aluminium alloy dataset and group them into clusters with similar compositions and properties. We applied standard scaling to the entire dataset after concatenating the converted categorical features (i.e., corrosion resistance, weldability, and anodising response) with the numerical features (i.e., compositions, yield strength, modulus of toughness, and fatigue strength) using a joining function. Standard scaling ensured that each property contributed equally during the clustering process. For instance, silicon content ranged from 0.35 to 1.35, which was considerably lower than the yield strength, ranging from 160 to 430 MPa. Similarly, the categorical ratings, ranging from 0 to 3, showed notable disparities. Standard scaling addressed these inconsistencies by transforming the dataset into a more uniform scale, assigning a mean of 0 and a standard deviation of 1 to all alloy attributes.

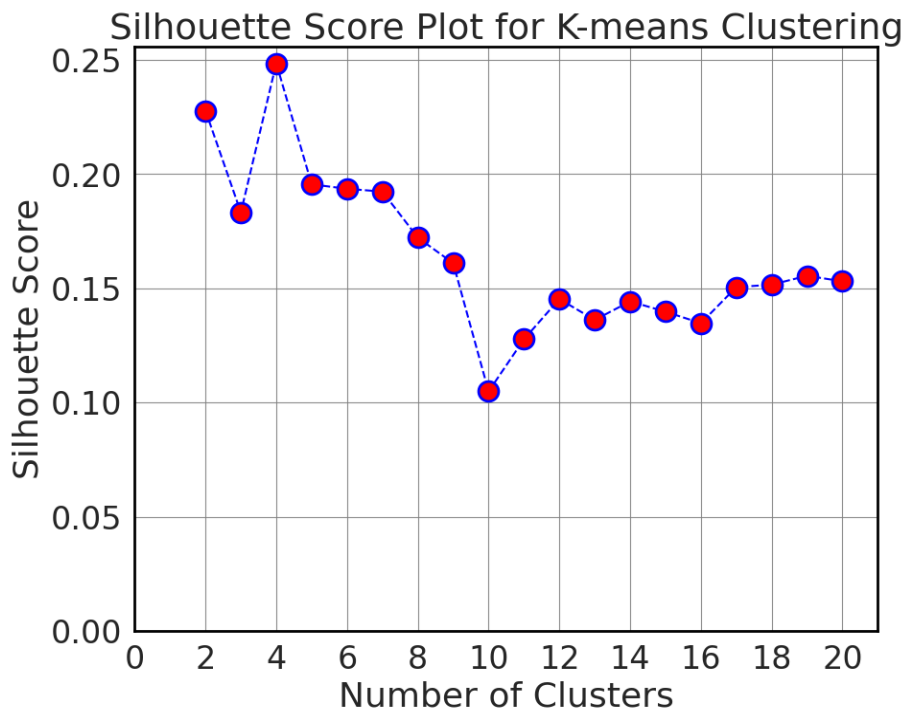
To clearly visualise the dataset and reduce its dimensionality, we employed a PCA variance plot (Figure 5.7(a)), to illustrate the relationship between cumulative explained variance (i.e., the proportion of variance explained by each component) and the number of principal components, (See **Section 2.2.3.6**). Four principal components were selected, corresponding to 80% of the dataset information, as indicated by the curve after performing hyperparameter tuning [151]. This reduced the dataset dimensionality from 11 features to 4 components, enabling easier cluster visualisation in a 2D space.



(a)



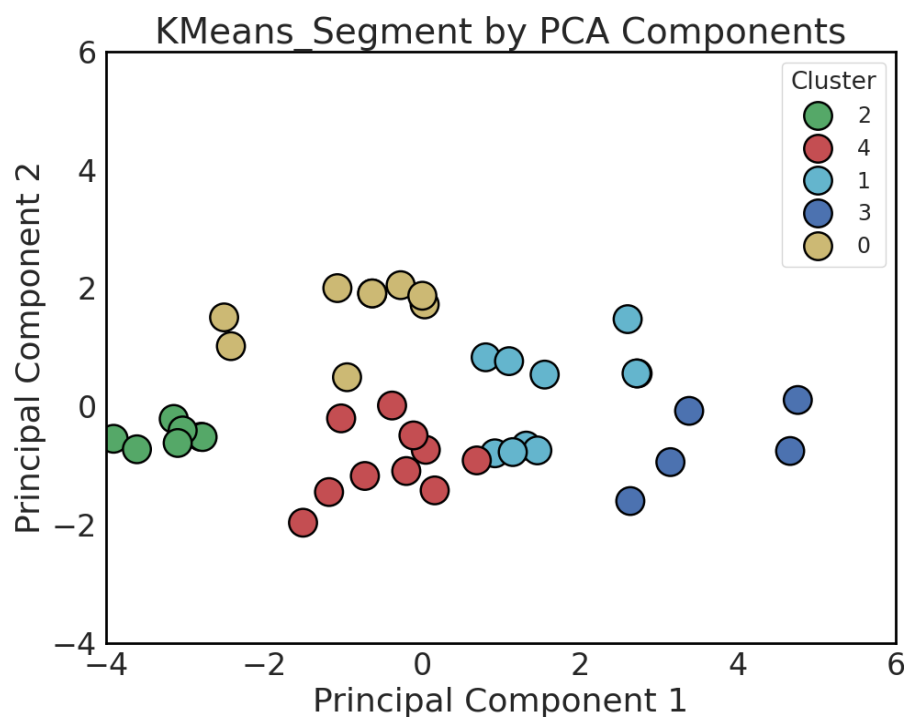
(b)



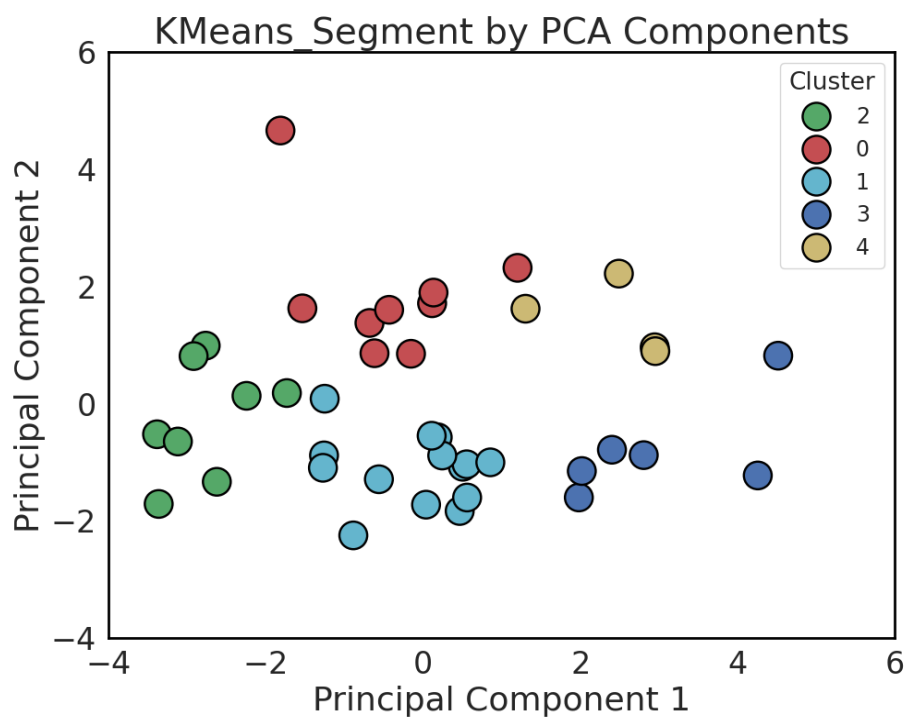
(c)

Figure 5.7. PCA variance plot (a), Elbow plot (b), and Silhouette plot (c) for determining the optimal number of clusters

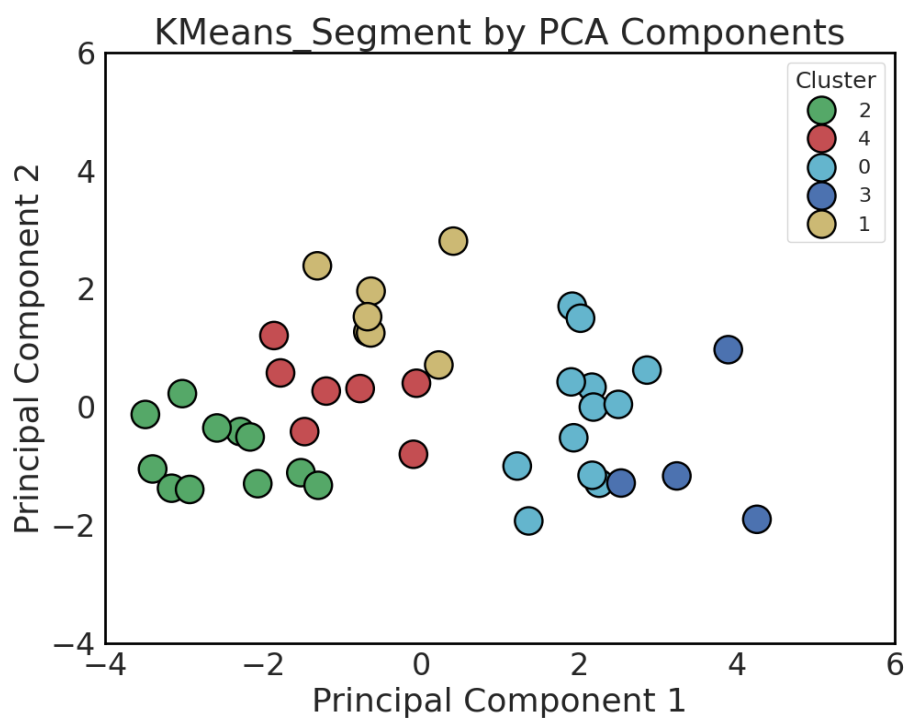
We employed an elbow plot on the low-dimensional feature dataset, as illustrated in Figure 5.7(b), showing WCSS vs the number of clusters. The point on the x-axis where a kink appeared in the curve indicated the number of clusters present in the dataset. As shown, the kink occurred between points, $K=4$ and 5 . After testing both values for the number of clusters, and using different combinations of principal components, we identified five well-defined clusters corresponding to point 5, with a silhouette score of 0.20 (Figure 5.7(c)). Figure 5.8(b) further illustrates these five clusters at T6 temper (see **Section 2.2.3.7.1** for methodological details).



(a) T5



(b) T6



(c) T7

Figure 5.8. Clustering results for (a) T5, (b) T6, and (c) T7 tempering conditions

Similar clustering patterns were observed under the T5 and T7 tempering conditions, as shown in Figures 5.8(a) and 5.8(c), with comparable alloy grades present within each cluster. The clustering results remained consistent across various processing conditions (i.e., rolling, forging, and extrusion) of 6xxx series aluminium alloys. This further supported our findings, demonstrating that the clustering outcomes were unaffected by variations in processing conditions, as shown in **Appendix Section 11.1**.

Despite the presence of outliers, as shown in Figure 5.2, K-means clustering produced well-defined segments, including those representing extreme material properties. Removing these outliers could alter the cluster boundaries and risk excluding rare but potentially valuable segments of the data.

Based on the ranges of alloy properties and composition values within each cluster, we found the clusters to be largely distinct, with only limited overlap in composition and property values, as summarised in Table 5.1.

Table 5.1. Composition and property variability across clusters

Property	Group 2 (Low)	Group 0 (Medium)	Group 1 (Medium)	Group 4 (Highest)	Group 3 (High)
Si, wt%	0.4 – 0.6	0.35 – 0.6	0.7 – 1.25	0.6 – 0.9	1.0 – 1.35
Fe, wt%	0.075 – 0.175	0.175 – 0.4	0.175 – 0.5	0.15 – 0.25	0.1 – 0.35
Cu, wt%	0.05 – 0.25	0.1 – 0.5	0.1 – 0.375	0.55 – 0.85	0.275 – 0.95
Mn, wt%	0.05 – 0.125	0 – 0.275	0.03 – 0.75	0.05 – 0.6	0.55 – 0.85
Mg, wt%	0.35 – 0.675	0.6 – 1.25	0.425 – 0.95	0.9 – 1.4	0.65 – 1.1
Yield strength, MPa	160 – 220	220 – 296	195 – 280	345 – 430	300 – 385
Modulus of Toughness (MoT)	15.4 – 27.3	15.9 – 46.5	16.66 – 32	35.035 – 49.5	13.8 – 36.55
Fatigue strength	62 – 88	89.6 – 119.3	88 – 110	125 – 180	94 – 130
Corrosion resistance (CR)	Good – Excellent	Good-Excellent	Good – Excellent	Fair- Good	Fair – Excellent
Weldability	Good – Excellent	Excellent	Good – Excellent	Good	Poor – Excellent
Anodising	Good – Excellent	Poor- Excellent	Good	Good	Fair – Good
Mg:Si	0.61 – 1.68	1.5 – 1.71	0.34– 1.07	1– 1.67	0.63 – 0.90
Mg+Si+Cu	1.2	1.9	2	2.7	2.5
Excess Si	Slight excess	Stoichiometry	Large excess	Excess	Large excess

The clusters were ranked in ascending order according to the concentration of critical alloying elements and their associated properties as follows:

YS = 2<1<0<3<4, MoT = 2<1<3<0<4, Fatigue Strength = 2<1<0<3<4, Si = 2<0<4<1<3, Cu = 2<1<0<4<3, Mg = 2<1<0<3<4, Mn = 2<0<4<1<3

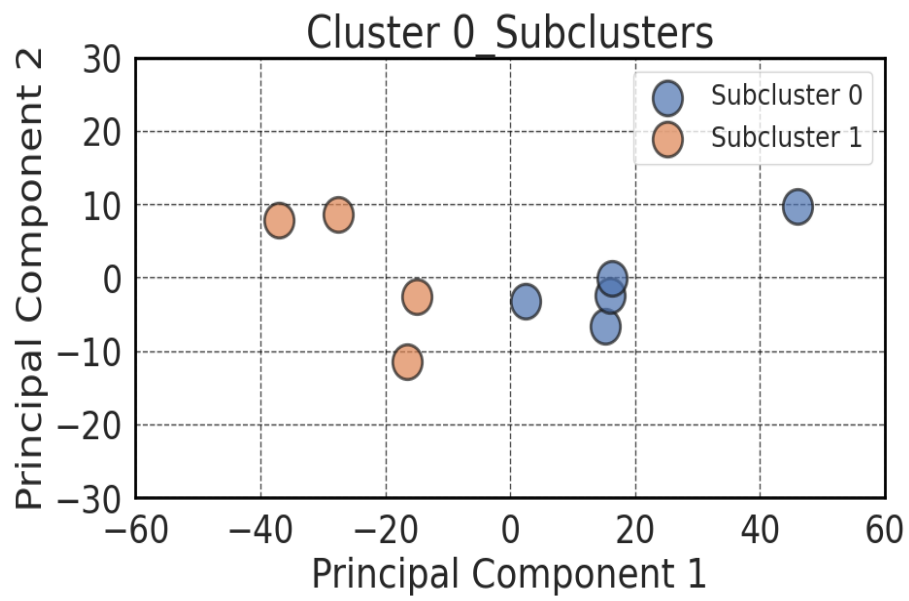
Cluster 4, which contains alloys with high amounts of solute elements ($Mg + Si + Cu \approx 2.7$), elevated Cu content, excess Si, and moderate Fe levels, exhibits the highest yield strength (YS). Cluster 0, characterised by medium levels of solute elements ($Mg + Si + Cu \approx 1.9$), moderate Cu content, and an Mg:Si ratio close to stoichiometry, shows medium YS values, with a correspondingly medium modulus of toughness (see Table 5.1). Cluster 1 contains medium levels of solute elements ($Mg + Si + Cu \approx 2$), low Cu content, medium Fe content, and an Mg:Si ratio of ≈ 1.071 , resulting in relatively lower YS and modulus of toughness.

Alloys in Cluster 2 exhibit the lowest YS, due to the low proportion of solute elements ($Mg + Si + Cu \approx 1.2$), low Cu content, and only a slight excess of Si. Cluster 3 alloys contain a high proportion of solute elements ($Mg + Si + Cu \approx 2.5$) with a large excess of Si, although the overall percentages of Cu and Mg are lower than in Cluster 4. Consequently, these alloys display the second highest YS.

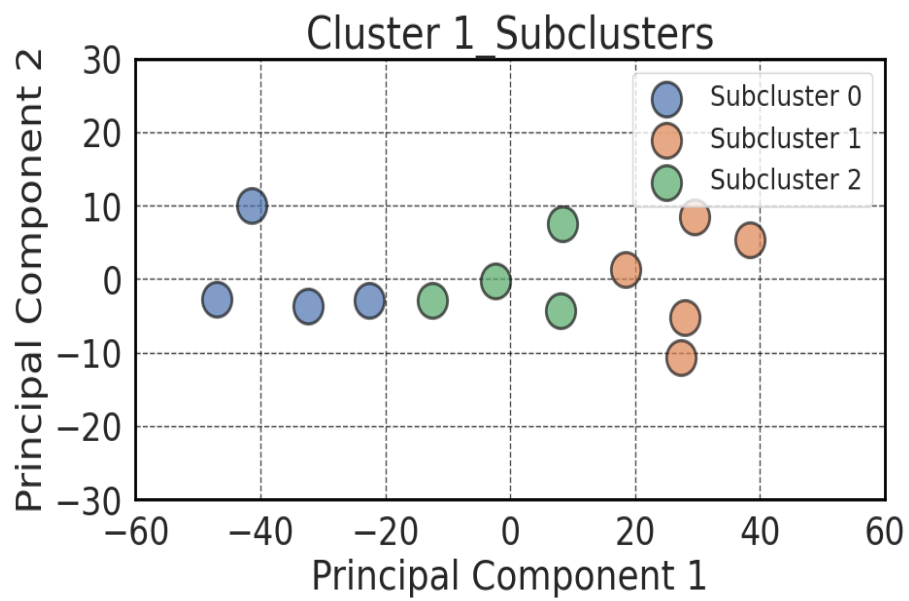
5.4 Sub-clustering of Clusters

After obtaining five well-defined clusters of alloys with similar properties and compositions, we further refined these clusters by applying the PCA and K-means clustering algorithm along with the mathematical model for sub-cluster division described in **Section 3.5**. This refinement aimed to generate finer, more compact sub-clusters with narrower property ranges, facilitating the selection of optimum alloys in the subsequent step. The sub-clustering process resulted in 11 sub-clusters, each characterised by a tight range of properties, as shown in Figure 5.9.

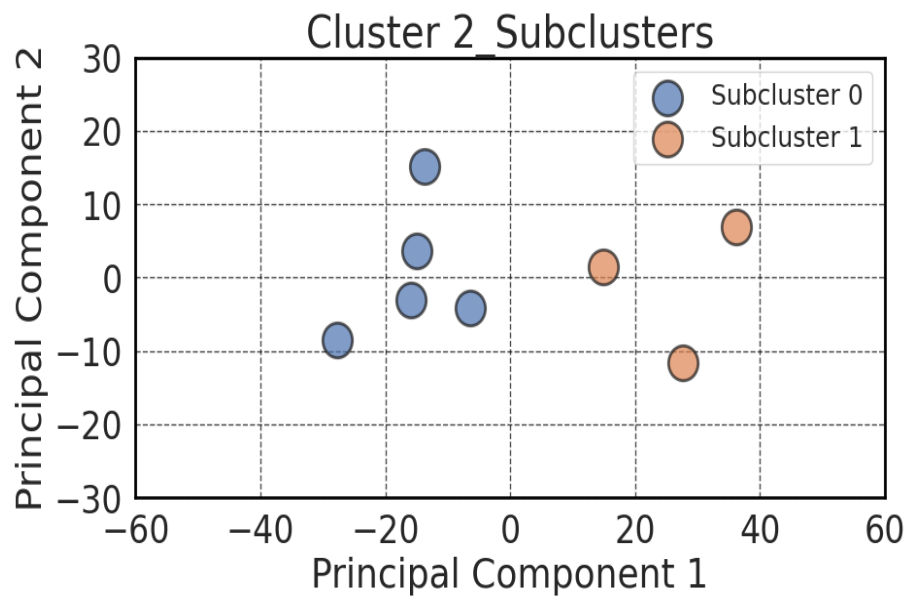
For example, Cluster 0 (Figure 5.9(a)) was subdivided into two sub-clusters: sub-cluster 00 and sub-cluster 01. Sub-cluster 01 exhibited a slightly lower range of tensile properties but superior weldability and corrosion resistance, whereas sub-cluster 00 displayed a higher range of tensile properties but moderate weldability and corrosion resistance. This pattern was consistent across all sub-clusters, demonstrating that the alloys were grouped according to a more precise set of characteristics.



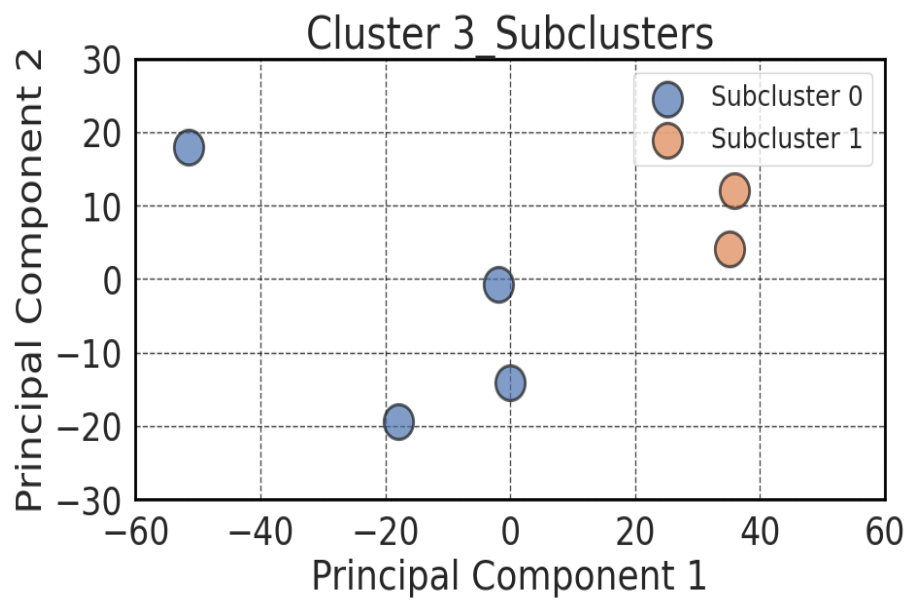
(a) Cluster 0, sub-clusters



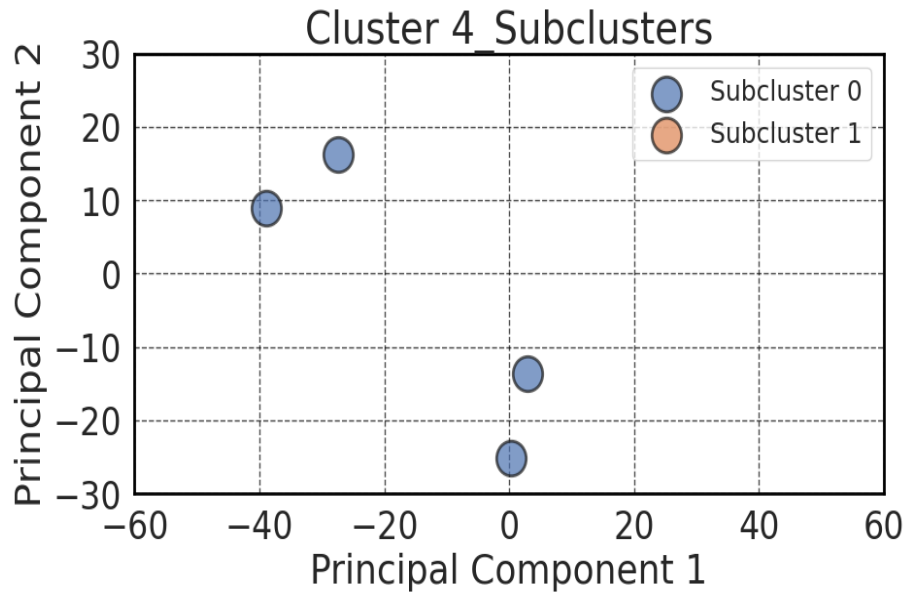
(b) Cluster 1, sub-clusters



(c) Cluster 2, sub-clusters



(d) Cluster 3, sub-clusters



(e) Cluster 4, sub-clusters

Figure 5.9. Sub-clustering analysis of clusters: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, (e) Cluster 4.

The next step involved the optimisation of these sub-clusters to identify representative alloys exhibiting the most favourable combination of properties.

5.5 Optimisation Algorithm

After obtaining sub-clusters with narrower property ranges, the next step was to optimise them by selecting alloys with the most favourable combination of properties using the TOPSIS optimisation algorithm as described in **Section 3.6**. TOPSIS is a robust multi-criteria decision-making method that ranks alloys based on their distance from ideal solutions, effectively balancing trade-offs among conflicting criteria. For example, sub-cluster 01 of Cluster 0 comprises five alloys with differing property ranges, as listed in Table 5.2.

Table 5.2. Actual and normalised values (in parentheses) of properties of alloys within sub-cluster 01 of Cluster 0

Property / Parameter	Alloy G1 (6040)	Alloy G2 (6261)	Alloy G3 (6162)	Alloy G4 (6262)	Alloy G5 (6065)
Yield Strength (MPa)	296 (1.00)	260 (0.00)	270 (0.28)	270 (0.28)	270 (0.28)

Modulus of Toughness (MPa%)	46.5 (1.00)	25.2 (0.00)	25.2 (0.00)	29.5 (0.20)	31.9 (0.32)
Fatigue Strength (MPa)	119.3 (1.00)	100 (0.00)	110 (0.52)	110 (0.52)	110 (0.52)
Corrosion Resistance	3 (1.00)	2 (0.00)	2 (0.00)	2 (0.00)	2 (0.00)
Weldability	3 (1.00)	3 (1.00)	2 (0.00)	3 (1.00)	3 (1.00)
Anodising Response	1 (0.33)	2 (0.67)	0 (0.00)	2 (0.67)	3 (1.00)
D_i[*] (Eq. 15)	$D_{G1}^* = \sqrt{\begin{matrix} (1-1)^2 + \\ (1-1)^2 + \\ (1-1)^2 + \\ (1-1)^2 + \\ (1-1)^2 + \\ (1-0.33)^2 \end{matrix}}$ ≈ 0.67	$D_{G2}^* = \sqrt{\begin{matrix} (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (1-1)^2 + \\ (1-0.67)^2 \end{matrix}}$ ≈ 2.01	$D_{G3}^* = \sqrt{\begin{matrix} (1-0.28)^2 + \\ (1-0)^2 + \\ (1-0.52)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 \end{matrix}}$ ≈ 2.18	$D_{G4}^* = \sqrt{\begin{matrix} (1-0.28)^2 + \\ (1-0.20)^2 + \\ (1-0.52)^2 + \\ (1-0)^2 + \\ (1-1)^2 + \\ (1-0.67)^2 \end{matrix}}$ ≈ 1.58	$D_{G5}^* = \sqrt{\begin{matrix} (1-0.28)^2 + \\ (1-0.32)^2 + \\ (1-0.52)^2 + \\ (1-0)^2 + \\ (1-1)^2 + \\ (1-1)^2 \end{matrix}}$ ≈ 1.49
D_i⁻ (Eq. 15)	$D_{G1}^- = \sqrt{\begin{matrix} (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 + \\ (0.33-0)^2 \end{matrix}}$ ≈ 2.26	$D_{G2}^- = \sqrt{\begin{matrix} (0-0)^2 + \\ (0-0)^2 + \\ (0-0)^2 + \\ (0-0)^2 + \\ (1-0)^2 + \\ (0.67-0)^2 \end{matrix}}$ ≈ 1.20	$D_{G3}^- = \sqrt{\begin{matrix} (0.28-0)^2 + \\ (0-0)^2 + \\ (0.52-0)^2 + \\ (0-0)^2 + \\ (0-0)^2 + \\ (0-0)^2 \end{matrix}}$ ≈ 0.59	$D_{G4}^- = \sqrt{\begin{matrix} (0.28-0)^2 + \\ (0.20-0)^2 + \\ (0.52-0)^2 + \\ (0-0)^2 + \\ (1-0)^2 + \\ (0.67-0)^2 \end{matrix}}$ ≈ 1.35	$D_{G5}^- = \sqrt{\begin{matrix} (0.28-0)^2 + \\ (0.32-0)^2 + \\ (0.52-0)^2 + \\ (0-0)^2 + \\ (1-0)^2 + \\ (1-0)^2 \end{matrix}}$ ≈ 1.56
C_i (Eq. 17)	0.77	0.37	0.21	0.46	0.51
Rank	1	4	5	3	2
Initial Mixing Ratio	0.95	0.012	0.0068	0.0148	0.0165
Final Normalised Ratio	0.9660	0.0081	0.0046	0.0101	0.0112

Note: The index *i* denotes the alloy grade (Alloys G1–G5), corresponding to 6040, 6261, 6162, 6262, and 6065, respectively.

The following steps were undertaken:

Step 1: Conversion of Categorical Ratings to Numerical Values – Ratings for corrosion resistance (CR), weldability, and anodising response were converted into numerical values as follows: Excellent = 3, Good = 2, Fair = 1, and Poor = 0.

Step 2: Construction of a Property Matrix – A property matrix was generated for all alloy properties of sub-cluster 01 of Cluster 0, as presented below and summarised in Table 5.2 for clarity.

296	260	270	270	270
46.5	25.2	25.2	29.5	31.9
119.3	100	110	110	110
3	2	2	2	2
3	3	2	3	3
1	2	0	2	3

Step 3: Normalisation of the Property Matrix – All properties were normalised using Equation (14) to ensure that no single property dominated the analysis. Normalised Property matrix:

1	0	0.28	0.28	0.28
1	0	0	0.20	0.32
1	0	0.52	0.52	0.52
1	0	0	0	0
1	1	0	1	1
0.33	0.67	0	0.67	1

After normalisation, each converted property value was indicated within blue brackets, as shown in Table 5.2.

Step 4: Determination of Ideal-Best and Ideal-Worst Solutions – The following criteria were used to identify the ideal-best and ideal-worst solutions for sub-cluster 01, namely:

Ideal-Best Solution T^* :

- Maximum for benefits: Yield strength, modulus of toughness, fatigue strength, weldability, corrosion resistance, and anodising response.
- Minimum for costs: None (all properties are beneficial)
- $T^* = (296, 46.5, 119.3, 3, 3, 3)$

Ideal-Worst Solution T^- :

- Minimum for benefits: Yield strength, modulus of toughness, fatigue strength, weldability, corrosion resistance, and anodising response.
- Costs: None (all properties are beneficial)
- $T^- = (260, 25.2, 100, 2, 2, 0)$

Step 5: Calculation of the Distance of Each Alloy from the Ideal-Best and Ideal-Worst Solutions – For each alloy (G1 through G5), the distance from the ideal-best solution (defined

as having the maximum values for all properties) was calculated using Equation (15), as follows:

$$D_{G1}^* = \sqrt{\frac{(1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2}{(1-0.33)^2}}$$

$$\approx 0.67$$

$$D_{G2}^* = \sqrt{\frac{(1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2}{(1-0.67)^2}}$$

$$\approx 2.01$$

$$D_{G3}^* = \sqrt{\frac{(1-0.28)^2 + (1-0)^2 + (1-0.52)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2}{(1-0)^2}}$$

$$\approx 2.18$$

$$D_{G4}^* = \sqrt{\frac{(1-0.28)^2 + (1-0.20)^2 + (1-0.52)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2}{(1-0.67)^2}}$$

$$\approx 1.58$$

$$D_{G5}^* = \sqrt{\frac{(1-0.28)^2 + (1-0.32)^2 + (1-0.52)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2}{(1-1)^2}}$$

$$\approx 1.49$$

The distance for each alloy from the ideal-worst solution, was calculated using Equation (16) as follows:

$$D_{G1}^- = \sqrt{\frac{(1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2}{(0.33-0)^2}}$$

$$\approx 2.26$$

$$D_{G2}^- = \sqrt{\frac{(0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0.67-0)^2}{(0.67-0)^2}}$$

$$\approx 1.20$$

$$D_{G3}^- = \sqrt{\frac{(0.28-0)^2 + (0-0)^2 + (0.52-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}{(0-0)^2}}$$

$$\approx 0.59$$

$$D_{G4}^- = \sqrt{\frac{(0.28-0)^2 + (0.20-0)^2 + (0.52-0)^2 + (0-0)^2 + (1-0)^2 + (0.67-0)^2}{(0.67-0)^2}}$$

$$\approx 1.35$$

$$D_{G5}^- = \sqrt{\frac{(0.28-0)^2 + (0.32-0)^2 + (0.52-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2}{(1-0)^2}}$$

$$\approx 1.56$$

Step 6: Calculation of the TOPSIS Closeness Score – The TOPSIS closeness score was calculated using Equation (17). The resulting closeness scores for all alloys are as follows:

$$C_{G1} = \frac{2.26}{0.67+2.26} \approx 0.77$$

$$C_{G2} = \frac{1.20}{2.02+1.20} \approx 0.37$$

$$C_{G3} = \frac{0.59}{2.18+0.59} \approx 0.21$$

$$C_{G4} = \frac{1.35}{1.58+1.35} \approx 0.46$$

$$C_{G5} = \frac{1.56}{1.49+1.56} \approx 0.51$$

Step 7: Ranking of the Alloys – The five alloys in sub-cluster 01 of Cluster 0 were ranked based on their TOPSIS closeness scores as follows: Alloy G1 (optimum alloy) > Alloy G5 > Alloy G4 > Alloy G2 > Alloy G3.

According to the TOPSIS analysis, Alloy A emerges as the optimum choice, as shown in Figure 5.10, indicating a balanced performance across all properties compared to the other alloys.

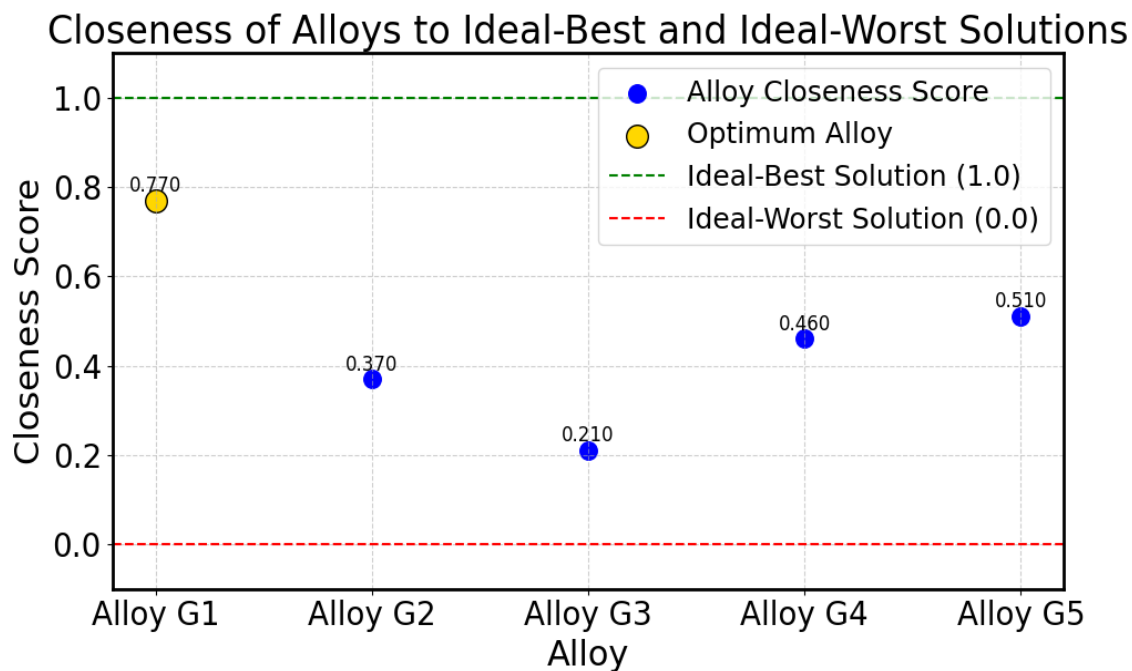


Figure 5.10. TOPSIS closeness scores of alloys

The selection of optimum alloys for the remaining sub-clusters followed the same methodology. The alloys identified as optimal for each sub-cluster and their respective parent clusters are summarised in Table 5.3. Through this optimisation process, the selected alloys were ensured to meet the specified criteria across a broad range of properties and potential applications. It is noteworthy that, for each sub-cluster, the optimum alloy corresponded to a standard alloy, as indicated in Table 5.3.

Table 5.3. Properties and applications of optimum alloys in each cluster

Clusters	Optimum alloys and their properties		Applications
Cluster 0	6064, Sub-cluster 00	Low strength, medium modulus of toughness, and good CR, excellent weldability, and anodising	Automotive braking systems, structural components for civil construction and heavy vehicles, and in the electronics industry [152]

	6040, Sub-cluster 01	Good strength, modulus of toughness, excellent CR, weldability, fair anodising	Automotive, small engine parts, sports products [153], [154]
Cluster 3	6055, Sub-cluster 30	High strength, medium modulus of toughness, fair weldability, good anodising and CR	Automotive, construction, and electronics industries [154]
	6070, Sub-cluster 31	High strength, modulus of toughness, weldability, moderate CR, anodising	Pipelines and heavy duty welded structures [155]
Cluster 2	6063, Sub-cluster 20	Low strength, modulus of toughness, excellent CR, weldability and anodising response	Transportation, architectural and industrial applications [156]
	6463, Sub-cluster 21	Moderate strength, modulus of toughness, and excellent CR, weldability, anodising response	Manufacturing extruded architectural and trim sections [157]
Cluster 4	6092, Sub-cluster 40	High strength, medium modulus of toughness, good CR and anodising, excellent weldability	Aerospace components and industrial wear parts [158]
	6011A, Sub-cluster 41	High strength, modulus of toughness, moderate CR, weldability, anodising	Automobile vehicles [159]
Cluster 1	6082, Sub-cluster 10	Medium strength, high modulus of toughness, excellent CR, good weldability, and anodising	High stress applications [160], [161]
	6009, Sub-cluster 11	Low strength, modulus of toughness, good strength, good CR, weldability and anodising response	Components of car body, such as door panels, roofs, front and rear bumpers, side skirts, and wheel arches [162], [163]
	6005, Sub-cluster 13	Medium strength, high modulus of toughness, good CR and anodising response, excellent weldability	Extrusions for tubing in furniture, railway and bus structures, pylons, platforms, and pipelines, as well as for ladder structures and components in trucks, trailers, and automotive vehicles [164]

5.6 Alloy Mixing within Sub-clusters: Recycling Algorithm

After selecting the optimum alloys using the optimisation algorithm, we proceeded with alloy mixing within the sub-clusters using the designed recycling algorithm as described in **Section 3.7**. This approach enables the prediction of desired property combinations by blending the optimum alloys with the remaining alloys in each sub-cluster. For example, consider sub-cluster 01 of Cluster 0:

As shown in Table 5.2, Alloy G1 was identified as the optimum alloy in sub-cluster 01, with a closeness score of 0.77. The remaining alloys in the sub-cluster had closeness scores of 0.37, 0.21, 0.46, and 0.51, respectively. The process proceeds as follows:

Step 1: Optimum Alloy Ratio – Alloy G1 (with a closeness score of 0.77) was assigned a minimum of 50% of the mixing ratio; in present case, it received 95%.

Step 2: Remaining Mixing Ratio – The remaining 50%, or less (here 5%), was distributed among the other alloys as shown in Table 5.2. The ratios for these alloys were calculated based on their TOPSIS closeness scores, (See **Section 5.5**), normalised so that the total sum of all mixing ratios equals 100%.

For instance, if the sum of the closeness scores of the remaining alloys ($0.37 + 0.21 + 0.46 + 0.51$) equals 1.55, the individual ratios for the alloys were calculated using Equation (25) as follows:

$$\text{Alloy G2 ratio} = (0.37/1.55) \times 0.05 \approx 0.012$$

$$\text{Alloy G3 ratio} = (0.21/1.55) \times 0.05 \approx 0.0068$$

$$\text{Alloy G4 ratio} = (0.46/1.55) \times 0.05 \approx 0.0148$$

$$\text{Alloy G5 ratio} = (0.51/1.55) \times 0.05 \approx 0.0165$$

Step 3: Normalisation of the Ratios – The final ratios were normalised so that their total sum equals 1, using Equation (26). This resulted in the following normalised mixing ratios: Alloy G1: 0.9660, Alloy G2: 0.0081, Alloy G3: 0.0046, Alloy G4: 0.0101, and Alloy G5: 0.0112.

Step 4: Application of the Mixing Ratios and Calculation of Final Composition and Property Values – The composition of the mixed alloy was calculated using Equation (27) as a weighted average of the individual alloys based on their respective mixing ratios. The final recycled optimum alloy composition, including Si, Fe, Mn, Mg, and Cu, was determined as follows:

$$\begin{aligned} \text{Final Si} &= (0.60 \times 0.9660) + (0.55 \times 0.0081) + (0.60 \times 0.0046) + (0.60 \times 0.0101) + (0.60 \times 0.0112) \\ &= 0.5997 \end{aligned}$$

$$\text{Final Fe} = (0.35 \times 0.9660) + (0.20 \times 0.0081) + (0.25 \times 0.0046) + (0.35 \times 0.0101) + (0.37 \times 0.0112) \\ = 0.3485$$

$$\text{Final Mn} = (0.15 \times 0.9660) + (0.275 \times 0.0081) + (0.10 \times 0.0046) + (0.15 \times 0.0101) + \\ (0.15 \times 0.0112) = 0.1508$$

$$\text{Final Mg} = (1.00 \times 0.9660) + (0.85 \times 0.0081) + (0.90 \times 0.0046) + (1.00 \times 0.0101) + (1.00 \times 0.0112) \\ = 0.9983$$

$$\text{Final Cu} = (0.50 \times 0.9660) + (0.275 \times 0.0081) + (0.20 \times 0.0046) + (0.275 \times 0.0101) + \\ (0.275 \times 0.0112) = 0.4920$$

Weighted average for each property – The weighted average of each property was calculated using Equation (27), in the same manner as for the alloy compositions, as follows:

$$\text{Final Yield Strength} = (296 \times 0.9660) + (260 \times 0.0081) + (270 \times 0.0046) + (270 \times 0.0101) + \\ (270 \times 0.0112) = 295.04 \text{ MPa}$$

$$\text{Final Modulus of Toughness} = (46.5 \times 0.9660) + (25.2 \times 0.0081) + (25.2 \times 0.0046) + \\ (29.5 \times 0.0101) + (31.9 \times 0.0112) = 45.86 \text{ MPa \%}$$

$$\text{Final Fatigue Strength} = (119.3 \times 0.9660) + (100 \times 0.0081) + (110 \times 0.0046) + (110 \times 0.0101) + \\ (110 \times 0.0112) = 118.91 \text{ MPa}$$

$$\text{Final Corrosion Resistance} = (3 \times 0.9660) + (2 \times 0.0081) + (2 \times 0.0046) + (2 \times 0.0101) + \\ (2 \times 0.0112) = 2.966 \rightarrow \text{Very close to "Excellent"}$$

$$\text{Final Weldability} = (3 \times 0.9660) + (3 \times 0.0081) + (2 \times 0.0046) + (3 \times 0.0101) + (3 \times 0.0112) = \\ 2.995 \rightarrow \text{"Excellent"}$$

$$\text{Final Anodising} = (1 \times 0.9660) + (2 \times 0.0081) + (0 \times 0.0046) + (2 \times 0.0101) + (3 \times 0.0112) = 1.036 \\ \rightarrow \text{Just above "Fair"}$$

The resulting recycled alloy composition and corresponding weighted properties are summarised in Table 5.4.

Table 5.4. Final recycled alloy composition and properties for sub-cluster 01

Parameter	Final Value
Chemical Composition	
Si	0.5997
Fe	0.3485
Mn	0.1508
Mg	0.9983
Cu	0.4920
Mechanical Properties	
Yield Strength, MPa	295.04
Modulus of Toughness, MPa%	45.86
Fatigue Strength, MPa	118.91
Technological Properties	
Corrosion Resistance	≈ Excellent
Weldability	≈ Excellent
Anodising Response	Fair

This mixing resulted in a recycled alloy that was closest to the optimum alloy selected by the optimisation algorithm, exhibiting nearly identical composition and only a minor deviation of 1 MPa in yield strength. All other properties remained unchanged. Similarly, the mixing ratios for the other sub-clusters of alloys can be calculated using the same algorithm, which will likewise produce the optimum alloys corresponding to their respective sub-clusters as determined by the optimisation algorithm.

5.7 Validation of Mixed Alloy Composition

By applying the recycling algorithm to each sub-cluster, mixed alloy compositions were generated, ensuring that the final mixed alloy met the desired performance criteria and closely resembled the optimum alloy identified by the optimisation algorithm. It is noteworthy that each optimum alloy corresponds to a standard alloy grade, which facilitates validation, as the properties of standard alloys are well documented. Table 5.5 presents a comparison of the predicted optimum alloy from sub-cluster 01 of Cluster 0 with the measured property values obtained from the MatWeb database [165]. It should be noted that the MatWeb database was not used during data collection and was entirely unseen, serving only for validation. The comparison reveals only minor differences between the two datasets. In this instance, Alloy G1, corresponds to grade 6040.

Table 5.5. Comparison of predicted vs. reported properties of optimum alloys for Sub-cluster 01

Property	Alloy G1 (Predicted)	Alloy G1 (Reported)
Yield strength	295	296
Elongation	15	15
Fatigue strength	119	119
Corrosion resistance	Excellent	Excellent
Weldability	Excellent	Excellent
Anodising	Good	Excellent

5.8 Understanding the Clustering Process: LIME Algorithm

As the model increases in complexity, its predictive accuracy improves; however, interpretability becomes more challenging. For this reason, clustering techniques are often regarded as “black box” algorithms. To ensure model reliability, it is important to assess how the model operates. Therefore, an explainable artificial intelligence (XAI) approach should be applied, with emphasis on the selected factors, or compositions, that significantly influence cluster formation. A black box clustering model, whose internal reasoning is obscured and difficult to understand, can be elucidated using the LIME algorithm (see **Sections 2.2.3.9** and **3.8**).

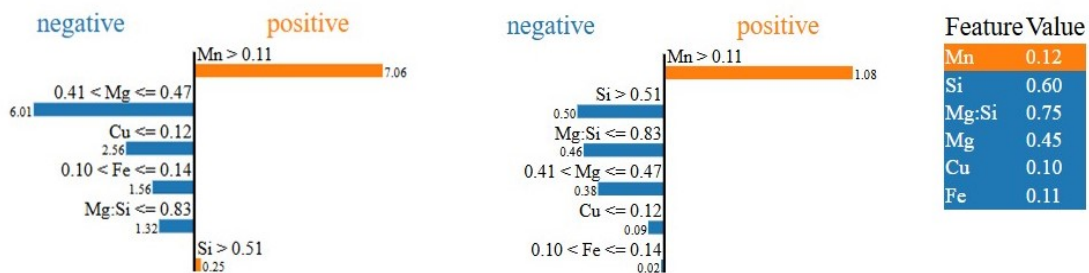
As described in **Section 5.3**, the clusters containing alloys with varying ranges of properties and compositions were predicted using the combined PCA and K-means clustering algorithm. To interpret the contribution of each alloying element to the properties, an additional step was necessary, as clustering alone cannot distinguish between independent and dependent variables. Furthermore, the predicted results required verification against established scientific knowledge.

The clusters obtained from the PCA and K-means clustering algorithm were interpreted using the LIME algorithm. For each cluster, a separate dataset was prepared. The LIME algorithm partitioned the dataset of alloy features into training and test sets and evaluated the prediction of target properties, such as yield strength (YS) and modulus of toughness, as shown in Figure 5.11. The figure highlights which alloying elements exert positive or negative effects on YS and modulus of toughness.

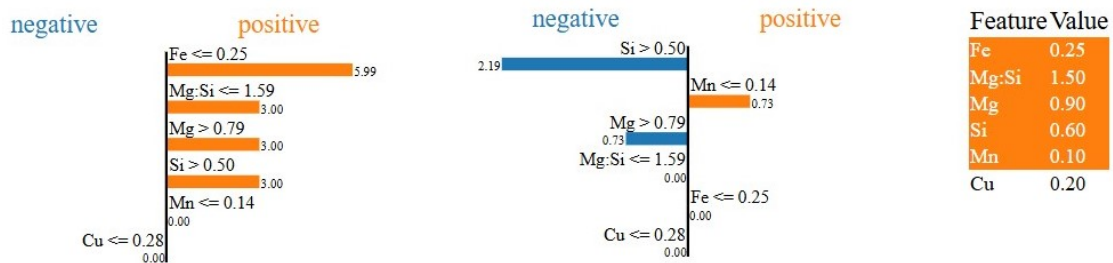
For Cluster 2, all alloying elements except Mn, and Si showing a slight positive effect, had a negative influence on YS, resulting in the lowest YS values among the clusters. In contrast,

almost all alloying elements except Mn had a positive effect on the modulus of toughness. However, because the concentrations of these elements were low in this cluster, their overall contribution was limited, and the modulus of toughness remained low.

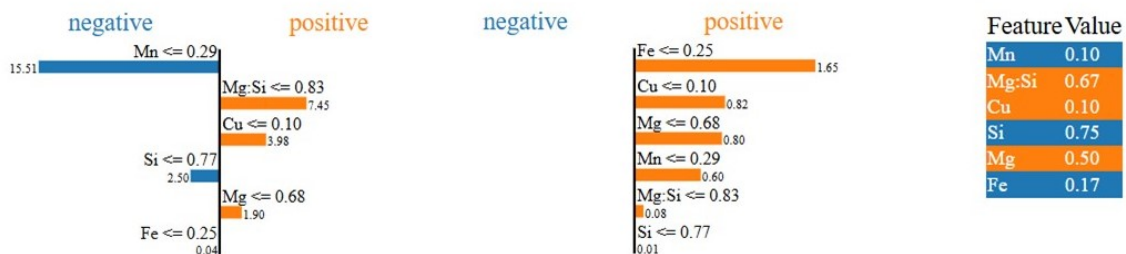
In Cluster 4, which exhibited the highest YS and a moderate modulus of toughness, the alloys contained high concentrations of Cu and Si. These elements primarily contributed to the increase in YS while simultaneously reducing the modulus of toughness. However, Mn had a strong positive effect on the modulus of toughness, which helped maintain it at a moderate level.



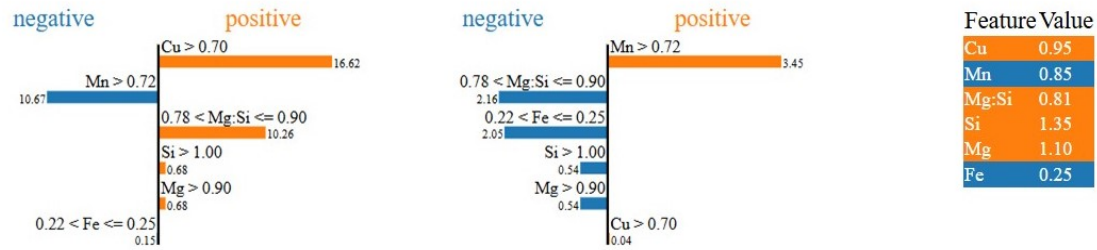
(a) Cluster 2, Yield strength and Modulus of Toughness



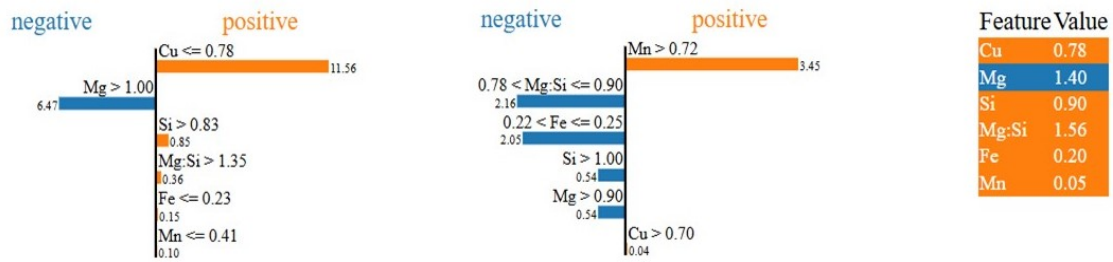
(b) Cluster 0, Yield strength and Modulus of Toughness



(c) Cluster 1, Yield strength and Modulus of Toughness



(d) Cluster 3, Yield strength and Modulus of Toughness



(e) Cluster 4, Yield strength and Modulus of Toughness

Figure 5.11. LIME plots of YS and modulus of toughness: (a) Cluster 2, (b) Cluster 0, (c) Cluster 1, (d) Cluster 3, and (e) Cluster 4

5.9 Metallurgical Reasoning behind Clustering

The clustering by ML methods will only be reliable if it has some metallurgical/physical meaning. The general metallurgical reasoning behind clusters obtained by the combined approach of PCA and K-means clustering, as well as the rationale behind optimum alloy selection, is discussed here.

According to the Al-Mg-Si system phase diagram, the precipitation events in ternary 6xxx series alloys proceed as follows: supersaturated solid solution (SSS) $\rightarrow$ GP zones $\rightarrow$ β'' $\rightarrow$ β' $\rightarrow$ β [149], [166]. β'' is the main hardening phase in 6xxx series aluminium alloys. An addition of Cu to Al-Mg-Si alloys introduces the possibility of forming a quaternary phase, Q, with some metastable variations, changing the kinetics and phase composition of precipitates, which is briefly outlined below for some ranges of alloying elements (with hardening phases in bold) [166], [166]:

At a lower Cu concentration: SSS – **GP zones – β'' (a modification with Cu)** – Si – numerous variations of β' (with Cu) including $\beta'C$ – Mg_2Si and Si.

At a higher Cu concentration: SSS – **GP zones – β'' (modifications with Cu) – θ' or/and Q'** – Si – numerous variations of β' (with Cu), including $\beta'C$, Q' , and θ' – Q (AlMgSiCu), Mg_2Si , Al_2Cu and Si.

These precipitation paths may also occur at intermediate copper concentrations during artificial ageing.

Based on this general understanding of the metallurgy of 6xxx series alloys, we can suggest the following metallurgical reasons behind grouping the alloys into high, medium, and low property clusters:

Cluster 4 (High Strength Alloys): Composition range (wt%): Si: (1.0-1.35), Fe: (0.1-0.35), Cu: (0.275-0.95), Mn: (0.55-0.85), Mg: (0.65-1.1) leads to the following phase formation: large number of β'' , increased number of Q' , Si, and β' .

High Cu content, ranging from 0.275% to 0.95%, along with high levels of Mg, Mn, and Si, contributes to the accelerated precipitation kinetics of strengthening precipitates such as Cu-modified β'' , and Q' [167], [168], [169], [170], [171], [172]. High intergranular corrosion (IGC) is due to excess silicon (i.e., low Mg:Si ratio), and AlFeSi (Mn, Cu) dispersoids formed at grain boundaries, creating a continuous corrosion channel by connecting with the grain-boundary Q' and Q phases [173], [174], [175], [176], [177]. Weldability is moderate due to higher liquation cracking caused by high Cu content (i.e., >0.8) and grain boundary segregation due to Si precipitates [174], [178], [179], [180], [174]. The anodising response is moderate due to the high presence of Si and Mg_2Si constitutive particles, significantly affecting surface hardness and gloss [176], [181], [178].

Cluster 0 (Medium Strength Alloys): Composition range (wt%): Si: (0.35-0.6), Fe: (0.175-0.4), Cu: (0.1-0.5), Mn: (0-0.275), Mg: (0.6-1.25) determines the formation of the following phases: medium amount of β'' , Q' , β' .

At a medium Cu content of 0.275% and with a small excess of Si as Mg:Si~1.67 (1.73 is the Mg_2Si stoichiometry), the main hardening phase is β'' and may be assisted at later ageing stages by the Q' phase [167], [168]. Corrosion resistance is rated as good to excellent, primarily attributable to the relatively high Mg:Si ratio, which influences the type, size, and distribution of grain boundary precipitates (GBPs), essential for combating IGC resistance [173], [174].

Weldability is excellent due to the low presence of liquation cracking-causing elements, i.e., Cu, and the moderate formation of Mg_2Si phases [178], [181], [174]. The anodising response is mostly excellent due to the moderate amount of constituent Si and Mg_2Si particles, resulting in softer surfaces and increased surface gloss [181], [178].

Cluster 2 (Lower Strength Alloys): Composition range (wt%): Si: (0.4-0.6), Fe: (0.08-0.18), Cu: (0.05-0.25), Mn: (0.05-0.13), Mg: (0.35-0.68) leads to the following phase formation: lower amount of β'' , Q', β' .

Due to the low concentration of solute elements, i.e., Cu, Si, and Mg, the number density of the hardening β'' and Q' phases are lower, leading to a lower strength [164], [173], [174]. On the other hand, corrosion resistance is high owing to magnesium excess (i.e., high Mg:Si), which improves the intergranular corrosion (IGC) resistance of alloys by forming discontinuous precipitates on the grain boundaries [149], [166]. Weldability is excellent due to low liquation cracking and shorter crack length caused by low copper content [167], [172], [175]. The anodising response is good due to lower surface hardness and low iron content, thereby reducing its effect on texture [176], [182], [178].

5.10 Thermo-Calc Analysis as a Basis for Cluster Classification

The clustering interpretations are further supported by Thermo-Calc single-point equilibrium analyses [183] with details of phase formation provided in Appendix **Section 11.3.2.2**. Under T6 tempering conditions at atmospheric pressure, the simulations revealed phase formations similar to those discussed in **Section 5.9**, across clusters representing low-, medium-, and high-property alloys, as illustrated in Figure 5.12. For ease of visualisation, the plotting in this figure was performed by averaging the formation of strengthening phases and the corresponding yield strength within each cluster.

As shown, Cluster 4 exhibits the highest volume fraction of Mg_2Si (β'') phase formation, together with a notable fraction of the Q phase. The combined presence of these strengthening phases contributes substantially to its high yield strength. By contrast, Cluster 2 displays a markedly lower yield strength, which can be attributed to the limited formation of both Mg_2Si and Q phases.

These results highlight the strong consistency between computational predictions and the literature-based interpretations.

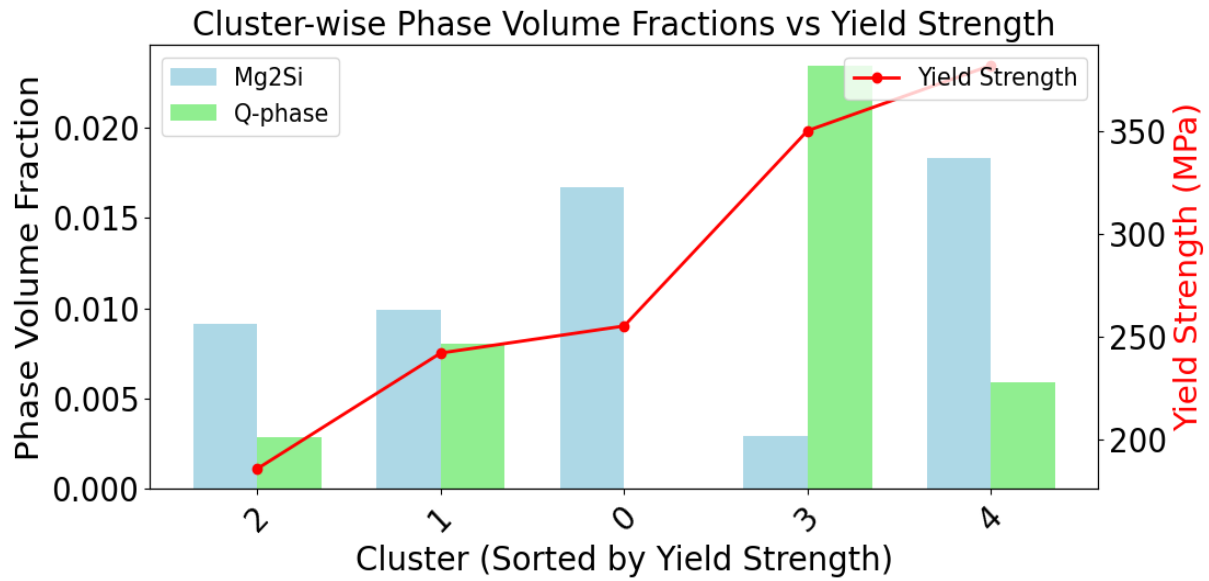


Figure 5.12. Volume phase formation within all cluster of 6xxx series

5.11 Conclusion on 6xxx Series Optimisation

This chapter presented a data-driven methodology to optimise 6xxx series aluminium alloys under T5, T6, and T7 tempering conditions, as well as processing routes such as extrusion, sheet, rolling, forging, and plate production. Starting from 279 datasets for 41 alloys, PCA and K-means clustering were applied to group the alloys into five distinct clusters representing low-, medium-, and high-property ranges. Each cluster was subsequently subdivided into more refined sub-clusters based on property proximity, resulting in a total of 11 sub-clusters.

Within each sub-cluster, the optimisation algorithm was employed to select the optimum alloy covering the broad range of properties and applications, based on mechanical, corrosion, and weldability characteristics. This process reduced the initial set of 41 alloys to 11 best-performing alloys, including 6005, 6082, and 6011, representing medium and high-performance clusters, respectively. To ensure that the optimum alloy received at least a 50% mixing ratio, a recycling algorithm was implemented within each sub-cluster. The remaining 50% of the mixing ratio was distributed among the other alloys within the sub-cluster proportionally according to their TOPSIS closeness scores. The resulting mixed alloys closely matched the optimum alloy and corresponded to standard alloy grades, with deviations of only 1 MPa in yield and fatigue strength and 0.001% in composition.

Each cluster was further analysed using LIME-based explainable AI, confirming that Mg, Cu, and Si exert a significant influence on key mechanical properties, thereby demonstrating a direct relationship between composition and alloy performance. Metallurgical analysis additionally revealed that clusters 3 and 4, which exhibited higher solute content (Cu + Mg + Si) and optimal Mg:Si ratios, achieved enhanced strength due to increased β'' phase precipitation. This was corroborated by Thermo-Calc single-point analysis under T6 conditions, which showed that higher volume fractions of these phases resulted in increased yield strength in the respective clusters.

This integrated framework not only simplifies alloy selection and classification but also ensures minimal property loss and improved recyclability. It provides a practical approach for industrial applications where alloy performance and sustainability are of equal importance.

CHAPTER 6: RESULTS AND DISCUSSION: OPTIMISATION OF 7xxx-SERIES ALUMINIUM ALLOYS

The aluminium alloys in the 7xxx series were optimised using the same methodology applied to the 6xxx series dataset. This systematic approach proved equally effective for the 7xxx series. Further details are not elaborated here, as they have already been discussed in **Chapter 5: Results and Discussion – Optimisation of 6xxx series Aluminium Alloys**. The key distinction in this case was that the 7xxx series dataset was initially divided into two groups: low Cu-containing alloys and high Cu-containing alloys. Following this initial division, a combined PCA and K-means clustering technique (**Section 3.4**), was employed to identify groupings within each subset. The subsequent optimisation steps, including sub-clustering (**Section 3.5**), the optimisation algorithm, (**Section 3.6**), and the recycling algorithm (**Section 3.7**), were carried out in the same manner as for the 6xxx series. The detailed procedures for this optimisation process are presented in the following sections.

6.1 Data Collection

The collected dataset comprises 78 alloy compositions, with properties recorded under T5, T6, and T7 tempering conditions (see **Section 2.1.5** for details). To facilitate analysis, the dataset was divided into two subsets based on copper content: 24 alloys with low-Cu and 54 alloys containing high-Cu. Copper is a critical alloying element in this series, as it significantly influences strength, corrosion resistance, and weldability, thereby determining the alloys' potential applications.

Low-Cu Alloys: These alloys exhibit medium-range strength characteristics, reflecting a distinct balance between mechanical performance, corrosion behaviour, and weldability. Their composition ranges are: $4 \leq \text{Zn} \leq 7.8 \%$, $0.02 \leq \text{Cu} \leq 0.3 \%$, and $0.75 \leq \text{Mg} \leq 3 \%$. The content of other alloying elements: $0.05 \leq \text{Fe} \leq 0.25 \%$, $0.04 \leq \text{Mn} \leq 0.7 \%$, and $0.1 \leq \text{Si} \leq 0.4 \%$, $0 \leq \text{Cr} \leq 0.175 \%$, $0.015 \leq \text{Ti} \leq 0.125 \%$, and $0 \leq \text{Zr} \leq 0.125 \%$.

These alloys exhibit properties in the following ranges: $290 \leq \text{YS} \leq 500 \text{ MPa}$, $120 \leq \text{Fatigue strength (FS)} \leq 180 \text{ MPa}$, and $31.07 \leq \text{Modulus of Toughness (MoT)} \leq 51.1 \text{ MPa } \%$. Corrosion resistance and weldability ratings span from poor to excellent.

High-Cu Alloys: These alloys exhibit higher strength characteristics, with composition ranges of $4.5 \leq \text{Zn} \leq 11.5 \%$, $1.1 \leq \text{Mg} \leq 3.15 \%$, and $0.5 \leq \text{Cu} \leq 2.4 \%$. Other elements fall within narrower ranges: $0.04 \leq \text{Fe} \leq 0.3 \%$, $0 \leq \text{Mn} \leq 0.8 \%$, $0.06 \leq \text{Si} \leq 0.5 \%$, $0 \leq \text{Cr} \leq 0.175 \%$, $0 \leq \text{Ti} \leq 0.1 \%$, and $0 \leq \text{Zr} \leq 0.3 \%$.

The corresponding properties fall within the following ranges: $378 \leq \text{YS} \leq 730 \text{ MPa}$, $120 \leq \text{FS} \leq 210 \text{ MPa}$, and $22.4 \leq \text{MoT} \leq 61.05 \text{ MPa } \%$. They represent the highest-strength aluminium alloys in the dataset, with corrosion resistance and weldability ratings ranging from poor to excellent.

By categorising the alloy dataset into low-Cu and high-Cu sub-sets, the structured data collection process facilitated a clearer understanding of Cu's role in determining the strength and other mechanical properties of the 7xxx series alloys. This approach establishes a foundation for further optimisation, allowing alloys to be selected according to specific performance criteria.

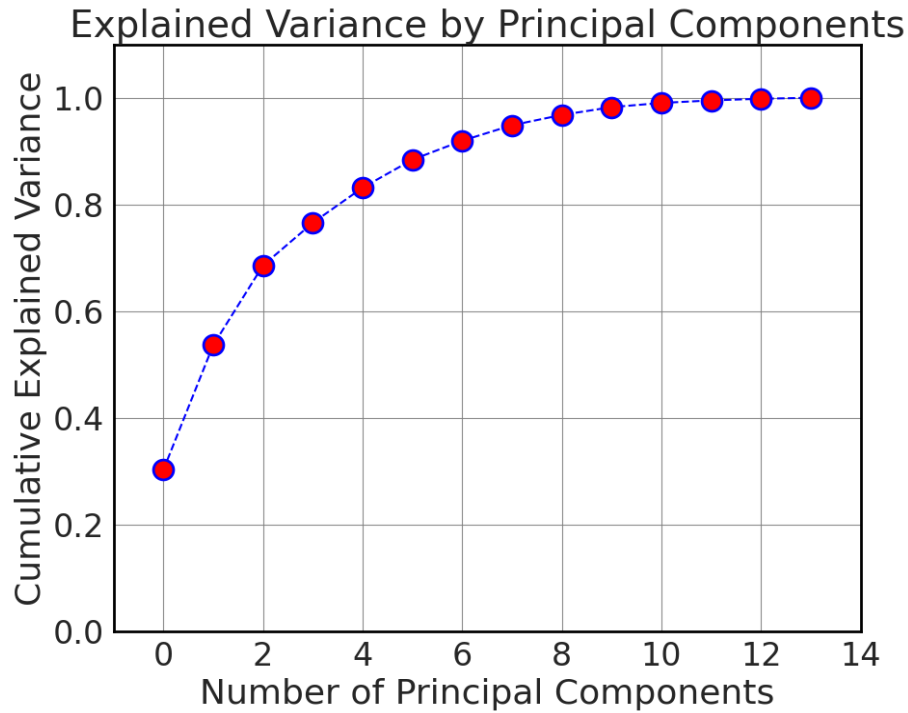
For transparency and reproducibility, the complete raw datasets, processed datasets at each step, and reproducibility notes are provided in **Appendix Sections 11.3.5.1, 11.3.5.2, and 11.3.5.3, 11.3.4.3**, respectively. All subsequent analyses in this chapter are based on these datasets.

6.2 Optimisation of Low-Cu 7xxx Series Aluminium Alloys

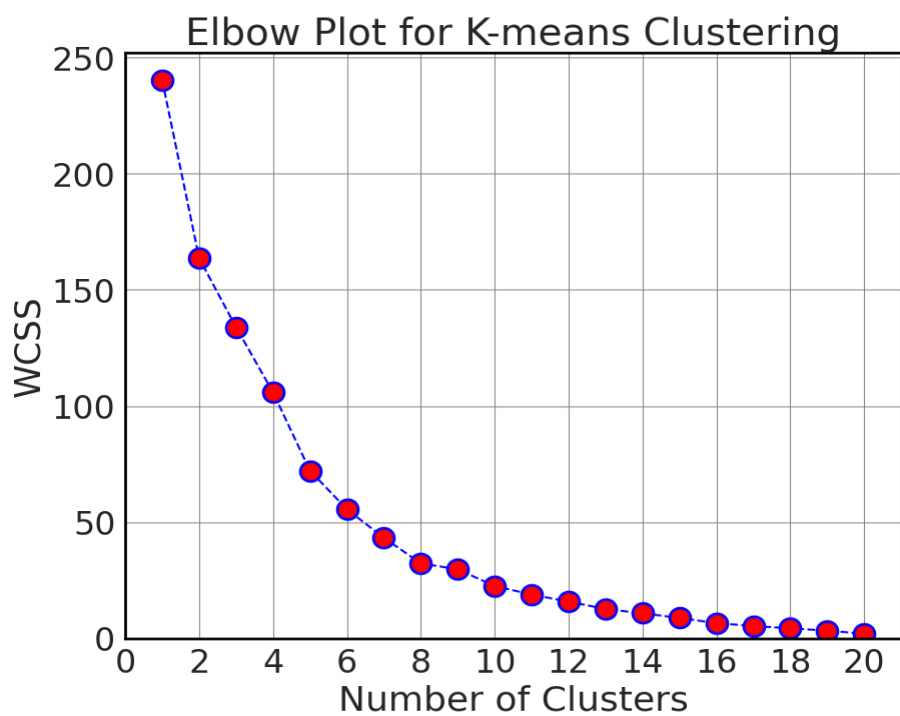
The full optimisation workflow, encompassing PCA, K-means clustering (**Section 3.4**), sub-clustering (**Section 3.5**), optimisation algorithm (**Section 3.6**), and the recycling algorithm (**Section 3.7**), was applied to the low-Cu 7xxx series alloys. The dataset was standardised to ensure all properties were expressed on a consistent scale. Subsequent clustering identified distinct groups, which were further refined into sub-clusters to capture narrower ranges of compositional and mechanical characteristics. Each sub-cluster was then subjected to optimisation algorithm to select the most representative optimum alloy, and the recycling algorithm was employed to predict mixed-alloy compositions that closely approximated the optimum alloy properties.

6.2.1 Clustering of Low Cu 7xxx Series Alloys

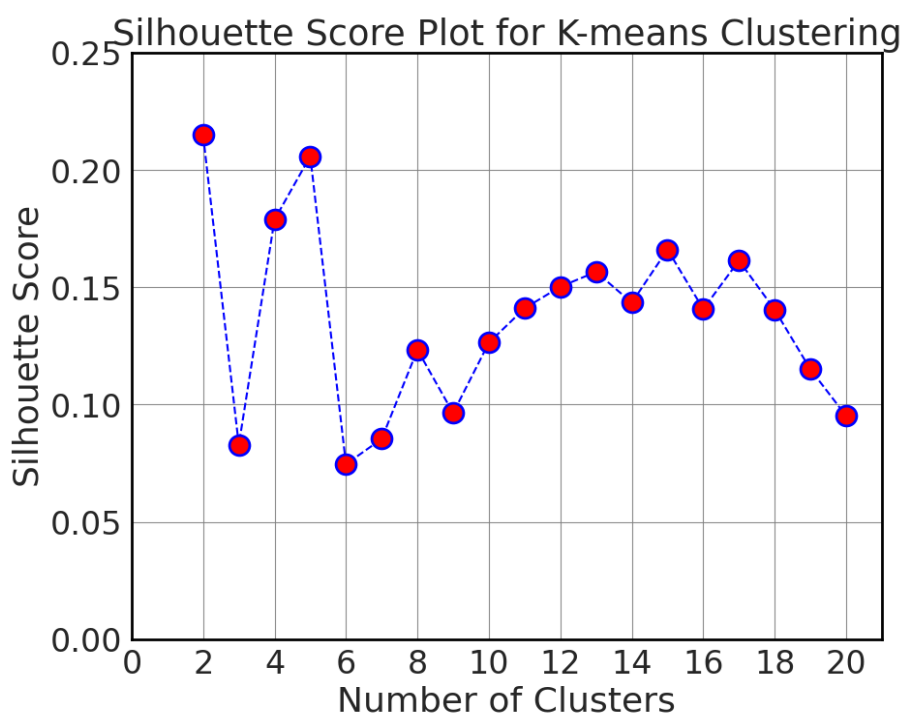
The dataset for the low-Cu 7xxx series alloys was standardised to ensure all properties were expressed on a consistent scale. Principal Component Analysis (PCA) was then performed to identify the number of components required to effectively represent the dataset (see **Section 2.2.3.6**). As shown in Figure 6.1(a), four principal components captured over 75% of the total variance, indicating that the dataset could be reliably represented in a reduced-dimensional space for subsequent clustering analyses.



(a)



(b)

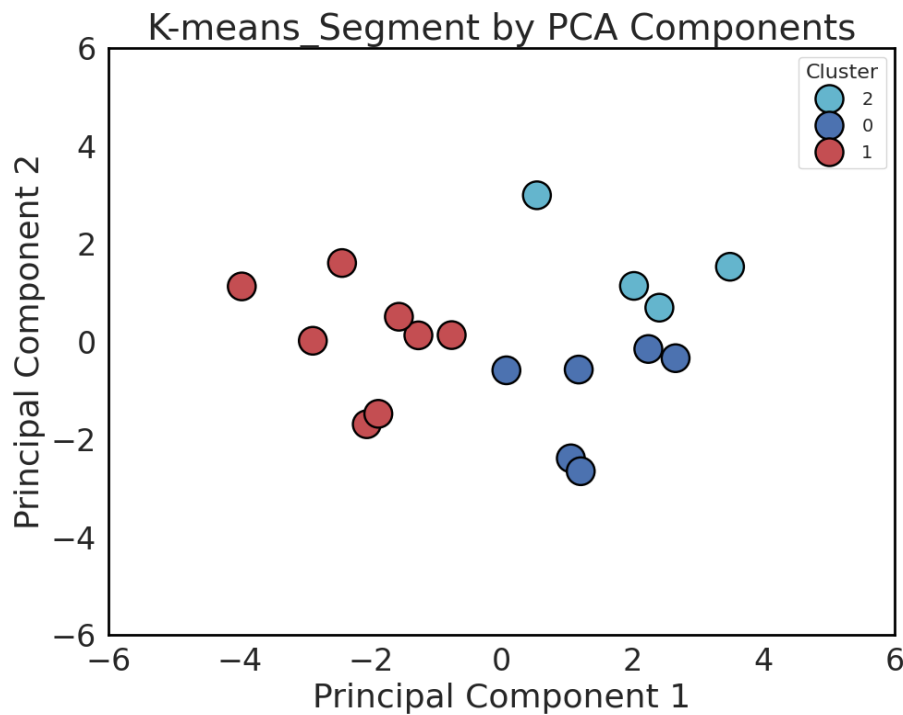


(c)

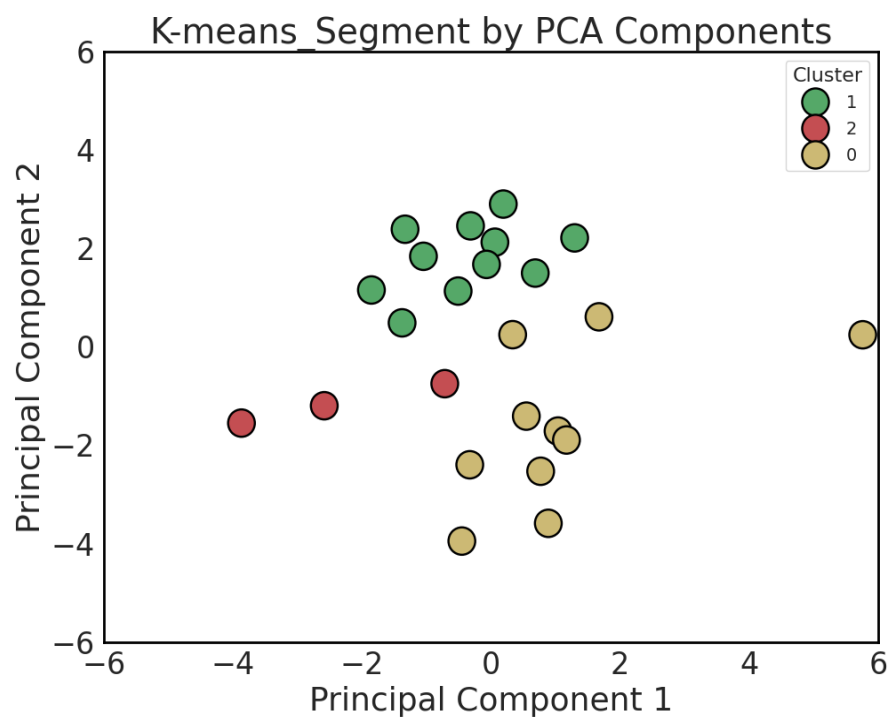
Figure 6.1. A PCA variance plot (a), Elbow plot (b), and Silhouette plot (c), for determining the optimal number of clusters in low-Cu 7xxx series alloys

Following PCA, the number of clusters in the dataset was determined. Although the silhouette scores (Figure 6.1(c)) were relatively low (0.1), the elbow plot (kink at points, $K=2$ and 3) in Figure 6.1(b), (as described in **Section 2.2.3.7.1**), suggested three well-separated clusters. After careful analysis, $K=3$ was selected, yielding reasonably well-defined clusters. The clustering results under the T6 tempering condition were visualised in Figure 6.2(b), showing distinct clusters with minimal overlap between groups.

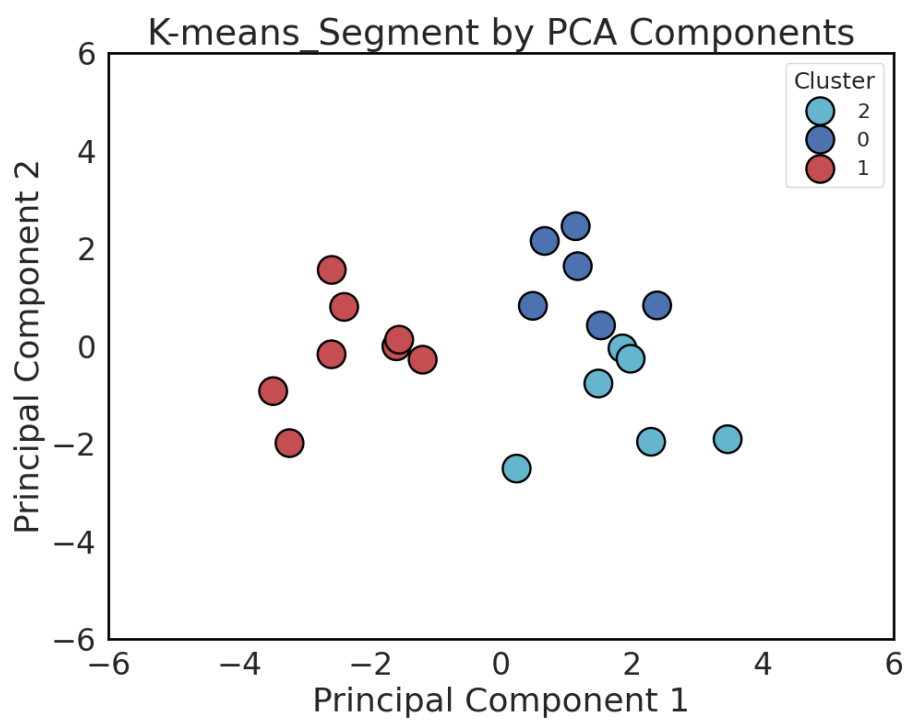
These results indicate that the alloys form distinct groupings, with alloys of each cluster displaying similar ranges of composition and properties. Comparable clustering patterns were observed under the T5 and T7 tempering conditions (Figures 6.2(a) and 6.2(c), respectively), further confirming the robustness of the clusters across different tempering conditions.



(a) T5



(b) T6



(c) T7

Figure 6.2. Clustering of low Cu 7xxx series aluminium alloys at (a) T5, (b) T6, and (c) T7 tempering condition.

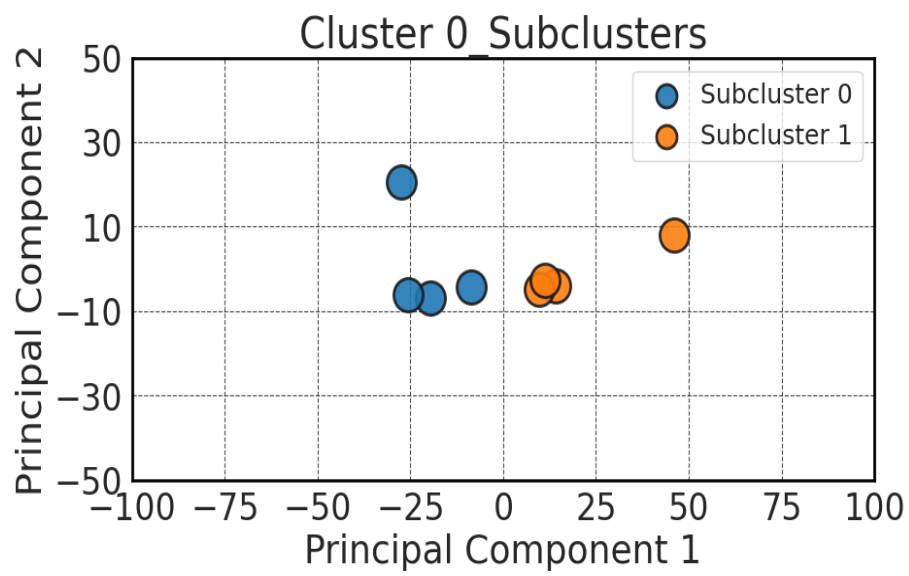
The variations in properties and composition ranges across the clusters are summarised in Table 6.1. Cluster 1 exhibits high yield strength, attributable to its elevated solute content, particularly Mg + Zn. Cluster 0 demonstrates medium yield strength, reflecting a more balanced solute content and a compromise between mechanical performance and corrosion resistance. In contrast, Cluster 2 displays comparatively lower yield strength, which can be linked to its reduced solute content and the resulting decline in mechanical performance. These trends confirm that the clustering algorithm successfully distinguished the alloys according to variations in composition and their associated property ranges.

Table 6.1. Range of properties and compositions in low Cu alloys

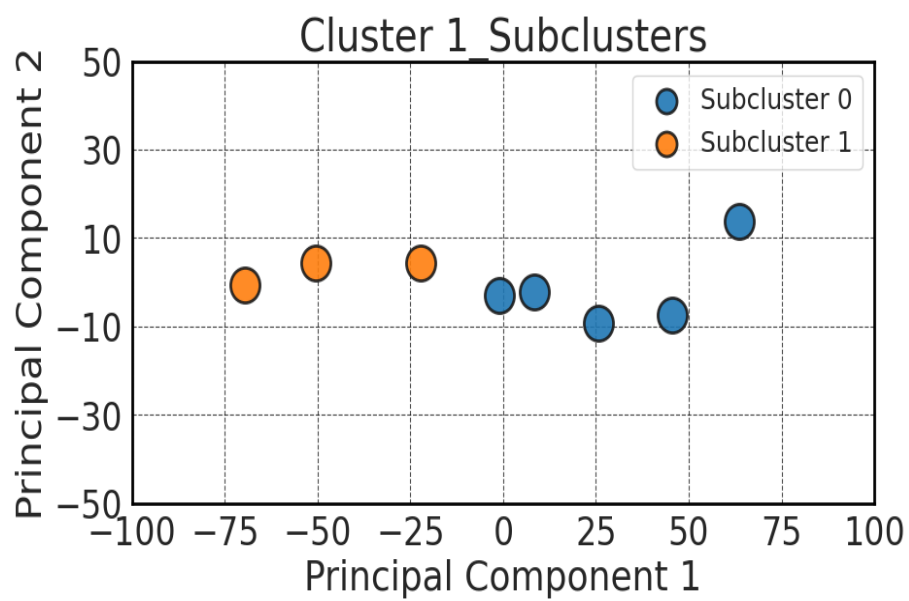
Property	Cluster 0 (Medium)	Cluster 1 (High)	Cluster 2 (Low)
Zn, wt%	4.0-4.9	4.9-7.8	4.0-5.75
Mg, wt%	1.15-2.8	1.2-3	0.75-1.5
Cu, wt%	0.025-0.2	0.02-0.3	0.025-0.1
Fe, wt%	0.05-0.25	0.075-0.25	0.05-0.2
Mn, wt%	0.1-0.7	0.04-0.5	0.05-0.7
Si, wt%	0.2-0.35	0.12-0.4	0.1-0.35
Cr, wt%	0.025-0.175	0.02-0.125	0-0.175
Ti, wt%	0.025-0.075	0.015-0.125	0.015-0.125
Zr, wt%	0-0.125	0.015-0.125	0-0.125
Yield Strength, MPa	330-403	372-500	290-326
Fatigue Strength MPa	135-160	143-180	120-135
Modulus of toughness, MPa %	35.3-51.1	31.1-46.3	33.6-42.3
Corrosion	Poor-Excellent	Poor- Good	Poor-Excellent
Weldability	Good-Excellent	Fair-Excellent	Excellent
Zn:Mg Ratio	1.5-3.5	1.63-5.58	4.75-6.5
Zn+Mg	5.15-7.1	6.6-10.4	3-7.6

6.2.2 Sub-clustering of Clusters

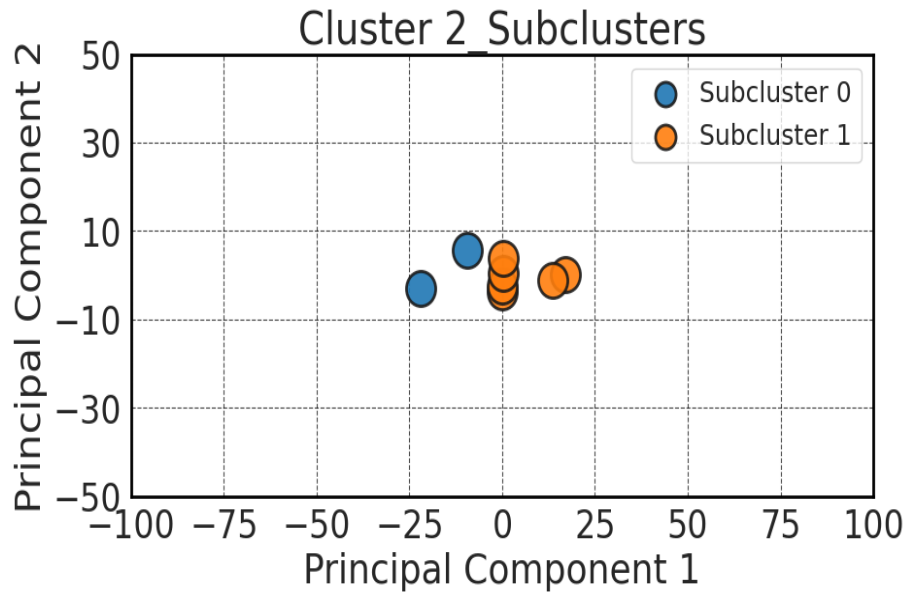
To refine the classification beyond the broader clusters, sub-clustering was performed (see **Section 3.5**), resulting in six distinct sub-clusters, as illustrated in Figure 6.3. This additional step allowed for a more compact grouping of alloys with narrower property ranges, thereby providing a finer range of alloy characteristics. For example, Sub-cluster 01 of Cluster 0 comprises alloys with moderate strength and good weldability but moderate corrosion resistance, whereas Sub-cluster 00 contains alloys with high strength and excellent weldability.



(a) Cluster 0, sub-clusters



(b) Cluster 1, sub-clusters



(c) Cluster 2, sub-clusters

Figure 6.3. Sub-clusters of low-Cu 7xxx series alloy: (a) Cluster 0, sub-clusters, and (b) Cluster 1, sub-clusters, and (c) Cluster 2, sub-clusters

6.2.3 Optimum Alloy Selection Algorithm

After identifying sub-clusters with narrower property ranges within the low-Cu alloy sub-set, the next step involved optimising each sub-cluster to identify the optimum alloys (see **Section 3.6**). These alloys represent the most balanced combination of properties within its respective sub-cluster. For instance, Sub-cluster 01 of Cluster 0 comprises four alloys with varying property ranges, as summarised in Table 6.2.

Table 6.2. Actual and normalised (in parentheses) values of properties of low-Cu alloys within Sub-cluster 01 of Cluster 0

Property	Alloy B1 (7039)	Alloy B2 (7004)	Alloy B3 (7025)	Alloy B4 (7015)
Yield Strength (MPa)	330 (0.4902)	337 (0.5006)	331 (0.4917)	348 (0.5170)
Fatigue Strength (MPa)	160 (0.5609)	135 (0.4733)	136 (0.4768)	138 (0.4838)
Toughness (MPa)	51.1 (0.6022)	39.483 (0.4653)	39.875 (0.4699)	37.95 (0.4472)
Corrosion Resistance	Good (0.50)	Good (0.50)	Poor (0.50)	Good (0.50)
Weldability	Excellent (0.6547)	Good (0.4364)	Good (0.4364)	Good (0.4364)
D_i^* (Eq. 15)	0.0268	0.2726	0.2699	0.2786
D_i^- (Eq. 16)	0.2817	0.02088	0.0230	0.02878

C_i (Eq. 17)	0.913	0.0711	0.0785	0.0936
Rank	1	4	3	2
Initial Mixing Ratio	0.75	0.0731	0.0807	0.0962
Final Normalised Ratio	0.9345	0.0191	0.0212	0.0252

The following steps were performed:

Step 1: The categorical variables for corrosion resistance and weldability were converted into numerical form using the following scale: excellent = 3, good = 2, fair = 1, and poor = 0. These numerical values were then used to construct a property matrix for all alloy properties, following the procedure outlined in Step 1 of **Section 2.2.3.8.3**.

Step 2: Each property value was then normalised using Equation (14). The corresponding normalised values are presented in parentheses in Table 6.2. Based on these values, the algorithm identified the ideal-best and ideal-worst solutions from the alloys within Sub-cluster 01, as follows:

Ideal-Best Solution T^* :

- Maximum for benefits: Yield strength, modulus of toughness, fatigue strength, weldability, and corrosion resistance
- Minimum for costs: None (all properties are beneficial)
- $T^* = (348, 160, 51.1, 3, 3)$

Ideal-Worst Solution T^- :

- Minimum for benefits: Yield strength, modulus of toughness, fatigue strength, corrosion resistance, weldability
- Costs: None (all properties are beneficial)
- $T^- = (330, 135, 37.95, 0, 2)$

Step 3: We then calculated the distance of each alloy from the ideal-best solution using Equation (15); for example, for Alloy B1, as follows:

$$\begin{aligned}
& D_{B1}^* \\
&= \sqrt{(0.4902 - 0.5170)^2 + (0.5609 - 0.5609)^2 + (0.6022 - 0.6022)^2 + (0.50 - 0.50)^2 + (0.6547 - 0.6547)^2} \\
&\approx 0.0268
\end{aligned}$$

Similarly, the distances of the other alloys from the ideal-best solution were calculated as follows: $D_{B2}^* = 0.2726$, $D_{B3}^* = 0.2699$, and $D_{B4}^* = 0.2786$.

Step 4: The distance of each alloy from the ideal-worst solution was calculated using Equation (16); for example, for Alloy B1, it was calculated as follows:

$$\begin{aligned}
& D_{B1}^- \\
&= \sqrt{(0.4902 - 0.4902)^2 + (0.5609 - 0.4733)^2 + (0.6022 - 0.4472)^2 + (0.50 - 0.50)^2 + (0.6547 - 0.4364)^2} \\
&\approx 0.2817
\end{aligned}$$

Similarly, the distances of the other alloys from the ideal-worst solution were: $D_{B2}^- = 0.02088$, $D_{B3}^- = 0.0230$, and $D_{B4}^- = 0.02878$.

Step 5: The TOPSIS score was then calculated using Equation (17):

$$\text{For alloy A: } C_{B1} = \frac{0.2817}{0.0268 + 0.2817} \approx 0.913$$

Similarly, the TOPSIS scores for other alloys were $C_{B2} = 0.0711$, $C_{B3} = 0.0785$, and $C_{B4} = 0.0936$.

Step 6: Ranking the five alloys in sub-cluster 01 of cluster 0 based on the TOPSIS score as follows: Alloy B1 (optimum alloy) > Alloy B4 > Alloy B3 > Alloy B2.

Based on the TOPSIS calculations above, Alloy B1 was identified as the optimum choice, as shown in Figure 6.4, indicating that it provides a balanced performance across all properties compared with the other alloys.

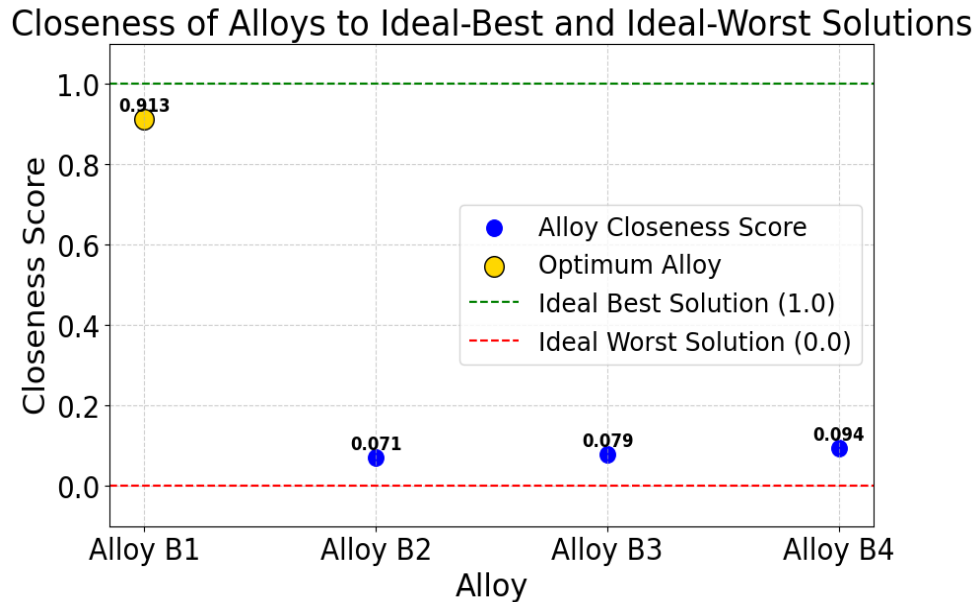


Figure 6.4. TOPSIS closeness scores of alloys A, B, C, and D

The optimum alloys for the remaining sub-clusters were selected using the same approach. This optimisation ensured that the chosen alloys met the specified criteria, encompassing a broad range of properties and applications relative to the other alloys within each sub-cluster, as detailed in Table 6.3.

Table 6.3. Optimum low-Cu 7xxx series alloys and their corresponding applications

Clusters	Optimum Alloys and their Properties	Applications
Cluster 0	7028, Sub-cluster 00: High strength, medium fatigue strength, and modulus of toughness and good corrosion resistance and weldability	Transportation and automotive applications [1], [184]
	7039, Sub-cluster 01: Medium strength, medium fatigue strength, high modulus of toughness, good corrosion resistance and excellent weldability	Aerospace armor alloy [185]
Cluster 1	7278, Sub-cluster 10: Very high strength, high fatigue, moderate modulus of toughness, good corrosion resistance and weldability	Military, automotive, aerospace applications [186], [187]
	7035, Sub-cluster 11: Medium strength, medium fatigue strength, high modulus of toughness, good corrosion resistance and weldability	Aerospace, automotive, and defence applications [1], [184]
Cluster 2	7005, Sub-cluster 20: Low strength, fatigue strength, medium modulus of toughness, excellent corrosion resistance and weldability	Bicycle frames, rapid trains, machine frames, and outdoor products [188]

	7024, Sub-cluster 21: Slightly high strength, fatigue strength, and modulus of toughness and excellent corrosion resistance and weldability	Aircraft parts, automotive components [189], [7], [1]

6.2.4 Alloy Mixing Within Sub-clusters: Recycling Algorithm

Following the selection of optimum alloys for the sub-clusters of the low-Cu alloy sub-set, alloy mixing was conducted within each sub-cluster to generate optimal compositions and properties of recycled alloy using the recycling algorithm. The following example demonstrates the procedure employed for this alloy mixing:

As previously calculated for Sub-cluster 01, the alloys received the following TOPSIS closeness scores using the optimisation algorithm: Alloy B1 (0.913), Alloy B4 (0.0936), Alloy B3 (0.0785), and Alloy B2 (0.0711). Alloy B1, having the highest closeness score, was identified as the optimum alloy. Consequently, it was assigned a minimum mixing ratio of 50% (in this instance, 0.75%) in accordance with the alloy recycling algorithm described in **Section 3.7**.

Since the sum of the closeness scores for the remaining alloys ($0.913 + 0.0936 + 0.0785 + 0.0711$) is 1.1562, the individual mixing ratios for these alloys were calculated using Equation (25): Alloy B1 – 0.75, Alloy B4 – 0.0962, Alloy B3 – 0.0807, and Alloy B2 – 0.0731.

The final mixing ratios for all alloys, after normalisation using Equation (26), were as follows: Alloy B1 – 0.9345, Alloy B4 – 0.0252, Alloy B3 – 0.0212, and Alloy B2 – 0.0191, summing to 1 (or 100%).

Finally, the compositions and properties of the mixed alloy were determined as weighted averages of the individual alloys' feature values and their respective mixing ratios, using Equation (27).

The final composition of the recycled alloy, comprising Zn, Cu, Mg, Fe, Mn, Si, Cr, Ti, and Zr, was calculated as follows:

$$\mathbf{Zn} = (4 \times 0.9345) + (4.9 \times 0.0252) + (4 \times 0.0212) + (4.2 \times 0.0191) = 4.0265$$

$$\mathbf{Cu} = (0.05 \times 0.9345) + (0.105 \times 0.0252) + (0.05 \times 0.0212) + (0.025 \times 0.0191) = 0.0509$$

$$\mathbf{Mg} = (2.8 \times 0.9345) + (1.7 \times 0.0252) + (1.15 \times 0.0212) + (1.5 \times 0.0191) = 2.7125$$

$$\mathbf{Fe} = (0.2 \times 0.9345) + (0.15 \times 0.0252) + (0.05 \times 0.0212) + (0.175 \times 0.0191) = 0.1951$$

$$\mathbf{Mn} = (0.4 \times 0.9345) + (0.1 \times 0.0252) + (0.6 \times 0.0212) + (0.7 \times 0.0191) = 0.4024$$

$$\mathbf{Si} = (0.3 \times 0.9345) + (0.2 \times 0.0252) + (0.3 \times 0.0212) + (0.25 \times 0.0191) = 0.2965$$

$$\mathbf{Cr} = (0.125 \times 0.9345) + (0.075 \times 0.0252) + (0.175 \times 0.0212) + (0.025 \times 0.0191) = 0.1229$$

$$\mathbf{Ti} = (0.05 \times 0.9345) + (0.05 \times 0.0252) + (0.05 \times 0.0212) + (0.025 \times 0.0191) = 0.0495$$

$$\mathbf{Zr} = (0 \times 0.9345) + (0.1 \times 0.0252) + (0 \times 0.0212) + (0.1 \times 0.0191) = 0.00443$$

The mechanical properties of the recycled alloy were calculated as follows:

$$\mathbf{YS} = (330 \times 0.9345) + (348 \times 0.0252) + (331 \times 0.0212) + (337 \times 0.0191) = 330.61 \text{ MPa}$$

$$\mathbf{Fatigue \text{ Strength}} = (160 \times 0.9345) + (138 \times 0.0252) + (136 \times 0.0212) + (135 \times 0.0191) = 158.46 \text{ MPa}$$

$$\mathbf{Modulus \text{ of Toughness}} = (51.1 \times 0.9345) + (37.95 \times 0.0252) + (39.875 \times 0.0212) + (39.483 \times 0.0191) = 50.29 \text{ MPa \%}$$

$$\mathbf{Corrosion \text{ Resistance}} = (2 \times 0.9345) + (2 \times 0.0252) + (0 \times 0.0212) + (2 \times 0.0191) = 1.9576 \approx \text{Round to closest rating i.e., Good}$$

$$\mathbf{Weldability} = (3 \times 0.9345) + (2 \times 0.0252) + (2 \times 0.0212) + (2 \times 0.0191) = 2.9345 \approx \text{Excellent}$$

The resulting recycled alloy composition and corresponding weighted properties are summarised in Table 6.4.

Table 6.4. Final Composition and Properties of the Recycled Alloy (Sub-cluster 01)

Parameter	Final Value
Chemical Composition (wt.%)	
Zn	4.03
Cu	0.051

Mg	2.71
Fe	0.195
Mn	0.402
Si	0.297
Cr	0.123
Ti	0.050
Zr	0.004
Mechanical Properties	
Yield Strength, MPa	330.61
Modulus of Toughness, MPa%	50.29
Fatigue Strength, MPa	158.46
Other Properties	
Corrosion Resistance	≈ Good
Weldability	≈ Excellent

Therefore, the alloy mixing from Sub-cluster 01 of Cluster 0 reproduces the optimum alloy predicted by the optimisation algorithm, with a composition error of less than 0.001%. Similarly, the optimum alloys for the remaining sub-clusters can be identified.

Using this approach, six optimum alloys were selected from the low-Cu 7xxx series, representing a reduction of over 50% from the original 24 standard alloys. This methodology ensures the selection of alloys with optimal compositions and properties, suitable for a broad range of applications across various fields.

6.2.5 Validation of Recycled Alloy Properties

After obtaining the optimum alloys with superior property combinations through the optimisation and recycling algorithms, it was essential to validate these predictions using an independent dataset. For this purpose, data from the MatWeb database, which was not included in the original dataset and was therefore completely unseen by the algorithm was employed. As shown in Table 6.5, the comparison demonstrates minimal differences, thereby confirming the accuracy and reliability of our predictions.

Table 6.5. Comparison of predicted vs reported properties of optimum alloys for Sub-cluster 01 of cluster 0 for low Cu containing alloys

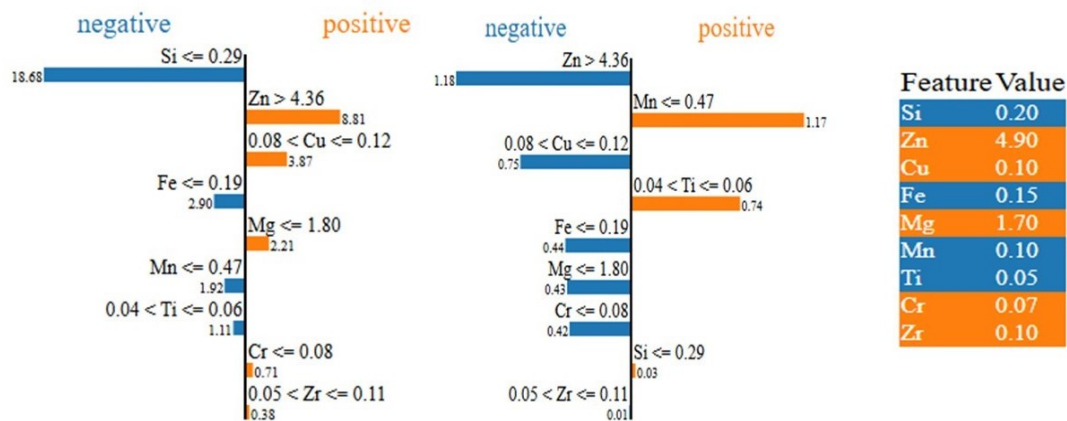
Property	Alloy B1 (7039) (Predicted)	Alloy B1 (7039) (Reported)
Yield strength, MPa	330.61	330
Elongation, %	14	14
Fatigue strength, MPa	158.5	160

We are confident that this methodology can be further refined and applied to enhance the properties and functionality of the alloys, while promoting the sustainable utilisation of resources.

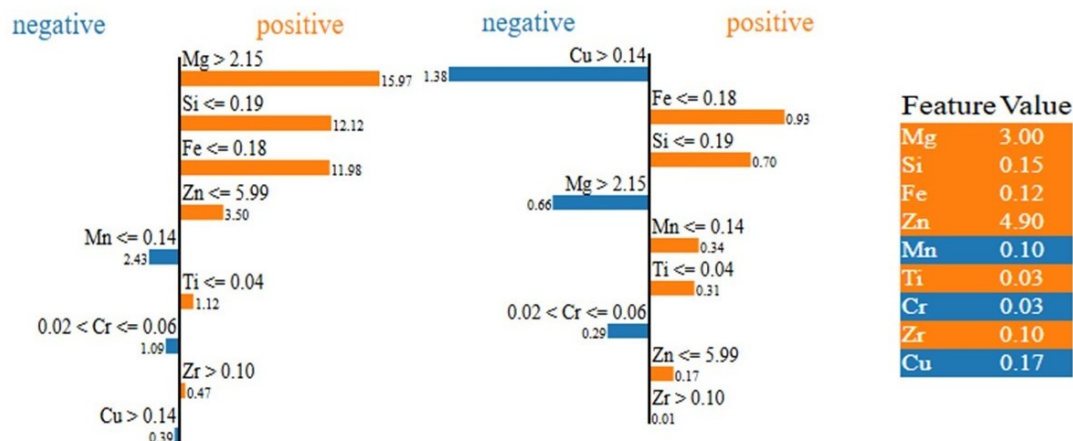
6.2.6 LIME Analysis to Explain the Clustering Results

After applying PCA and K-means clustering to the low-Cu 7xxx-series aluminium alloys, the dataset was grouped into three clusters as discussed in **Section 6.2.1**. However, as the clustering algorithm functions as a “black box”, the specific factors driving variations in properties within each cluster, and how the compositions influence these properties remain unclear.

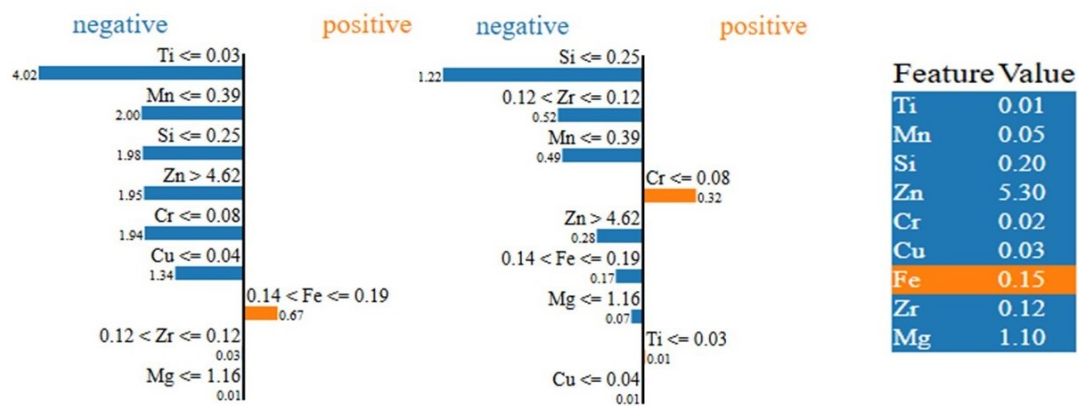
To elucidate this, the LIME algorithm (see **Section 2.2.3.9**) was employed to determine the contribution of individual alloying elements to key properties within each cluster. For instance, in Cluster 1 (Figure 6.5(b)), the high yield strength can be attributed to the positive impact of most of the alloying elements. The modulus of toughness was also elevated, as most elements exerted a positive effect, with the exception of Cu, which is present at very low levels (~0.17%) in this cluster. In contrast, Cluster 2 exhibits lower yield strength, as all alloying elements, except Fe, negatively influence the yield strength, while the modulus of toughness is slightly reduced due to the negative effect of most elements, except Cr. These observations explain why Cluster 2 comprises low-strength alloys.



(a) Cluster 0, YS and Modulus of toughness



(b) Cluster 1, YS and Modulus of Toughness



(c) Cluster 2, YS and Modulus of toughness

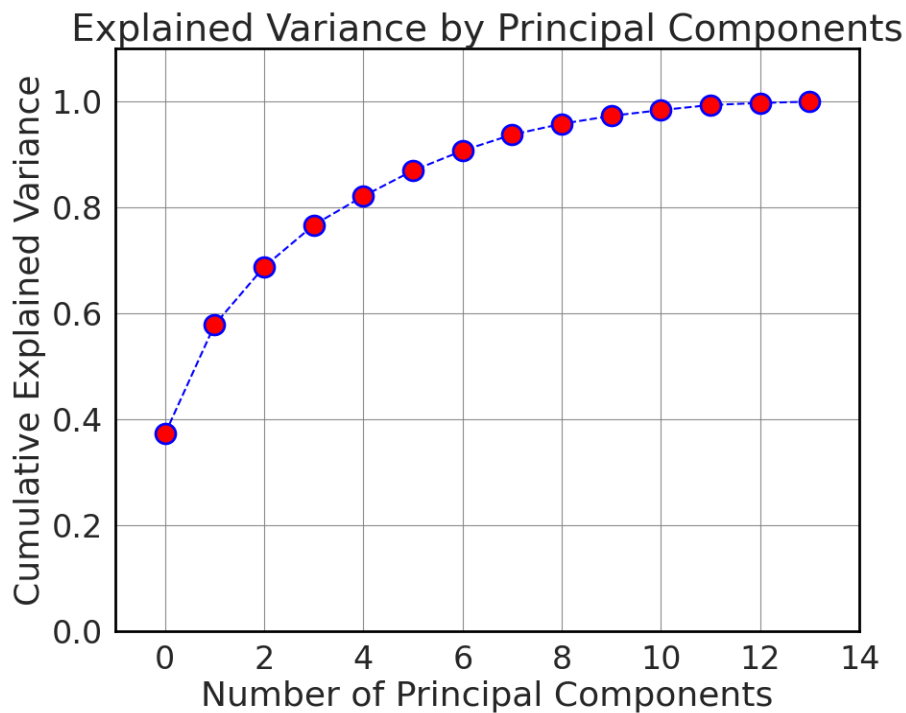
Figure 6.5. LIME results for YS and modulus of toughness for (a) Cluster 0, (b) Cluster 1, and (c) Cluster 2

6.3 Optimisation of High Cu-Containing 7xxx Series Alloys

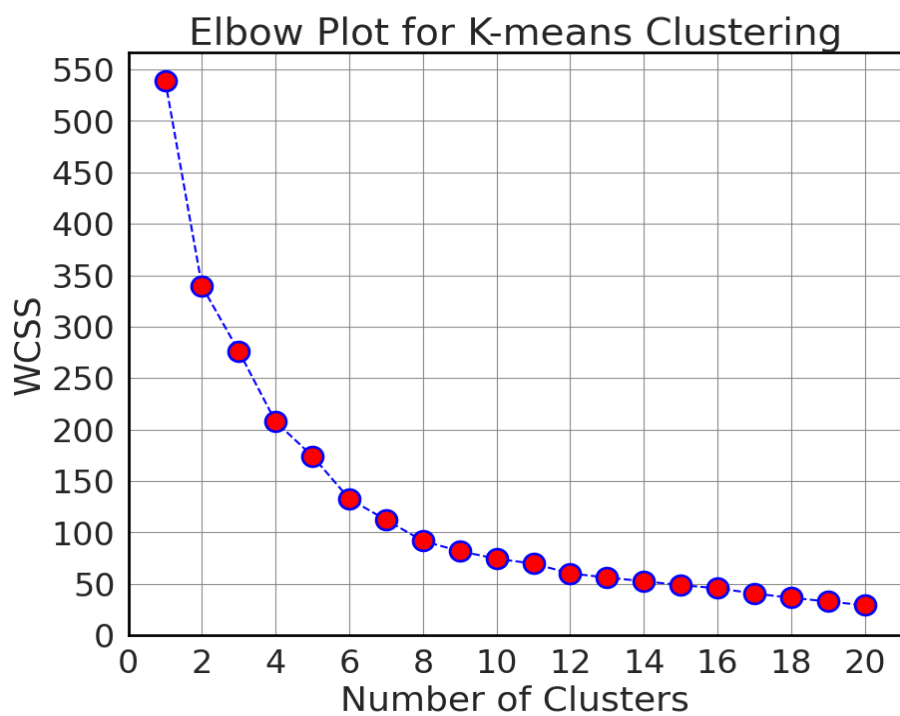
This section focuses on the optimisation of high Cu-containing 7xxx series alloys. The methodology follows a similar approach to that used for the low-Cu alloys, including clustering (Section 3.4), sub-clustering analysis (Section 3.5), selection of optimum alloys (Section 3.6), and recycling analysis (Section 3.7). The following sub-sections presents the main clustering findings and subsequent optimisation of clusters.

6.3.1 Clustering of High Cu-Containing 7xxx Series Alloys

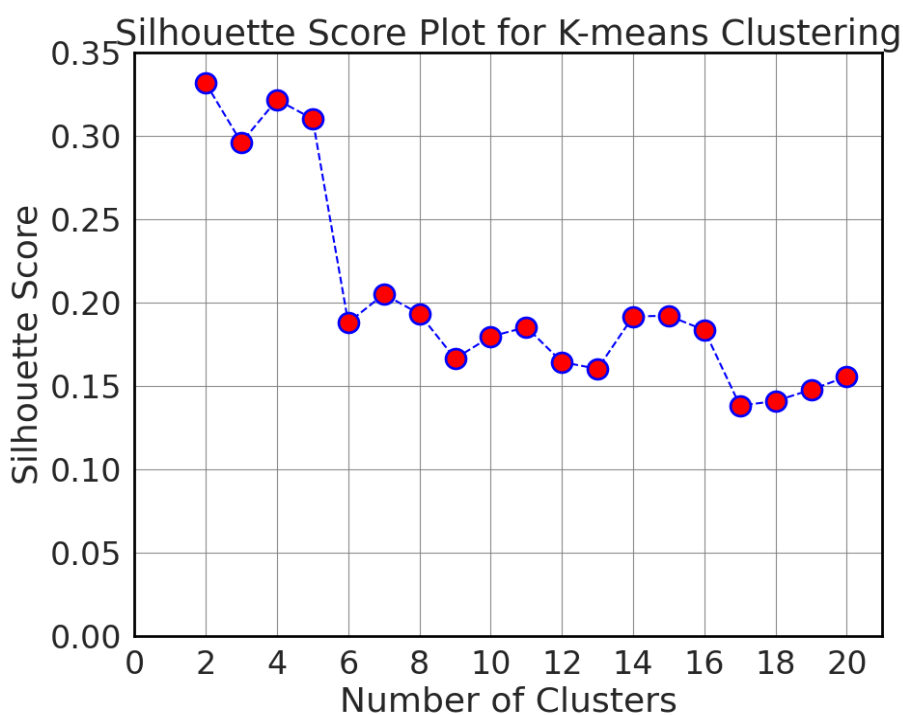
Data standardisation was performed to place all features on a comparable scale. Principal Component Analysis (PCA) as described in **Section 2.2.3.6**, was then conducted to determine the number of components required to effectively represent the dataset in a 2D space. The PCA variance plot in Figure 6.6(a) indicates that over 75% of the dataset's information can be captured using four principal components, demonstrating that this number is sufficient to explain the majority of variance in the high-Cu alloys. This highlights the efficiency of PCA in reducing dimensionality while retaining essential information for subsequent analysis. The Elbow method, in conjunction with a Silhouette score of 0.32 (Figures 6.6(b) and 6.6(c)), was then applied to identify the optimal number of clusters, revealing the presence of K=4 distinct clusters (See **Section 2.2.3.7.1**).



(a)



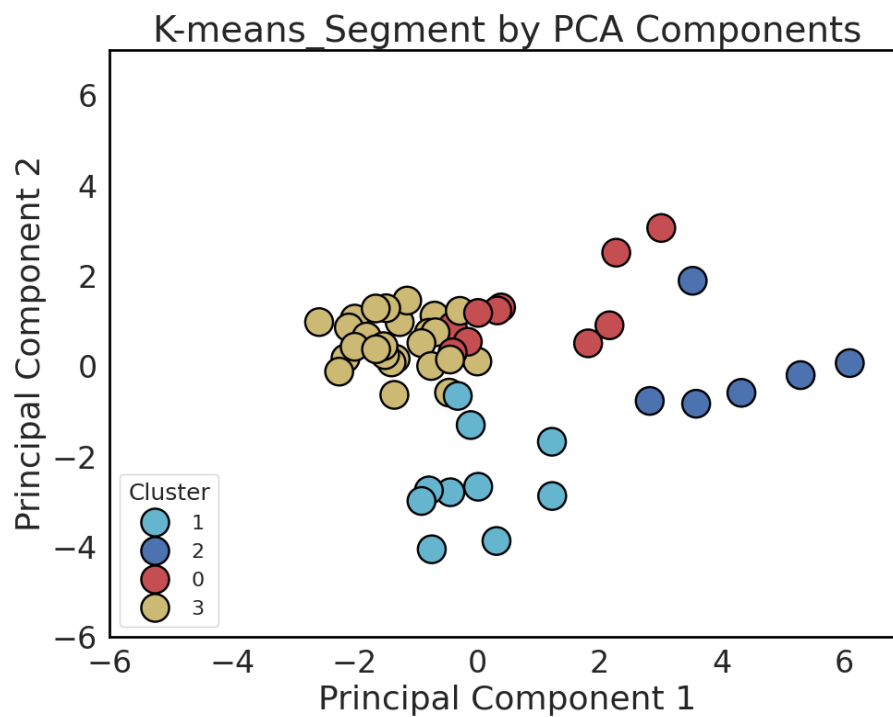
(b)



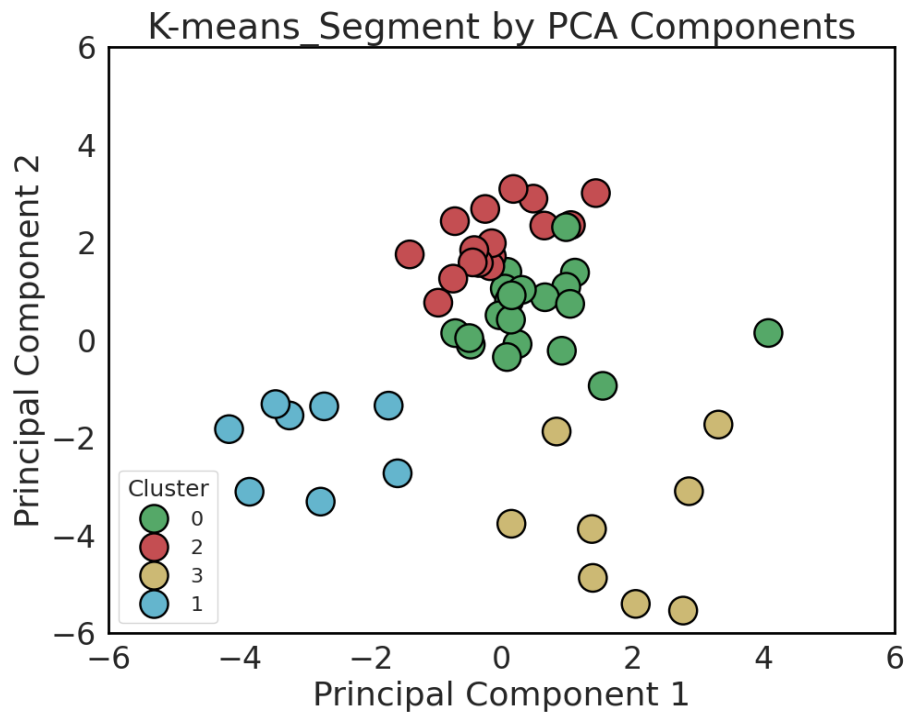
(c)

Figure 6.6. PCA variance plot (a), Elbow plot (b), and (c) Silhouette score plot for high Cu 7xxx series alloys

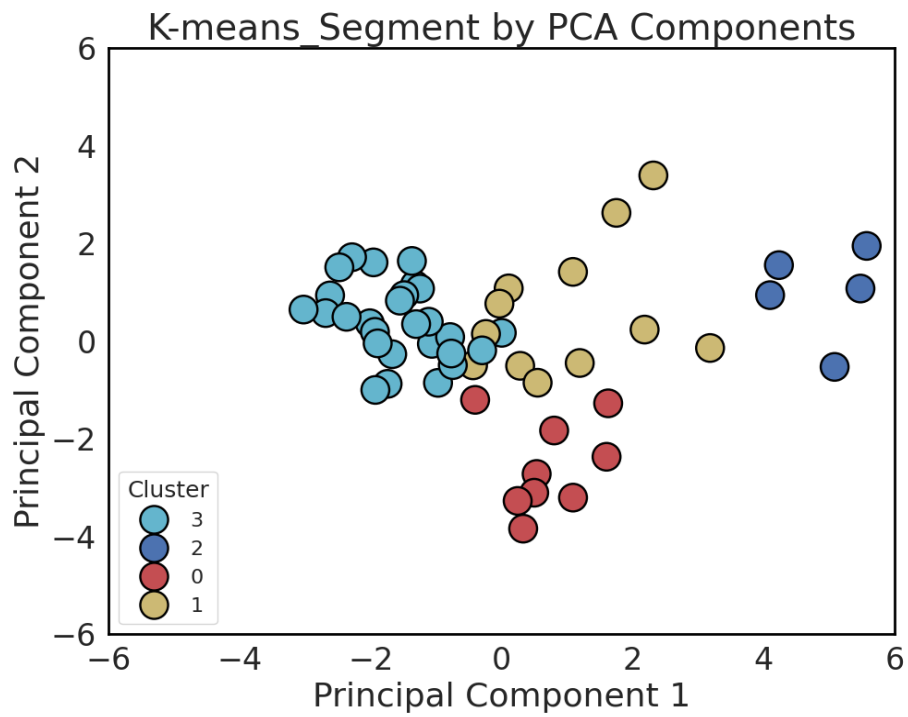
To examine these clusters in greater detail, they were visualised under the widely used T6 tempering condition. As illustrated in Figure 6.7(b), the clusters were well defined, exhibiting minimal overlap. Comparable clustering patterns were observed under the T5 (Figure 6.7(a)) and T7 (Figure 6.7(c)) tempering conditions, thereby reinforcing the robustness of the clustering results across different heat treatments. Collectively, these findings demonstrate the presence of four distinct groupings in the high-Cu 7xxx series alloys, each defined by similar alloy compositions and corresponding properties.



(a) T5



(b) T6



(c) T7

Figure 6.7. Clustering of high Cu 7xxx series aluminium alloys at (a) T5, (b) T6, and (c) T7 tempering condition

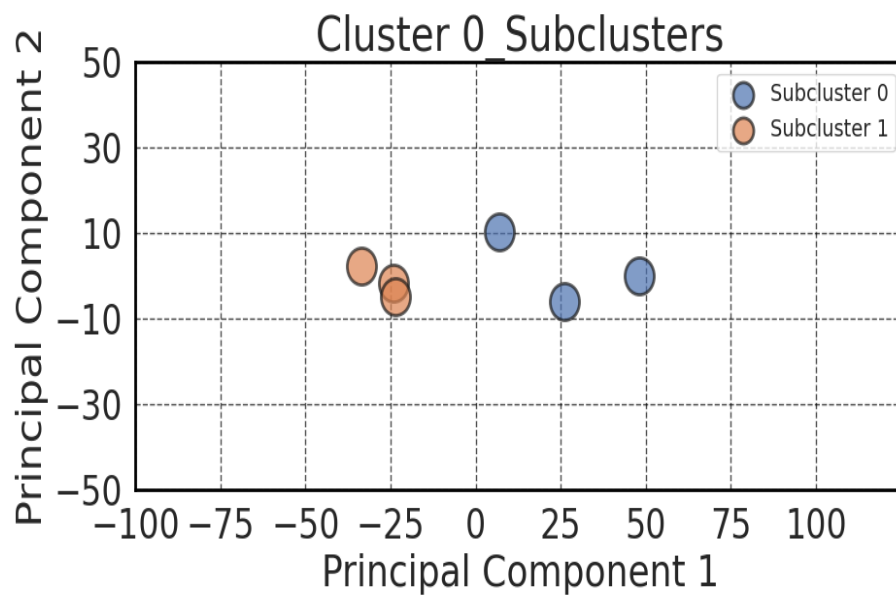
The variations in properties and composition ranges within each cluster were summarised in Table 6.6. For instance, Cluster 2 comprised alloys with high yield strength but a slight compromise in corrosion resistance, whereas Cluster 0 included alloys that provided a balanced trade-off between corrosion resistance and yield strength, exhibiting moderate levels in both properties. Similar patterns were observed across the remaining clusters, indicating that the clustering algorithm effectively distinguished alloys based on variations in their mechanical, corrosion properties.

Table 6.6. Range of properties and compositions in high Cu alloys

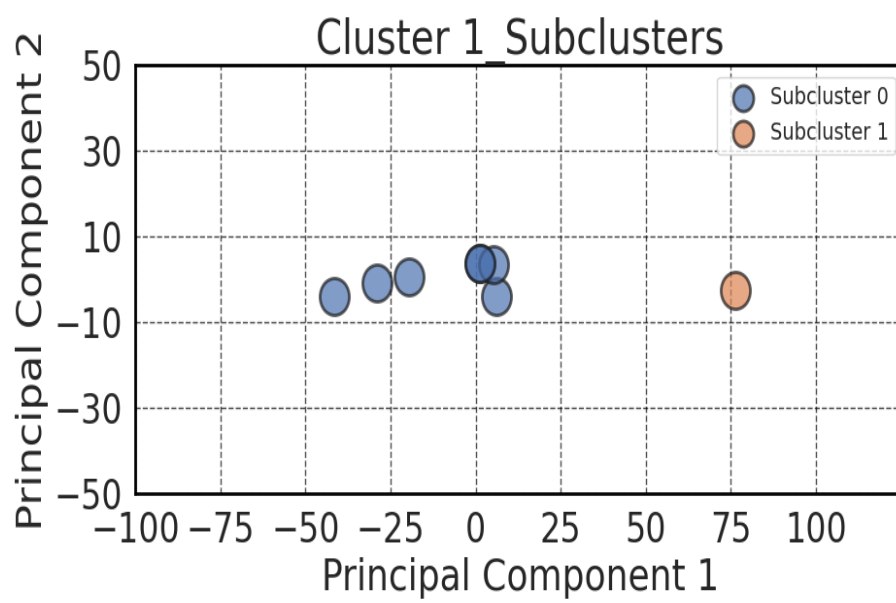
Groups	Cluster 0 (Medium)	Cluster 1 (Low)	Cluster 2 (Highest)	Cluster 3 (High)
Zn, wt%	4.75-7.5	4.5-6.2	6.2-11.5	5.6-8.4
Mg, wt%	1.6-3.15	1.1-1.9	1.5-2.75	1.5-3
Cu, wt%	0.5-0.75	0.65-1.0	0.95-2.4	0.85-2.2
Zn+Mg+Cu, wt%	8.25-9.75	6.3-8.75	10.15-15	9.25-12.5
Fe, wt%	0.175-0.3	0.04-0.15	0.04-0.125	0.05-0.2
Mn, wt%	0.1-0.8	0.03-0.2	0-0.4	0.02-0.5
Si, wt%	0.25-0.5	0.06-0.15	0.06-0.25	0.1-0.4
Cr, wt%	0-0.175	0-0.1	0-0.125	0.02-0.175
Ti, wt%	0-0.1	0.015-0.05	0-0.05	0.015-0.1
Zr, wt%	0-0.125	0-0.075	0-0.25	0-0.3
Yield strength, MPa	420-501	378-494	556-730	520-625
Fatigue strength MPa	154-170	144-163	169-180	150-210
Modulus of toughness, MPa	30.69-41-8	36.88-45.89	33.45-59.2	44-61.05
Corrosion	Good- Excellent	Good-Excellent	Fair-Excellent	Fair-Excellent
Weldability	Fair-Excellent	Good- Excellent	Poor-Fair	Poor-Good
Zn:Mg Ratio	1.51-4.69	2.8-4.27	2.51-5.41	2.4-5.411

6.3.2 Sub-clustering of High Cu Alloys Clusters

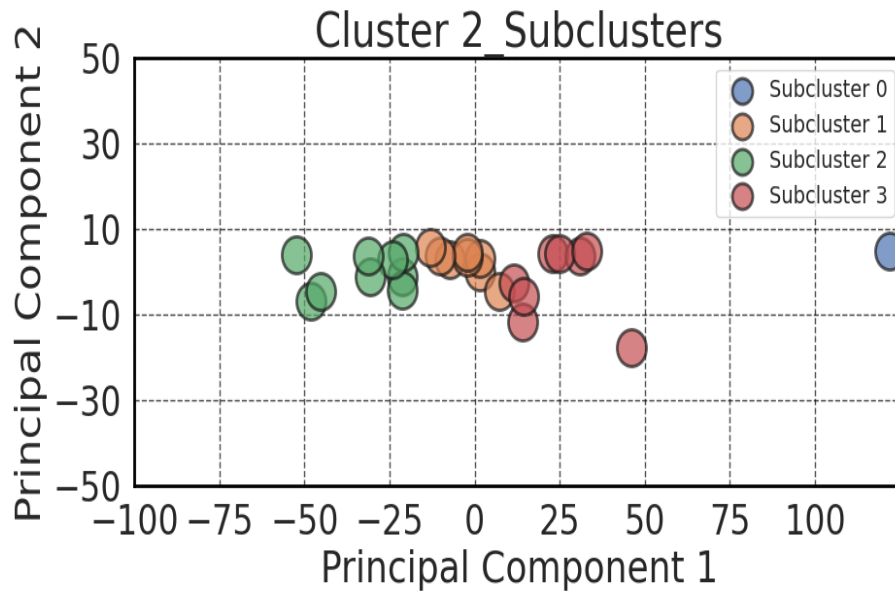
After identifying the primary clusters, further refinement was achieved through sub-clustering to obtain narrower property ranges (see **Section 3.5**). The four main clusters were subdivided into 11 distinct sub-clusters, as illustrated in Figure 6.8. For instance, Sub-cluster 20 of Cluster 2 comprised high-strength alloys, whereas Sub-cluster 21 of Cluster 2 contained alloys with moderate corrosion resistance. Sub-cluster 22 included alloys exhibiting very high yield strength. This sub-clustering ensured that each group displayed a more restricted range of properties, thereby highlighting the specific characteristics of the alloys within it. The same procedure applied to the remaining clusters produced similarly well-defined sub-clusters, each characterised by distinct property ranges.



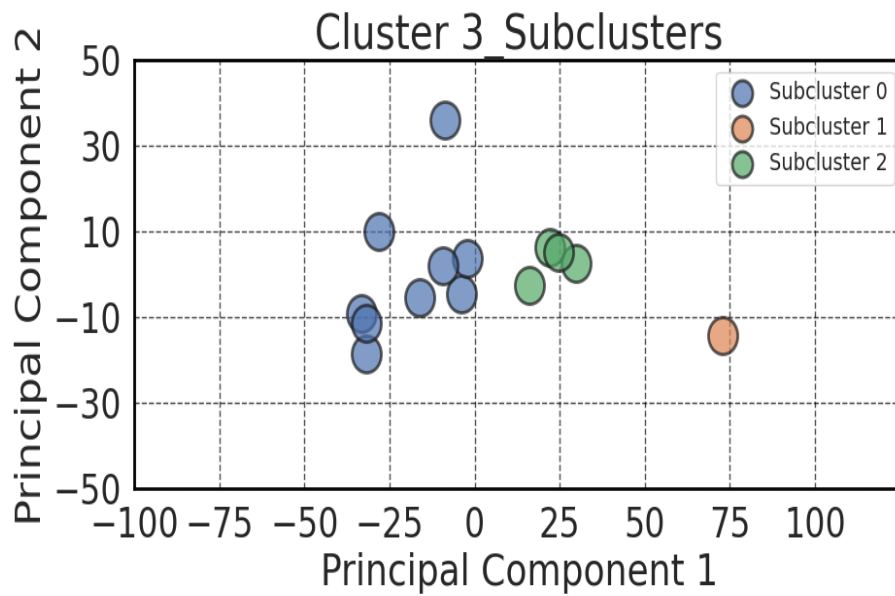
(a)



(b)



(c)



(d)

Figure 6.8. Sub-clusters of high-Cu 7xxx series alloy: (a) Cluster 0, sub-clusters, (b) Cluster 1, sub-clusters, (c) Cluster 2, sub-clusters, and (d) Cluster 3, sub-clusters

6.3.3 Optimum Alloy Selection and Recycling Analysis

After the sub-clustering analysis, which resulted in 11 sub-clusters with more compact property ranges, the optimisation algorithm, as described in **Section 3.6**, was applied to identify the

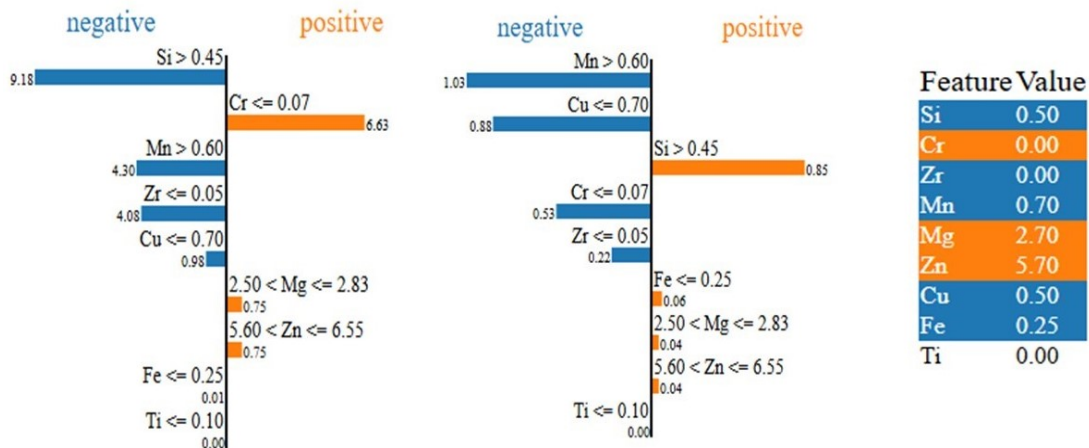
optimum alloys for the high-Cu 7xxx series alloys. This approach ensured that the selected optimum alloys adequately represented the full range of properties and compositions within each sub-cluster. The resulting set of optimum alloys for each cluster and sub-cluster is summarised in Table 6.6. These optimum alloys were subsequently refined using the recycling algorithm within each sub-cluster, as detailed in **Section 3.7**, in which a minimum mixing ratio of 50% was assigned to the optimum alloy and the remaining 50% was proportionally distributed among the other alloys in the sub-cluster. This recycling procedure produced recycled alloys with minimal deviation from the optimum alloys identified by the optimisation algorithm, as shown in Table 6.7.

Table 6.7. Optimum high-Cu 7xxx series alloys and their corresponding applications

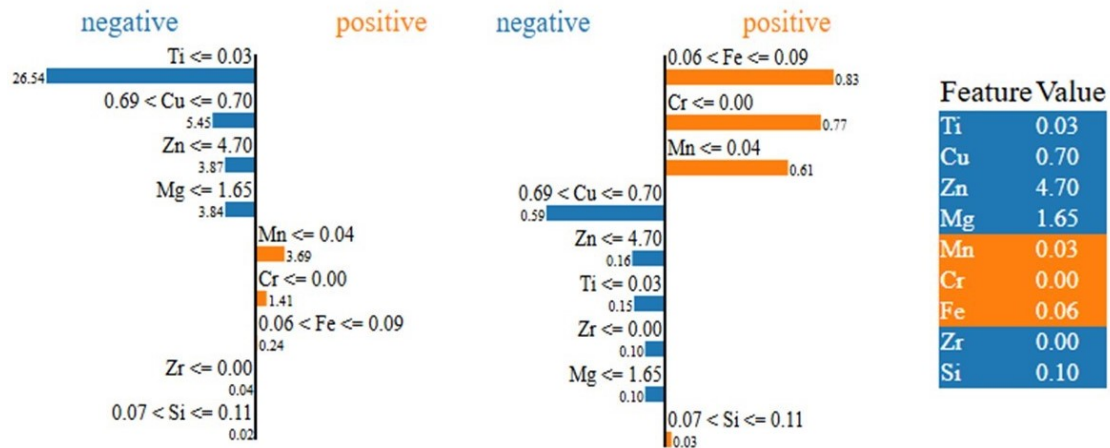
Clusters	Optimum Alloys and their Properties	Applications
Cluster 0	7023, Sub-cluster 00: Moderate strength, moderate fatigue strength, moderate modulus of toughness	Automotive and aerospace industries [190], [1], [184]
	7014, Sub-cluster 01: Medium to high strength, fatigue strength and modulus of toughness	Aerospace and automotive industries [191]
Cluster 1	7041, Sub-cluster 10: High strength, fatigue strength and high modulus of toughness	Aerospace, military, and nuclear industries [192]
	7033, Sub-cluster 11: Medium strength, medium modulus of toughness, good corrosion resistance and weldability	Automotive applications [193]
Cluster 2	7136, Sub-cluster 20: Very high strength, fatigue strength, and modulus of toughness but fair weldability, good corrosion resistance	Primarily used in the aerospace industry [194]
	7249, Sub-cluster 21: High strength, fatigue strength, and modulus of toughness, fair weldability, and excellent corrosion resistance	Automotive and aerospace applications [195], [196]
	7085, Sub-cluster 22: High strength, fatigue strength, and very high modulus of toughness, fair weldability, and excellent corrosion resistance	Thick and ultra-thick section products for both aerospace and non-aerospace applications [197]
	7034, Sub-cluster 23: Superior strength, fatigue strength, and modulus of toughness and poor weldability	Manufacturing structure components of aerospace equipment [198]
Cluster 3	7149, Sub-cluster 30: High strength, fatigue strength, and modulus of toughness and good corrosion resistance and weldability	Aerospace and defence sectors [199]
	7081, Sub-cluster 31: High strength, fatigue strength, and modulus of toughness and excellent corrosion resistance and weldability	Aeronautic applications, such as wing components or fuselage frames [200]
	7001, Sub-cluster 32: Very high strength, medium fatigue strength, and high modulus of toughness and excellent corrosion resistance and good weldability	Tent poles, bicycle frames, and aircraft components [201], [7]

6.3.4 Understanding the Clustering Process: LIME Algorithm

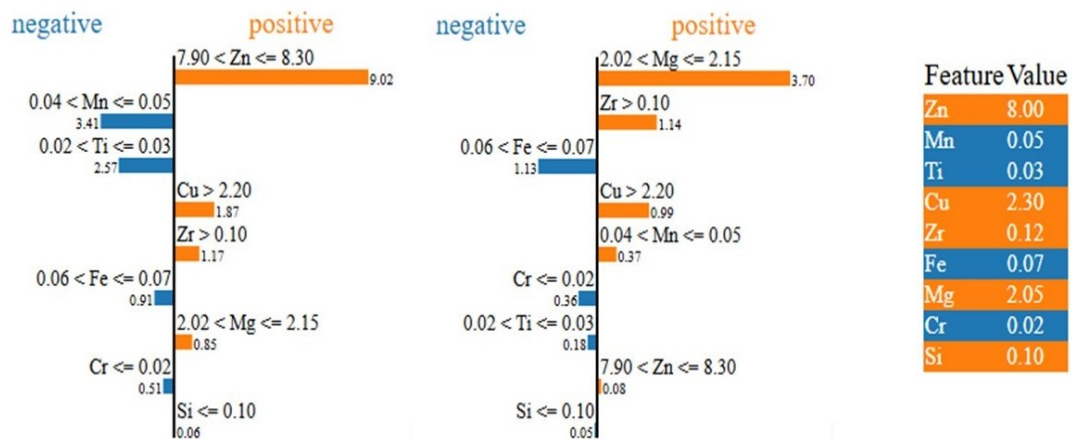
After the clustering analysis and optimisation of clusters, the LIME algorithm was employed to identify which alloying elements exerted positive or negative effects on the properties within each cluster, as illustrated in Figure 6.9 (see **Sections 2.2.3.9** and **3.8**). In Cluster 2 (Figure 6.9(c)), the high yield strength was primarily attributed to the positive contributions of Zn, Mg, Cu, and Zr, whereas Fe and Mn had a negative effect. In contrast, for the modulus of toughness, Mg, Mn, Zn, Cu, and Zr contributed positively, while Fe had a minor negative influence; however, its low concentration allowed the cluster to maintain a high modulus of toughness. Cluster 0 (Figure 6.9(a)) exhibited moderate yield strength due to the positive effects of Zn, Mg, and Cr, while Cu, Mn, and Zr showed negative effects. The modulus of toughness was also moderate, influenced positively by Si, Mg, and Zn, and negatively by Mn, Cu, and Cr.



(a) Cluster 0, YS and Modulus of toughness



(b) Cluster 1, YS and Modulus of toughness



(c) Cluster 2, YS and Modulus of toughness

Figure 6.9. LIME results for (a) Cluster 0 (Medium properties), (b) Cluster 1 (Low properties), and (c) Cluster 2 (High properties)

A similar procedure was applied to the remaining clusters, demonstrating that the clustering approach was meaningful and provided clear interpretations of the groupings. These observations will be further supported in the following section through detailed metallurgical reasoning.

6.4 Metallurgical Reasoning Behind Clustering of 7xxx Series Alloys

The precipitation behaviour and mechanical properties of 7xxx series aluminium alloys vary markedly depending on the presence or absence of copper. In Al–Zn–Mg alloys without

copper, the precipitation sequence proceeds as follows: supersaturated solid solution (SSSS) $\rightarrow$ Guinier–Preston (GP) zones $\rightarrow \eta'$ (MgZn_2) $\rightarrow \eta$ (MgZn_2) [202]. The primary strengthening phase is the semi-coherent η' precipitate, which provides high strength through its fine, evenly distributed particles that interact with dislocations. GP zones act as the initial strengthening sites, while η precipitates are coarse and incoherent, leading to over-ageing. The mechanical performance of these alloys is therefore dominated by the η' precipitates, resulting in a favourable strength-to-weight ratio alongside moderate resistance to stress corrosion cracking (SCC). In over-aged conditions, the absence of copper enhances SCC resistance and improves the modulus of toughness; however, the overall strength remains lower than that of alloys containing copper [202].

In contrast, the precipitation sequence in Al–Zn–Mg–Cu alloys is more complex owing to the presence of copper: SSSS $\rightarrow$ GP zones $\rightarrow \eta'$ (MgZn_2) + T' ($\text{Al}_2\text{Mg}_3\text{Zn}_3\text{Cu}$) $\rightarrow \eta$ (MgZn_2) + T ($\text{Al}_2\text{Mg}_3\text{Zn}_3\text{Cu}$) [203]. The addition of copper refines the distribution of η' precipitates and promotes the formation of the T' phase, a copper-rich phase that contributes significantly to both strength and hardness. The synergistic interaction between the T' and η' phases enhance strength retention, particularly at elevated temperatures, and delays the onset of over-ageing. Consequently, these alloys exhibit superior mechanical properties, including higher yield strength and hardness, which underpin their suitability for high-strength applications. However, increasing copper content, while beneficial for strengthening, also raises susceptibility to intergranular corrosion and reduces stress corrosion cracking (SCC) resistance when compared with Al–Zn–Mg alloys. Thus, although copper improves strength, it simultaneously renders the alloys more vulnerable in corrosive environments [204]. The following section builds upon these insights by linking the metallurgical reasoning to the clustering results for both high- and low-Cu-containing 7xxx series alloys, with particular emphasis on clusters associated with high- and low-strength properties.

6.4.1 Reasoning for Low-Cu 7xxx Series Alloys

Cluster 1 (Medium to High Strength), as shown in Table 6.1 and Figure 6.2: This cluster comprises alloys that exhibit medium to high strength, achieved through a combination of elevated Zn, Mg, and Cu contents; low levels of Mn, Si, Cr, and Ti; medium concentrations of Fe and Zr; and an optimal total Zn+Mg content of approximately 6.6–10%. The elevated Zn and Mg contents increase the volume fraction of MgZn_2 (η') precipitates, and, together with their finer size, significantly enhance YS [204]. A medium Zr content facilitates the formation of

fine, coherent Al_3Zr dispersoids, refining grain size and improving both strength and modulus of toughness, while also reducing quench sensitivity [205]. Low Fe and Si concentrations minimise coarse void formation, further contributing to strength [206], [207].

Corrosion resistance is moderate, owing to the high Zn/Mg ratio, which produces broader precipitate-free zones (PFZs) and elevates Zn content in the matrix. This condition increases susceptibility to SCC through hydrogen embrittlement and anodic dissolution of MgZn_2 precipitates [208], [209]. Nevertheless, Zr additions refine grains and mitigate PFZ formation, thereby enhancing resistance to SCC by homogenising slip modes and strengthening grain boundary cohesion [210].

Weldability is moderate, as solidification cracking arising from the high Zn/Mg ratio, coarse grains, and limited grain refiners remains a constraint. The presence of Cu tends to exacerbate grain boundary cracking through the formation of coarse precipitates, while higher Zn/Mg ratios promote weld-zone microcracking, thereby limiting crack healing during solidification [209].

Cluster 2 (Low Strength): This cluster consists of alloys with lower property ranges, characterised by medium Zn and Fe, high Mn, low Mg and Cu, high Si, medium Cr, and elevated Ti and Zr contents. These alloys exhibit reduced strength due to the limited formation of strengthening precipitates such as G.P. zones and η' , which results from a Zn+Mg content of approximately 7% and a medium Zn/Mg ratio [209]. The high Fe and Si levels exacerbate coarse voiding during deformation, resulting in an increased fraction of coarse Fe- and Si-rich particles, which contributes to transgranular fracture and decreased modulus of toughness [206], [209].

Nevertheless, the presence of grain refiners such as Ti and Zr enhances grain refinement during recrystallisation through the formation of coherent Al_3Zr dispersoids, resulting in a noticeable increase in both modulus of toughness and strength [205].

These alloys exhibit excellent SCC resistance, except for some with medium Zn/Mg, which causes wider PFZs at grain boundaries. Additionally, higher concentrations of Zn in the matrix significantly increase SCC susceptibility [208], [209]. For the majority of alloys, the low Mg content limits the formation of Mg-bearing phases at grain boundaries, reducing hydrogen absorption and diffusion, and thereby decreasing SCC initiation [208]. The presence of microalloying elements such as Cr, Ti, and Zr further enhances SCC resistance through mechanisms

including slowing hydrogen transport, stabilising the deformation microstructure, promoting discontinuous η -phase particle growth, and refining PFZs. Grain refinement by Zr and Ti reduces intergranular fracture susceptibility by promoting homogeneous slip modes and lowering electrochemical disparities at grain boundaries. Moreover, the refinement of Fe- and Si-rich phases minimises their cathodic activity, further enhancing SCC resistance [206], [211].

High weldability of these alloys is achieved through three mechanisms: (1) **Grain refinement on solidification:** In particular, grain refiners such as Ti and Zr refine the microstructure into fine equiaxed grains that significantly decrease crack susceptibility (typically achieved by adding the standard grain refiner Al-Ti-B) [212]. (2) **Segregation of alloying elements to grain boundaries:** Mn additions reduce the mushy zone, thus decreasing the tendency for solidification cracking, whereas Zr forms stable dispersoids and increases the strength in grain-boundary regions [213], [211]. (3) **Change of the solidification range:** Lower Mg and medium Zn levels narrow the solidification range, thereby reducing crack length. Together, these mechanisms improve weldability by significantly enhancing resistance to solidification cracking [211].

6.4.2 Reasoning Behind High-Cu 7xxx Series Alloys

Cluster 2 (High Strength Properties): As shown in Table 6.5 and Figure 6.7, these alloys possess superior strength characteristics due to high levels of Zn, Mg, Cu, and Zr, which together result in a high solute content (Zn+Mg+Cu); low concentrations of Fe and Si; moderate to low amounts of Mn and Cr; and moderate Ti content, leading to the extensive precipitation of fine strengthening phases, namely GP zones, η' , and η .

The high Zn content enhances the precipitation of MgZn_2 (η'), increasing the volume fraction of matrix precipitates that impede dislocation motion and thereby raise yield strength via mechanisms such as Orowan strengthening [204]. Cu plays a critical role in supersaturation, accelerating the formation of GP zones and η' phases. Elevated Mg content further increases precipitate volume while reducing their size, enhancing yield strength [214]. Moreover, a $\text{Cu/Mg} > 1$ promotes the formation of finer precipitates, which improves both modulus of toughness and ductility [210]. The high Zr content promotes grain refinement by forming coherent Al_3Zr dispersoids during recrystallisation, which further enhances strength and reduces quench sensitivity. However, excessive alloying may cause negative effects, such as grain boundary segregation and the development of coarse intermetallic compounds. These drawbacks are partially mitigated through the use of grain refiners [205].

Despite the strength benefits conferred by the formation of η and η' precipitates, their presence increases the chemical potential difference between the matrix and grain boundaries. This leads to the formation of grain boundary precipitates primarily η (MgZn_2) and T ($\text{Al}_2\text{Mg}_3\text{Zn}_3\text{Cu}$) phases, which act as preferential sites for anodic dissolution, increasing the susceptibility to stress corrosion cracking (SCC) and hydrogen embrittlement. The high Zn content further exacerbates SCC susceptibility through enhanced anodic activity. However, Cu-rich precipitates reduce hydrogen generation and slow electrochemical reaction rates, providing some resistance to SCC [208]. Additionally, elevated Zr levels help stabilise deformation substructures and promote the formation of low-angle sub-grains alongside discontinuous η -phase distributions, collectively reducing SCC susceptibility [212], [215]. Thus, the overall SCC resistance of these alloys is moderate, reflecting the interplay of both beneficial and detrimental effects of the alloying elements.

The weldability of these alloys is low, particularly when processed through fusion welding. The high Cu content increases susceptibility to grain boundary cracking and promotes the formation of coarse precipitates within the weld zone, resulting in a reduction in hardness as coarsened η phases replace the finer η' precipitates. Furthermore, the high Zn/Mg ratio correlates with an increased occurrence of micro-cracking [211]. Low grain refinement, due to Ti content ($<0.05\%$), leads to coarse grain structures in the weld zone. Although the inclusion of Zr affords partial grain refinement, its effect is limited. Consequently, while weldability may be considered fair due to Zr-induced refinement, maintaining mechanical integrity in welded joints remains a significant challenge, leading to overall low weldability [212].

Cluster 1 (Low Properties), as shown in Table 6.5 and Figure 6.7, consists of alloys with low Zn, medium Mn and Si, low Cr, Ti, and Zr, medium Fe, Cu, and Mg, along with a low combined Zn+Mg+Cu content and a low Zn:Mg ratio leads to low to medium yield strength due to the limited volume fraction and number density of strengthening precipitates (GP zones, η' , η) [204]. While Medium Mn levels support the improved mechanical properties by forming manganese dispersoids that enhance dislocation glide, promote uniform slip, and increase strength without compromising elongation [216]. Also, a Cu/Mg ratio of 0.67 supports a balance between moderate strength and ductility by fine precipitation phases [210]. However, with medium levels of Fe and Si, coarse precipitates of $\text{Al}_7\text{Cu}_2\text{Fe}$ and Mg_2Si tend to form, and thus the alloys are crack-prone; these intermetallics reduce the availability of hardening causing elements (by binding Cu and Mg) and act as crack initiation sites [217]. Additionally, the absence

of sufficient anti-recrystallising agents, such as Ti, Cr, and Zr, results in coarse grains that further diminish strength and modulus of toughness [205].

The excellent SCC resistance is partly due to the low volume fraction of nano-scale precipitates η , η' , and T' , and the low Zn+Mg+Cu content, which together minimise anodic dissolution at the grain boundaries. Additionally, the medium Cu content improves SCC resistance by promoting the formation of Cu-rich grain boundary precipitates that reduce electrochemical activity related to hydrogen generation rates. Cu dissolution in the matrix also increases nobility, decreasing hydrogen embrittlement and slowing crack growth, leading to significant enhancements in SCC resistance [208].

The weldability is good to excellent, governed primarily by three factors: grain refinement, segregation of alloying elements, and the solidification range. Grain refinement reduces the risk of solidification cracking, with Mn promoting the formation of globular grains, which leads to finer grains. However, the refinement effect is limited due to the low concentrations of Cr, Ti, and Zr [212].

The low Zn content and low combined Zn+Mg+Cu reduce the formation of anodic $MgZn_2$ precipitates, minimising the likelihood of solidification-related cracking. The low Zn:Mg ratio contributes to a narrower solidification interval, further enhancing weldability. Conversely, medium Cu levels may slightly increase the risk of grain boundary cracking due to the potential formation of coarse intermetallic. Overall, the combination of low Zn, low total Zn+Mg+Cu, and moderate Mn content provides a favourable balance, resulting in good to excellent weldability [212], [213].

6.5 Conclusions on 7xxx Series Optimisation

In this chapter, we applied the same methodology used for the 6xxx series aluminium alloys to optimise the 7xxx series alloys, yielding comparably robust outcomes for the 7xxx series dataset. Initially, we collected 78 datasets encompassing chemical compositions, mechanical, corrosion, and weldability properties. This dataset was divided into two subsets: 54 high-Cu-containing alloys and 24 low-Cu-containing alloys. Principal Component Analysis (PCA) and K-means clustering were then performed on both subsets, resulting in four distinct clusters for the high-Cu subset and three clusters for the low-Cu subset.

Sub-clustering analysis was subsequently conducted to identify more compact property ranges within each cluster, producing 11 sub-clusters for the high-Cu-containing subset and six sub-clusters for the low-Cu-containing subset. The optimisation algorithm was then applied to identify the alloys exhibiting the best combination of properties within each sub-cluster. For the high-Cu subset, 11 optimum alloys were selected, while 6 optimum alloys were identified for the low-Cu-containing subset. These optimum alloys demonstrated superior properties compared to others within the same sub-clusters, enabling a reduction of over 50% in the number of 7xxx series aluminium alloys in each subset.

The recycling algorithm was subsequently applied to recycle alloys within the sub-clusters of both subsets. Mixing ratios were determined based on the TOPSIS closeness scores, assigning at least 50% to the optimum alloys identified by the optimisation algorithm, with the remaining 50% distributed among the other alloys. This procedure generated optimised compositions and properties alloys, further corroborating the findings of the optimisation algorithm.

Validation of the optimum alloys was performed using an unseen database from the MatWeb website, which had not been included in our study. This validation generally demonstrated less than 1 MPa error in yield strength and fatigue strength properties, and less than 0.001% error in compositions.

Furthermore, the clustering results were corroborated through metallurgical reasoning. For both high- and low-Cu-containing subsets, we applied the LIME algorithm to interpret the “black-box” clustering results, alongside literature-based metallurgical analysis, which produced consistent findings. The clustering was primarily governed by the sum and ratios of the principal alloying elements namely (Cu+Mg+Zn) and Zn:Mg. For example, Cluster 2 of the high-Cu subset comprises high-strength alloys, where the elevated solute content (Cu+Mg+Zn) promotes the formation of the high η' phase, whereas Cluster 1 of the low-Cu subset exhibits medium strength owing to a medium-high Mg+Zn content, which also leads to medium η' phase formation.

CHAPTER 7: CASE STUDY – OPTIMISATION OF 2xxx, 3xxx, AND 5xxx SERIES ALUMINIUM ALLOYS AND MIXED SCRAP RECYCLING COMPARISON

In this chapter, we applied our optimisation framework, as outlined in **Chapter 3: Methodology of Study**, to the 2xxx, 3xxx, and 5xxx series aluminium alloys, following the same procedure previously employed for the optimisation of the 6xxx and 7xxx series alloys in **Chapters 5 and 6**. The aim was to evaluate the effectiveness and utility of our approach across different aluminium series, each possessing distinct characteristics. To this end, we focused on refining the 2xxx, 3xxx, and 5xxx series alloys and identifying the optimum alloy within each series. These series were not examined in as much depth as the 6xxx series; rather, the primary objective was to assess the broad applicability of our design framework and to analyse the global results for these alloy series.

The procedure commences with the collection of data on the chemical composition, mechanical properties, corrosion properties, and technological properties of the 5xxx and 3xxx series alloys, with the addition of creep properties for the 2xxx series alloys. These data were subsequently subjected to clustering using a combined approach of principal component analysis (PCA) and K-means clustering (see **Section 3.4**), followed by sub-clustering as described in **Section 3.5**, in order to group alloys exhibiting similar property ranges.

Following clustering and sub-clustering, an optimisation algorithm (**Section 3.6**) was employed to identify the optimum alloy within each sub-cluster. The objective of this process was to select alloys that displayed a balanced combination of properties within their respective sub-clusters. Furthermore, a recycling algorithm (see **Section 3.7**) was applied within each sub-cluster to refine the selection, ultimately yielding the optimum alloys predicted by the optimisation algorithm.

In addition, a comparative analysis was performed between the recycling of mixed industrial scrap from the 3xxx, 5xxx, and 6xxx series alloys and the recycling of the optimum alloys selected by our design framework (see **Section 3.2**) for the same series. This analysis assessed the potential for cross-series recycling using our design framework, highlighting opportunities to leverage optimum alloys for sustainable alloy reuse.

7.1 Data Collection for 2xxx, 3xxx, and 5xxx Series Aluminium Alloys

The analysis began with the comprehensive collection of data on 59 alloys from the 2xxx series, 36 alloys from the 3xxx series, and 71 alloys from the 5xxx series. The database encompasses chemical compositions, mechanical properties, corrosion, and weldability characteristics. Tempering and hardening conditions have also been incorporated to reflect the most utilised industrial standards for each alloy series.

As the 2xxx series alloys are heat-treatable, data for this series were collected under the T6 tempering condition at 210°C, appropriate for high-temperature applications. In contrast, the 3xxx and 5xxx series alloys, which undergo work hardening rather than heat treatment, were analysed under the H32 hardening condition, the most prevalent for these alloys.

For the 2xxx series, the composition was determined by averaging the minimum and maximum values of Cu, Si, Ni, Ag, Mg, and Fe, while the maximum value of Mn was considered due to its positive effect on mechanical properties. These alloys are primarily used in structural applications across a wide temperature range, although they exhibit limited corrosion resistance and weldability. Consequently, mechanical properties such as creep and rupture strength were prioritised during data collection for this series.

For the 3xxx series alloys, the composition was computed as the average of Mn, Mg, Si; half of the maximum values of Fe; and a high Zn, Ti, Cr content. For the 5xxx series, the composition was determined by taking the maximum of Si, Cu, Zn, Cr, and Ti; averaging the contents of Mg and Mn; and using half of the maximum values of Fe.

Unlike the 2xxx series, the 3xxx and 5xxx series alloys are not intended for high-temperature applications but are valued for their high corrosion resistance and good weldability. Therefore, data collection for these alloys focused on corrosion resistance, weldability, and mechanical properties. Table 7.1 presents the chemical compositions and properties of the 2xxx, 3xxx, and 5xxx series alloys, with “N/A” indicating properties not considered for a given alloy series.

Table 7.1. Chemical compositions and properties range’s for 2xxx, 3xxx and 5xxx series data

Series	Comp %	Strength (MPa)	EI (%)	MoT (MPa %)	Corrosion Resistance	Weldability	Temp/ Hard
2xxx	Si: 0.05–1.3, Fe: 0.05–1.6, Cu: 0.9–	YS: 140–462,	10–27	26.5–54.69	N/A	N/A	T6

	6.3, Mn: 0–1.3, Mg: 0–1.55, Ni: 0–2, Ag: 0–0.35	Rupture strength: 190–330, Creep strength: 170–325					
3xxx	Si: 0.08–1.8, Fe: 0.05–0.5, Cu: 0.03–0.9, Mn: 0.15–1.5, Mg: 0–1.5, Cr: 0–0.2, Zn: 0–1, Ti: 0–0.175	YS: 34–170, Fatigue strength: 34–120	3.45–19.57	12.95–28.62	Fair–Excellent	Good–Excellent	H32
5xxx	Si: 0–0.7, Fe: 0–0.4, Cu: 0–0.6, Mn: 0–1.4, Mg: 0.4–5.6, Cr: 0–0.25, Zn: 0–2.8, Ti: 0–0.1	YTS: 160–310, Fatigue strength: 90–180	13–29	12.95–28.62	Good–Excellent	Good–Excellent	H32

Following data collection, the subsequent sections present the global optimisation and recycling results for each series, detailing the outcomes for the 2xxx, 3xxx, and 5xxx alloys, including the selected optimum alloys, their key properties, and the associated recycling mixing ratios.

7.2 Optimisation of 2xxx series Aluminium Alloys

The 2xxx series aluminium alloys are widely used in aerospace, automotive, and structural applications due to their high-temperature performance, including excellent creep and rupture strength. Their properties are strongly influenced by composition and microstructure, making them well-suited for systematic optimisation.

To facilitate the identification of optimum alloys with balanced properties, this series was initially analysed and grouped according to similarities in composition and performance characteristics. This clustering established a structured framework for subsequent detailed optimisation.

The following sub-sections presents the optimisation of the 2xxx series, including the selection of optimum alloys (**Section 3.6**), recycling algorithm (**Section 3.7**), and metallurgical insights derived from the analysis.

7.2.1 Optimum Alloy Selection in 2xxx Series

The alloys were subjected to clustering and sub-clustering, as described in **Sections 3.4** and **3.5** respectively, resulting in 6 primary clusters and 15 sub-clusters, as detailed in **Appendix Section 11.2.1**. Following the refinement of these sub-clusters into more compact property ranges, the TOPSIS optimisation algorithm, as described in **Section 3.6**, was applied to identify the alloys whose closeness scores were nearest to the ideal-best solution, representing the best combination of properties. This procedure ensured that the selected optimum alloys exhibited a balanced combination of properties closely matching the ideal-best solution.

These alloys are particularly noted for their high-temperature performance, including creep and rupture strength, and are widely employed in aerospace, automotive, and structural applications. Table 7.2 presents the optimum alloys from each cluster and sub-cluster, along with their corresponding properties and potential applications across various sectors. The results highlight the broad property range exhibited and demonstrate the alloys' versatility and effectiveness in diverse industrial contexts.

Table 7.2. Optimum alloys for 2xxx series aluminium alloys

Cluster	Sub-cluster	Optimum Alloy	Key Properties	Primary Applications
Cluster 0	Sub-cluster 01	2124	High yield strength, Creep, and rupture strength, and modulus of toughness	Wing skins, fuselage bulkheads. high-performance military aircraft, moderately elevated temperature applications [218], [219], [196]
	Sub-cluster 02	2027	High yield strength and modulus of toughness, medium rupture and creep strength	Construction, transportation, energy, and wing skin applications [217], [220], [7]
	Sub-cluster 03	2021	High yield, rupture and creep strength, and medium modulus of toughness	Automotive, aerospace manufacturing [221]
Cluster 1	Sub-cluster 11	2117	Low yield strength, modulus of toughness, moderate creep and rupture strength	Aerospace applications [222]
	Sub-cluster 12	2013	Moderate yield, creep and rupture strength, and high modulus of toughness	Automotive components, aerospace fittings, precision electronic parts [1], [184]
	Sub-cluster 13	2006	Moderate yield, creep, and rupture strength and high modulus of toughness	Aerospace, automotive, marine hardware, and construction [1], [184]
	Sub-cluster 14	2010	High to moderate yield, creep, and rupture	Automotive, construction, and energy [1], [223], [224]

			strength and modulus of toughness	
Cluster 2	Sub-cluster 21	2218	Moderate yield strength, high creep, rupture strength and low modulus of toughness	Transportation and structural materials, such as building facades and window frames [225], [196]
	Sub-cluster 22	2018	Low to moderate yield, creep, and rupture strength and low modulus of toughness	Manufacture of aircraft engine cylinders and heads and pistons [196]
		2618	Moderate yield, creep, and rupture strength and high modulus of toughness	Aerospace, automotive, and other high-stress applications [226], [196]
Cluster 3	Sub-cluster 31	2012	Moderate yield strength, high creep, rupture strength and low modulus of toughness	Free machining applications, Hydraulic components, transportation [7], [224], [227]
	Sub-cluster 32	2011	Low yield strength, high creep, and rupture strength and high modulus of toughness	Machine parts, automotive components, weaponry & munitions, fasteners / fittings [224], [228], [196]
Cluster 4	Sub-cluster 41	2040	High strength, moderate creep, and rupture strength and high modulus of toughness	3D printers, plotters, housing and other lightweight constructions, aerospace, defense [229], [184]
	Sub-cluster 42	2029	High-moderate strength creep, rupture strength and moderate modulus of toughness	Aerospace application, particularly used for fuselage and wing structures [230]
Cluster 5	Sub-cluster 51	2122	Moderate strength creep, and rupture strength and modulus of toughness	Aerospace applications [1]
	Sub-cluster 52	2001	High strength and moderate creep, and rupture strength and high modulus of toughness	Automotive, marine, and aerospace [1], [184]
		2004	High to moderate strength and moderate creep, rupture strength and modulus of toughness	Aerospace industry [231], [184]

7.2.2 Recycling Algorithm for Alloy Mixing of 2xxx Series Alloys

After identifying the optimum alloys in each sub-cluster using the optimisation algorithm (see **Section 3.6**), the next step involved mixing alloys within each sub-cluster using recycling algorithm (see **Section 3.7**), so that the resulting recycled alloy compositions and properties closely approximated those of the optimum alloys.

In this approach, the optimum alloy within each sub-cluster was assigned a minimum mixing ratio of 50%, with the remaining proportion distributed among the other alloys in the same sub-cluster according to their TOPSIS closeness scores, ensuring that the total mixing ratio summed to 100%. Blending the alloys in this way produced a recycled composition and property set closely matching those of the optimum alloys identified in **Section 7.2.1**.

Certain alloys exhibiting a unique combination of properties were found exclusively within specific sub-clusters. These were retained in their original form, without mixing, and designated as optimum alloys for targeted applications, due to their superior performance in niche scenarios where their distinct characteristics offered particular advantages.

Table 7.3 presents the mixing ratios for each alloy within the sub-clusters. The table also demonstrates a notable improvement in overall properties across the sub-cluster, emphasising the effectiveness of alloy mixing in enhancing performance for a wide range of industrial applications.

Table 7.3. Mixing ratios and key property improvements within the sub-clusters

Cluster	Sub-cluster	Optimum alloy mixing ratio ($\geq 50\%$)	Other mixed alloys (remaining)	Resultant recycled alloy properties
Cluster 0	Sub-cluster 01	2124: 0.897	2023: 0.0057, 2224: 0.0035, 2324: 0.0066, 2424: 0.0126, 2524: 0.0228, 2624: 0.0179, 2724: 0.0204, 2824: 0.0066, 2026: 0.0068	High yield strength, Creep, and rupture strength, and modulus of toughness
	Sub-cluster 02	2027: 0.5861	2009: 0.1551, 2034: 0.1425, 2056: 0.1163	High yield strength and modulus of toughness, medium rupture and creep strength
	Sub-cluster 03	2021: 0.9513	2024: 0.0487	High yield, rupture and creep strength, and medium modulus of toughness
Cluster 1	Sub-cluster 11	2117: 1	N/A	Low yield strength, modulus of toughness, moderate creep and rupture strength
	Sub-cluster 12	2013: 0.9057	2002: 0.0094, 2007: 0.0275, 2028: 0.0275, 2038: 0.0024, 2045: 0.0275	Moderate yield, creep and rupture strength, and high modulus of toughness
	Sub-cluster 13	2006: 0.9315	2005: 0.0088, 2008: 0.0085, 2015: 0.0103, 2017: 0.0058, 2025: 0.0068, 2030: 0.0108, 2033: 0.0065, 2044: 0.0109	Moderate yield, creep, and rupture strength and high modulus of toughness

	Sub-cluster 14	2010: 0.6921	2014: 0.0859, 2214: 0.0279, 2036: 0.0981, 2037: 0.0959	High to moderate yield, creep, and rupture strength and modulus of toughness
Cluster 2	Sub-cluster 21	2218: 0.9303	2032: 0.0697	Moderate yield strength, high creep, and rupture strength and low modulus of toughness
	Sub-cluster 22	2018: 0	N/A	Low to moderate yield, creep, and rupture strength and low modulus of toughness
	Sub-cluster 22	2618: 0.975	2031: 0.025	Moderate yield, creep, and rupture strength and high modulus of toughness
Cluster 3	Sub-cluster 31	2012: 0.8787	2111: 0.0427, 2219: 0.0359, 2041: 0.0427	Moderate yield strength, high creep, and rupture strength and low modulus of toughness
	Sub-cluster 32	2011: 1	N/A	Low yield strength, high creep, and rupture strength and high modulus of toughness
Cluster 4	Sub-cluster 41	2029: 0.8957	2039: 0.073, 2139: 0.0313	High-moderate strength creep, and rupture strength and moderate modulus of toughness
	Sub-cluster 42	2040: 1	N/A	High strength, moderate creep, and rupture strength and high modulus of toughness
Cluster 5	Sub-cluster 51	2122: 0.9153	2319: 0.0193, 2419: 0.0267, 2519: 0.0064, 2022: 0.0131, 2042: 0.0193	Moderate strength creep, and rupture strength and modulus of toughness
	Sub-cluster 52	2001: 1	N/A	High strength and moderate creep, and rupture strength and high modulus of toughness
	Sub-cluster 52	2004: 0	N/A	High to moderate strength and moderate creep, rupture strength and modulus of toughness

7.2.3 Metallurgical Reasoning Behind Clustering of 2xxx Series Alloys

The phase composition of Al–Cu–Mg alloys is largely governed by the Cu/Mg ratio, which plays a critical role in determining the sequence and stability of phase precipitation. Alloys with high Cu/Mg ratios typically exhibit two distinct precipitation sequences. The first follows the conventional sequence observed in Al–Cu alloys: supersaturated solid solution (SSSS) → Guinier–Preston (GP) zones (Al_3Cu) → θ'' (Al_3Cu) → θ' (Al_2Cu) → θ (Al_2Cu) [232]. The presence of Mg and Si introduces the possibility of forming additional minor phases, such as S (Al_2CuMg), Ω (Al_2Cu), and σ ($\text{Al}_5\text{Cu}_6\text{Mg}_2$) [4]. (Note: S and S' are difficult to distinguish due to their similar crystal structures and will henceforth be referred to collectively as the S phase).

The second sequence is initiated by the formation of $\{111\}\alpha$ clusters, which subsequently lead to the precipitation of the Ω phase (Al_2Cu) and ultimately the θ phase (Al_2Cu). The GP zones, θ'' , and θ' phases are critical to alloy strengthening, typically forming as disk- or plate-shaped

structures on the $\{001\}\alpha$ planes and thereby contributing significantly to the mechanical properties of the alloy.

In contrast, alloys with lower Cu/Mg ratios, and correspondingly higher Mg content, preferentially form the S' phase (Al_2CuMg). This phase originates from Cu–Mg clusters and may nucleate heterogeneously on dislocations or at interfaces with the T phase ($\text{Al}_{20}\text{Cu}_2\text{Mn}_3$), which itself forms through Mn addition. Both S' and θ' phases play a crucial role in enhancing the strength and ductility of the alloy, with θ' generally exhibiting a plate-like morphology and S' adopting a lath-like morphology. These phases are principally responsible for controlling the mechanical behaviour of the alloy, particularly its strength and modulus of toughness [4].

Additional alloying elements, including Fe, Ni, Ag, and Mn, further influence phase formation and the ultimate properties of Al–Cu–Mg alloys. Notably, an Ag content of 0.1–0.5% promotes Ω phase (Al_2Cu) development, enhancing age hardening and mechanical performance, particularly when combined with Mg. Minor additions of Ni and Fe improve thermal stability by forming aluminide particles such as Al_9FeNi , which enhance resistance to high-temperature degradation. Mn contributes to microstructural stability through T phase formation, thereby supporting overall alloy performance [232], [233].

These metallurgical insights underscore the critical importance of alloy composition and phase-forming mechanisms in optimising the mechanical behaviour of 2xxx series alloys. Such alloys are particularly vital for high-temperature applications, where enhanced strength, ductility, and thermal stability are required. A comprehensive analysis of clustering behaviour in relation to alloy composition and phase transformations provides valuable insights into how these elements influence phase stability and alloy properties, as discussed below.

Cluster 4 (High Strength Alloys, see Appendix Figure 11.3, and Table 11.1): The high strength of this group of alloys results from a combination of high Cu and Ag, medium Mg and Mn, and a moderate Cu/Mg ratio, which promotes the formation of Ω , S' , S'' , θ' , and T phases. The Ω phase contributes to enhanced strength, modulus of toughness, and creep strength through dispersion strengthening. It is more stable than θ' , possesses high antiphase boundary energy, and is therefore resistant to dislocation shearing.

High Ag content is particularly important as a catalyst for Ω -phase formation. Ag diffuses into vacancies in regions of high Mg content, thereby reducing the nucleation and misfit energies

with the Al matrix. Owing to the high binding energy between Mg and Ag atoms, Ag effectively traps Mg atoms, forming Mg–Ag co-clusters. These co-clusters gradually evolve into a well-defined Ω phase through enrichment with Cu atoms. This process facilitates a uniform and dense distribution of the Ω phase, suppresses its coarsening, and consequently strengthens the alloy. Mg is a key element for the nucleation of the Ω phase, while high Cu content further stabilises and refines it, thereby enhancing overall strength [232], [234].

The S-phase in these alloys develops from Cu–Mg clusters within the Al matrix and nucleates at dislocations and at interfaces between the matrix and the T phase ($\text{Al}_{20}\text{Cu}_2\text{Mn}_3$), a process promoted by medium Mn content [232]. However, the elongation of these alloys is reduced due to the dense distribution of secondary phase particles, such as $\text{Al}_{11}\text{Cu}_5\text{Mn}_3$ and Al_2Cu [235], [236]. Under tensile stress, these particles act as stress concentrators, leading to the formation of microcracks, which serve as sites for crack initiation. Cracks propagate rapidly along these particles, thereby reducing ductility. The T phase, present discontinuously along grain boundaries, also affects crack growth and further restricts elongation [236].

Cluster 2 (Medium Strength Alloys): This cluster is characterised by high levels of Si, Fe, Mg, and Ni; medium Cu; and low Mn content, which promotes the formation of Guinier–Preston–Bagaryatskii (GPB) zones, as well as S' , S, and θ phases [234]. The presence of Fe and Ni enhances microstructural stability under thermal exposure through the formation of intermetallic particles, such as Al_9FeNi aluminides, which help regulate grain size and impede dislocation movement, thereby improving the overall performance of the alloy [233].

Moderate θ'' and θ' phases form due to the intermediate Cu content, while S' , β'' (Mg_2Si), and σ ($\text{Al}_5\text{Cu}_6\text{Mg}_2$) phases develop as a consequence of the high Mg and Si content, resulting in medium strength. High Si content also facilitates S' phase precipitation, particularly when $\text{Cu/Mg} < 5$, by increasing vacancy concentration and stabilising GPB zones. This promotes nucleation sites for the S' (Al_2CuMg) phase and increases its volume fraction, thereby enhancing age-hardening. However, the coarsening and uneven distribution of these precipitates limit their overall contribution to hardening [232], [237].

The elevated Si content further encourages S-phase development, which inhibits grain growth at high temperatures and consequently enhances strength. Under applied stress, the size and density of S-phase precipitates increase, improving both strength and creep resistance.

Nevertheless, elongation is reduced due to microvoid formation and the potential for cleavage failure under high stress, which accelerates fracture [238].

Cluster 3 (Low-Moderate Strength Alloys): The high Cu and Fe content in this cluster of alloys leads to the formation of GP zones, θ , θ'' , Cu_2FeAl_7 , and $\text{Cu}_2\text{Mn}_3\text{Al}_{20}$ phases. Once the GP zones reach their critical size, they transform into the θ'' phase, a primary hardening phase responsible for strengthening these alloys. The fine θ'' precipitates effectively impede dislocation motion, thereby increasing both strength and yield strength. Although precipitates such as Cu_2FeAl_7 and $\text{Cu}_2\text{Mn}_3\text{Al}_{20}$ are present in the microstructure, they contribute minimally to the precipitation process and overall strengthening. Apart from Cu, these alloys contain very few additional strengthening elements, which limits their overall strength [239].

While some θ' and Ω phases also form, they do not significantly contribute to strength. Instead, they play a role in reducing crack propagation along second-phase particles, such as Al_2Cu and T phases, thereby improving elongation [237].

Cluster 0 (High Strength Alloys): The combination of medium Cu with high Mn and Mg in these alloys promotes the formation of multiple strengthening phases, including θ' , S' , and S'' . With a Cu/Mg ratio < 5 , the S' phase, characterised by a rod- or lath-shaped morphology, becomes the dominant strengthening phase. It nucleates heterogeneously on dislocations and at interfaces with the T phase ($\text{Al}_{20}\text{Cu}_2\text{Mn}_3$), a dispersoid promoted by high Mn content, thereby contributing significantly to the alloy's strength [236].

Although some θ' and Ω phases also form, they do not substantially increase strength. Instead, they assist in reducing crack propagation along second-phase particles, such as Al_2Cu and T phases, thereby improving elongation [237]. High Mg content further facilitates Ω -phase formation, which, together with the dense S' phase structure, enhances creep resistance. Additionally, the elevated Mg content reduces grain size and dendrite arm spacing, resulting in a fine and homogeneous grain structure that further strengthens these alloys [237].

Cluster 1 (Moderate Strength Alloys): This cluster comprises alloys with moderate Cu and Mg content, together with the presence of Mn and Si that promote the formation of several precipitate phases, including S' , θ'' , and θ' . The relatively high Cu/Mg ratio (>6) indicates that θ' is the dominant strengthening phase, although its quantity remains moderate due to the

intermediate Cu levels. Some S-phase precipitates also form; however, their quantity is limited owing to the low Mg content [240].

The high Si content contributes to structural refinement and increases the number of nucleation sites for S-phase (Al_2CuMg) precipitates, thereby enhancing their volume fraction and contributing to medium strength. Stable intermetallic particles, such as Al_9FeNi , help control grain size and impede dislocation movement, though these coarse intermetallic provide minimal precipitation strengthening [233].

The S' phase can nucleate heterogeneously on dislocations and at interfaces with the T phase ($\text{Al}_{20}\text{Cu}_2\text{Mn}_3$), with nucleation promoted by the high Mn content. Mn not only facilitates a finer θ' precipitate distribution but also slows their growth at elevated temperatures due to preferential segregation at Al/ θ interfaces. Consequently, the addition of Mn and Fe stabilises θ' precipitates, resulting in a fine precipitate distribution that mitigates crack propagation and improves alloy elongation [240].

Overall, the medium Cu and Mg content supports moderate phase formation, yielding moderate creep strength in these alloys.

Cluster 5 (Medium Strength Alloys): This group is characterised by high Cu and low Mg content, with a Cu/Mg ratio exceeding 18. Consequently, the θ' phase is primarily responsible for strengthening, significantly enhancing the alloy's strength by impeding dislocation glide. Although the high Cu content shifts the fracture mode from ductile transgranular to intergranular, both intergranular cleavage and dimple size increase with temperature exposure, leading to reduced plasticity in these alloys [241]. The θ' phase possesses semi-coherent interphase boundaries, which maintain overall strength despite plasticity falling below 20%.

Low Si and Mg contents reduce the quantity and size of quenched-in dislocation loops, limiting heterogeneous nucleation sites for the S' phase. Partial development of the Ω phase occurs due to the high Cu content; however, the absence of Mg restricts its formation, contributing only moderately to creep strength [232].

The presence of Mn promotes a finer and denser distribution of θ' precipitates while also slowing their growth, thereby enhancing resistance to coarsening. Fine $\text{Al}_{11}\text{Cu}_5\text{Mn}_3$ particles form a dense distribution. Under tensile stress, these particles act as sites for crack initiation and

propagation, resulting in reduced elongation due to rapid crack growth along second-phase particles. Although the low Mg content helps improve rupture strength by reducing closed porosity, overall plasticity remains compromised [234], [241].

7.2.4 Thermo-Calc Analysis as a Basis for Cluster Classification

For the 2xxx-series clusters, Thermo-Calc phase fraction analysis at single-point equilibrium was performed at 210 °C and atmospheric pressure under T6 tempering conditions. This computational approach provides insight into the equilibrium distribution of strengthening phases across the different clusters, enabling direct comparison with the precipitation trends identified in **Section 7.2.3**. By averaging phase formation and yield strength within each cluster, the results offer a robust validation of the metallurgical interpretations.

As shown in Figure 7.1, the low-strength clusters, such as Cluster 3, are dominated by the Al_2Cu phase, with only a minimal fraction of fine strengthening precipitates. This outcome reinforces the literature-based understanding that the absence of substantial precipitation strengthening leads to reduced mechanical performance. By contrast, high-strength clusters, such as Clusters 0 and 4, exhibit a significant fraction of the S phase together with other phases. The presence of these phases is consistent with metallurgical principles, as they act to hinder dislocation motion and thereby substantially increase yield strength.

Collectively, the Thermo-Calc results align closely with the literature-derived correlations, providing strong confirmation that the clustering approach reliably captures the metallurgical basis of strength variation across the 2xxx series clusters.

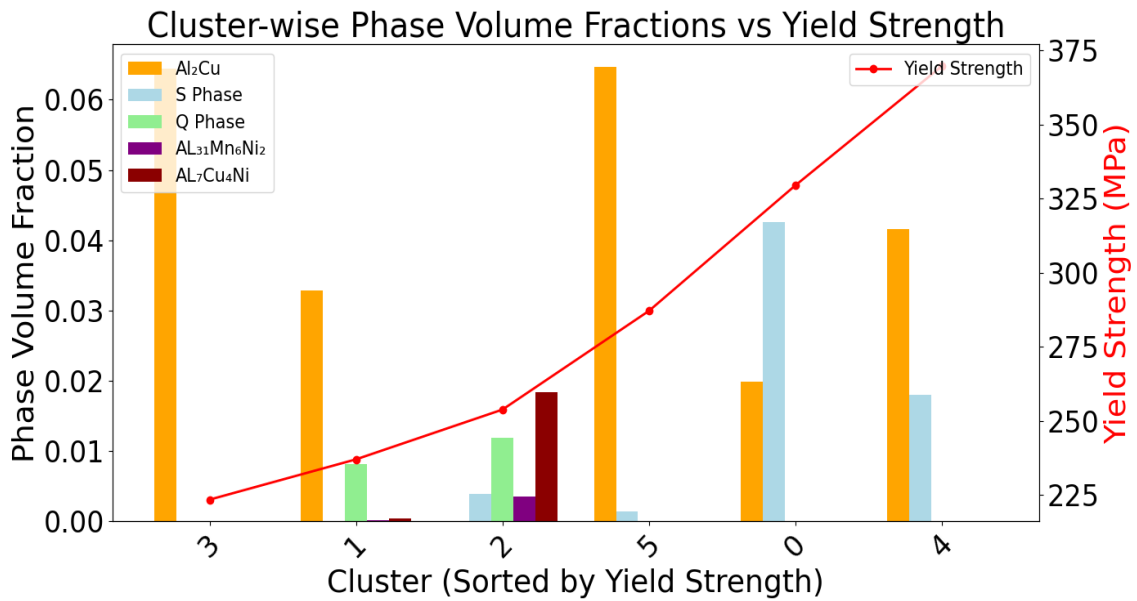


Figure 7.1. Thermo-Calc analysis of 2xxx series clustering

7.3 Optimisation of 5xxx Series Aluminium Alloys

The 5xxx series aluminium alloys are widely used in marine, automotive, and structural applications due to their excellent corrosion resistance and moderate-to-high strength. Their properties are predominantly governed by magnesium content, which strongly influences mechanical performance and workability, making systematic optimisation highly valuable.

To identify alloys with balanced properties, they were initially grouped according to composition and performance characteristics, establishing a structured framework for subsequent optimisation and recycling. The following sections present this optimisation, including the identification of optimum alloys, recycling mixing ratios and the metallurgical insights derived from this analysis.

7.3.1 Optimum Alloys Selection for the 5xxx Series Alloys

We conducted a detailed clustering (Section 3.4) and sub-clustering analysis (Section 3.5) of these alloys, as described in Appendix Section 11.2.2, which yielded 3 clusters and 11 sub-clusters. Following the sub-clustering analysis, the optimisation algorithm was applied, as described in Section 3.6. Each alloy within the sub-clusters was assigned a TOPSIS closeness score, reflecting its proximity to the ideal-best solution and its distance from the ideal-worst solution. Alloys positioned closest to the ideal-best solution and furthest from the ideal-worst solution were identified as optimum, exhibiting the most favourable combination of properties

and compositions. These alloys find extensive application across a range of industries, including marine, transportation, and structural sectors, due to their weldability, high strength, and excellent corrosion resistance. Table 7.4 presents the optimum alloys identified by the optimisation algorithm, together with their corresponding properties and applications.

Table 7.4. Optimum alloys for 5xxx series aluminium alloys

Cluster	Sub-cluster	Optimum Alloy	Key Properties	Primary Applications
Cluster 0	Sub-cluster 01	5043	Low-medium strength, low fatigue strength, moderate modulus of toughness, good weldability and excellent corrosion resistance	Filler material for welding aluminium parts [242]
	Sub-cluster 02	5140	Medium strength, fatigue strength, high modulus of toughness, excellent corrosion resistance, and weldability	Marine, Automotive, structural applications [1]
	Sub-cluster 03	5351	Moderate strength, high fatigue strength, excellent corrosion resistance, and weldability	Marine transportation, automotive applications [1]
	Sub-cluster 04	5449	Low strength, fatigue strength, modulus of toughness, good corrosion resistance, and excellent weldability	Door & window and industrial applications [243]
		5151	Low strength, fatigue strength, modulus of toughness, excellent corrosion resistance, and weldability	Boats, paneling and pressings, marine structures, aircraft parts, vehicle panels, furniture, tubing, silos, containers [243]
Cluster 1	Sub-cluster 11	5352	High strength, fatigue strength, modulus of toughness, excellent corrosion resistance, and weldability	Packaging, Transportation [244]
	Sub-cluster 12	5954	Superior strength, modulus of toughness, medium fatigue strength, good corrosion resistance, and excellent weldability	Marine and structural applications [1], [243]
	Sub-cluster 13	5483	High-Medium strength, fatigue strength, high modulus of toughness, good corrosion resistance, and excellent weldability	Transportation, construction, food and chemical processing, packaging [245]
	Sub-cluster 14	5083	High strength, medium fatigue strength, high modulus of toughness, moderate corrosion resistance, and excellent weldability	Shipbuilding, vehicle bodies, pressure vessels, and rail cars [246], [196]
Cluster 2	Sub-cluster 21	5556	Medium-low strength, medium fatigue strength, low modulus of toughness, excellent corrosion resistance, and weldability	Filler metal for welding [247], [196]
	Sub-cluster 22	5052	Medium strength, high fatigue strength, medium modulus of toughness, good corrosion resistance, and excellent weldability	Marine, automotive, aerospace applications [248]

	Sub-cluster 23	5051	Low strength, medium fatigue strength, modulus of toughness, good corrosion resistance, and excellent weldability	Manufacturing [249]

7.3.2 Recycling Algorithm for Alloy Mixing of 5xxx Series Alloys

The alloys within each sub-cluster were recycled using the recycling algorithm described in **Section 3.7**. This process involved mixing the alloys according to the mixing ratios calculated from their TOPSIS closeness scores obtained via the optimisation algorithm (**Section 3.6**), with at least 50% allocated to the optimum alloys and the remaining 50% distributed among the other alloys within the sub-cluster based on their respective TOPSIS scores, thereby enhancing recyclability and producing recycled alloys whose compositions and properties closely approximated those of the optimum alloy.

However, some alloys within the sub-clusters could not be mixed due to their unique combinations of composition and properties; these alloys remained separate. Table 7.5 presents the mixing ratios of alloys within sub-clusters alongside their corresponding optimum alloys. The resulting recycled alloys exhibit minimal deviations in composition and properties from the optimum alloys identified by the optimisation algorithm, demonstrating that the mixing process effectively optimises the alloys within each sub-cluster.

Table 7.5. Mixing ratios of 5xxx series alloys

Cluster	Sub-cluster	Optimum alloy (OA) ($\geq 50\%$)	Other mixed alloys (Remaining)	Resultant recycled alloy properties
Cluster 0	Sub-cluster 01	5043: 0.6399	5006: 0.0619, 5106: 0.0478, 5017: 0.049, 5040: 0.0465, 5042: 0.0333, 5249: 0.0155, 5349: 0.0648, 5150: 0.0413	Low-medium strength, low fatigue strength, moderate modulus of toughness, good weldability and excellent corrosion resistance
	Sub-cluster 02	5140: 0.9509	5018: 0.0083, 5041: 0.01, 5451: 0.0093, 5254: 0.0051, 5354: 0.0098, 5454: 0.0066	Medium strength, fatigue strength, high modulus of toughness, excellent corrosion resistance, and weldability
	Sub-cluster 03	5351: 0.9105	5016: 0.0149, 5021: 0.0142, 5028: 0.0093, 5049: 0.0137, 5149: 0.01, 5654: 0.0104, 5754: 0.0169	Moderate strength, high fatigue strength, excellent corrosion resistance, and weldability
	Sub-cluster 04	5449: 0	N/A	Low strength, fatigue strength, modulus of toughness, good corrosion resistance, and excellent weldability

		5151: 1	N/A	Low strength, fatigue strength, modulus of toughness, excellent corrosion resistance, and weldability
Cluster 1	Sub-cluster 11	5352: 0.9183	5554: 0.034, 5854: 0.0477	High strength, fatigue strength, modulus of toughness, excellent corrosion resistance, and weldability
	Sub-cluster 12	5954: 0.9633	5119: 0.0026, 5023: 0.0027, 5026: 0.0026, 5056: 0.0034, 5356: 0.0025, 5456: 0.0028, 5657: 0.0027, 5058: 0.0017, 5182: 0.0026, 5183: 0.0031, 5283: 0.0026, 5383: 0.0021, 5186: 0.0018, 5087: 0.0018, 5187: 0.0018	Superior strength, modulus of toughness, medium fatigue strength, good corrosion resistance, and excellent weldability
	Sub-cluster 13	5483: 0.9597	5022: 0.0096, 5027: 0.0039, 5059: 0.0081, 5070: 0.009, 5086: 0.0097	High-Medium strength, fatigue strength, high modulus of toughness, good corrosion resistance, and excellent weldability
	Sub-cluster 14	5083: 0.9421	5024: 0.0179, 5180: 0.0279, 5082: 0.0121	High strength, medium fatigue strength, high modulus of toughness, moderate corrosion resistance, and excellent weldability
Cluster 2	Sub-cluster 21	5556: 0.5298	5205: 0.019, 5305: 0.0418, 5505: 0.0418, 5605: 0.0418, 5110: 0.0418, 5210: 0.0418, 5310: 0.0418, 5050: 0.0471, 5251: 0.0412, 5257: 0.0261, 5457: 0.0262, 5557: 0.0596	Medium-low strength, medium fatigue strength, low modulus of toughness, excellent corrosion resistance, and weldability
	Sub-cluster 22	5052: 0.7421	5252: 0.2579	Medium strength, high fatigue strength, medium modulus of toughness, good corrosion resistance, and excellent weldability
	Sub-cluster 23	5051: 1	N/A	Low strength, medium fatigue strength, modulus of toughness, good corrosion resistance, and excellent weldability

7.3.3 Metallurgical Reasoning Behind Clustering of 5xxx Series Alloys

In the non-age-hardenable 5xxx series alloys, the addition of up to 6% Mg provides solute hardening, which is complemented by efficient strain hardening, rendering them among the strongest non-age-hardenable aluminium alloys while maintaining excellent formability. Work hardening strengthens alloys by increasing the density of dislocations, whereas solid-solution hardening arises when solute atoms create strain fields due to size and modulus mismatches with the aluminium matrix. The initial strength of these alloys results from a combination of

solid-solution hardening, primarily due to solute Mg, and dispersion hardening, facilitated by Mn and other dispersoid-forming elements such as Fe.

The principal strengthening phases in these alloys are Al_6Mn and Al_3Fe , with Fe and Mn able to substitute for one another within these phases across a wide compositional range. Mn plays a particularly significant role in enhancing mechanical strength and intergranular corrosion resistance. The solubility of Mg in the solid α -Al phase is 17.4% in the binary Al–Mg system, decreasing to 14% in ternary Al–Mg–Mn alloys. At higher Fe and Mn concentrations, the Fe-rich phases shift from $\text{Al}_6(\text{FeMn})$ to $\alpha\text{-Al}(\text{FeMn})\text{Si}$ (α -Fe) phases [250].

Although Mn contributes to solid solution strengthening and improved mechanical properties, excessive Mn can be detrimental by increasing the size and volume fraction of $\text{Al}_6(\text{FeMn})$ phases. Consequently, Mn additions are generally limited to below 1.0% to maintain an appropriate balance.

In addition to Mg and Mn, trace elements and impurities such as Si, Cu, and Zn may also be present. While these elements are not intentionally added as major alloying constituents, they can influence phase formation and properties. For instance, Zn may promote the formation of the T-phase ($\text{Mg}_{32}(\text{Al,Zn})_{49}$), which can contribute to certain strengthening effects, whereas Ti-containing particles such as $\text{Al}_{18}\text{Ti}_2\text{Mg}_3$ and TiAl_3 can assist in grain refinement and overall alloy stability [251]. However, excessive levels of these impurities may adversely affect mechanical performance and corrosion resistance.

This metallurgical investigation demonstrates a clear correlation between composition, structural features, and properties, as shown below, providing support for the clustering results and illustrating how Mg–Mn interactions govern the characteristics of the 5xxx series alloys within each cluster.

Cluster 1 (High-Strength Alloys, see Appendix Figure 11.6 and Table 11.2): This cluster comprises alloys characterised by high Mg and moderate Mn and Zn contents, resulting in elevated strength, moderate corrosion resistance, and reasonable weldability. The high Mg content promotes the formation of strengthening phases such as $\beta\text{-Al}_8\text{Mg}_5$ and Mg_2Si (the latter formed due to Si impurities), which impede dislocation mobility and facilitate effective work hardening [252]. Additionally, the presence of moderate Zn enhances the formation of the MgZn_2 phase, which further strengthens these alloys by pinning dislocations [253].

Mn plays a critical role in grain refinement through the formation of Al_6Mn dispersoids, which increase grain boundary surface area and contribute to a more uniform distribution of intermetallic precipitates within the alloy matrix [252]. Grain refinement is further aided by the presence of Cr and Ti, which lead to the formation of $\epsilon\text{-Al}_{18}(\text{Cr,Mn})_2\text{Mg}_3$, as well as Ti-containing particles such as $\text{Al}_{18}\text{Ti}_2\text{Mg}_3$ and TiAl_3 , all of which refine the grain structure and enhance alloy hardness [254], [255].

Corrosion resistance in these alloys is generally good. Mg_2Si and Al_3Mg_2 phases act as anodic sites relative to the matrix, leading to a minor increase in corrosion [255]. Conversely, the MgZn_2 phase, with a self-corrosion potential close to that of the aluminium matrix, mitigates potential differences and reduces the likelihood of corrosion cell development [252].

Weldability is moderate due to the relatively high Fe, Cu, and Mg contents, which promote the formation of β -phases and intermetallics such as $\text{Al}_6(\text{Mn,Fe})$, Al_6Mn , and Al_3Fe , all of which are susceptible to cracking during welding [255], [252].

Cluster 0 (Moderate Strength Alloys): This cluster comprises medium-strength alloys characterised by moderate levels of Mg, Mn, and Cr, along with trace elements such as Si, Zn, and Cu. These elements facilitate the formation of the Mg_2Si phase, which impedes dislocation motion and enhances alloy strength [256]. The moderate Mg content supports Mg_2Si precipitation, thereby pinning dislocations and further increasing strength [253]. Mn contributes to grain refinement through the formation of Al_6Mn dispersoids, which increase grain boundary area and promote a uniform distribution of intermetallic precipitates within the matrix [252]. Cr, in the form of $\epsilon\text{-Al}_{18}(\text{Cr,Mn})_2\text{Mg}_3$, together with Ti-containing particles such as ($\text{Al}_{18}\text{Ti}_2\text{Mg}_3$, TiAl_3), refines the grain structure, producing a finer and stronger microstructure [254], [255], [256], [257]. The potentially detrimental effects of high Fe content are mitigated by Mn through the formation of less harmful $\text{Al}_6(\text{Mn,Fe})$ intermetallics [252]. Overall, the moderate levels of these elements maintain the alloy strength at a moderate level.

Corrosion resistance is moderate, as Mg_2Si and Al_3Mg_2 phases act as anodic sites relative to the matrix, leading to a slight acceleration of corrosion [253].

Weldability is also moderate, owing to the presence of Cu, Fe, and Mg at medium levels, which promote the formation of β -phase and other intermetallic precipitates such as $\text{Al}_6(\text{Mn,Fe})$,

Al_6Mn , and Al_3Fe . These phases increase the susceptibility of the alloys to cracking during welding [255].

Cluster 2 (Low Strength Alloys): The alloys in this cluster exhibit limited properties due to their low concentrations of alloying elements. With fewer alloying additions, there are correspondingly fewer strengthening phases to impede dislocation motion or reinforce the grain structure, resulting in reduced strength, hardness, and other mechanical properties. The lower alloying content also adversely affects both corrosion resistance and weldability.

7.3.4 Thermo-Calc Analysis as a Basis for Cluster Classification

Thermo-Calc single-point equilibrium analysis was conducted to validate the literature-based metallurgical rationale underlying the clustering obtained via the combined PCA and K-means clustering approach (see Section 3.4) for the 5xxx series alloys in the H32 condition and atmospheric pressure. The analysis involved plotting the average phase volume fraction against the average yield strength for each cluster. As illustrated in Figure 7.2, owing to this averaging, Cluster 2 appears to exhibit higher strength than Cluster 0, despite the latter having a higher phase volume fraction. In reality, the presence of the T-phase and Al_6Mn dispersoids in Cluster 0 confers greater strength; however, this is obscured by the averaging, making the curve appear lower. Cluster 1 demonstrates high strength, consistent with the formation of strengthening phases such as T (Al_2Mg_3) and $\text{Al}_{18}\text{Mg}_3\text{Ti}_2$. The phase distributions in these clusters show good correlation with their respective mechanical properties, confirming both the effectiveness and the metallurgical relevance of the clustering approach.

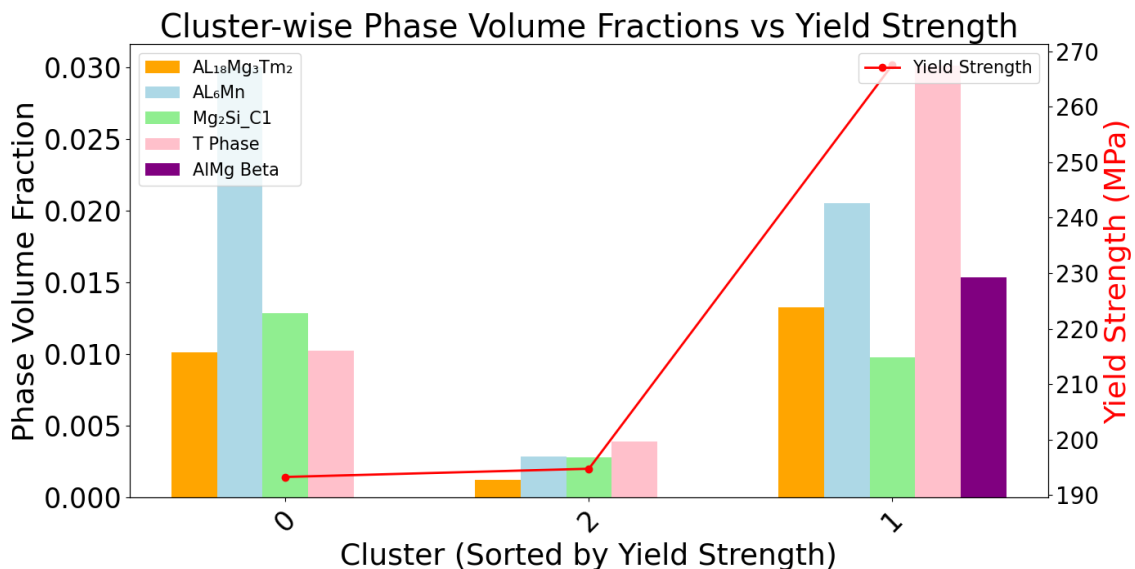


Figure 7.2. Thermo-Calc analysis of clusters of 5xxx series alloys

7.4 Optimisation of 3xxx Series Aluminium Alloys

The 3xxx series aluminium alloys are widely used in automotive, construction, and packaging applications due to their excellent corrosion resistance, moderate strength, and good formability. Their properties are strongly influenced by composition and microstructural characteristics, making them ideal candidates for systematic optimisation aimed at achieving an optimal balance of mechanical performance and recyclability.

To identify alloys with optimised properties, the series was initially analysed and grouped into clusters and sub-clusters based on compositional and property similarities. This clustering established a structured framework for subsequent detailed optimisation and recycling.

The following sections present the optimisation of the 3xxx series, including the selection of optimum alloys (see **Section 3.6**), the implementation of the recycling algorithm (see **Section 3.7**), and the metallurgical insights derived from the analysis.

7.4.1 Optimum Alloy Selection for the 3xxx Series

We first performed the clustering (**Section 3.4**) and sub-clustering (**Section 3.5**) of the 3xxx series alloys, as detailed in **Appendix Section 11.2.3**. This process resulted in 4 distinct clusters and 9 sub-clusters.

Next, the optimisation algorithm, was applied within each sub-cluster. Each alloy was assigned a closeness score, reflecting how closely it approximated the ideal combination of property solutions. Alloys with the highest closeness scores within each sub-cluster were selected as the optimum alloys, as they represented the most balanced combination of properties within their respective sub-clusters. Table 7.6 presents the properties of the selected optimum alloys of the 3xxx series, including corrosion resistance, strength, weldability, and modulus of toughness, along with their broad range of industrial applications across sectors such as architecture, food processing, and transportation.

Table 7.6. Optimum alloy for 3xxx series aluminium alloys

Cluster	Sub-cluster	Optimum Alloy	Key Properties	Primary Applications
Cluster 0	Sub-cluster 01	3307	Medium strength, fatigue strength, medium modulus of toughness, excellent weldability, corrosion resistance	Architectural and general industrial applications [1]
	Sub-cluster 02	3020	Low strength, fatigue strength, low modulus of toughness, excellent weldability, corrosion resistance	Architectural applications [1]
Cluster 1	Sub-cluster 11	3103	Medium strength, fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance	Heat exchangers, air condition evaporators, aircraft and military components, motor vehicle radiators, freezer linings, cladding alloy, office equipment, tubing and piping [258], [196]
	Sub-cluster 12	3009	Low-medium strength, fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance	Packaging, general industrial applications [1]
Cluster 2	Sub-cluster 21	3102	Low strength, fatigue strength, high modulus of toughness, good weldability, excellent corrosion resistance	Kitchenware, condenser tubes, and heat exchangers [259], [196]
	Sub-cluster 22	3130	Medium strength, fatigue strength, high modulus of toughness, fair weldability, good corrosion resistance	Transportation: Used in aircraft, military components, vehicle radiators, and power transmission lines, Construction: Used in heat exchangers, heating and cooling systems, and corrugated roofing Food packaging: Used in drinks cans and other food packaging applications, Kitchenware: Used in cooking utensils and kitchen equipment, Electrical: Used in domestic wiring and power transmission lines, other applications: Used in freezer linings, office equipment, tubing, piping, pressure vessels, and signs [1], [228]
Cluster 3	Sub-cluster 31	3403	Moderate strength, high fatigue strength, low modulus of toughness, good weldability, excellent corrosion resistance	Marine, construction, heat exchangers, industrial machinery components [1]
	Sub-cluster 32	3104	Superior strength, medium fatigue strength, low modulus of toughness, good weldability, excellent corrosion resistance	Manufacture of the body section of full-aluminium beverage cans [260]
	Sub-cluster 33	3016	High strength, moderate fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance	Transportation tools, door & window, glass wall, heat sink, and decorations [1], [243]

7.4.2 Recycling Algorithm for Alloy Mixing of the 3xxx Series Alloys

The recycling potential within the 3xxx series sub-clusters was assessed using the optimisation algorithm in conjunction with the recycling algorithm (See **Section 3.7**), which determined the mixing ratios of alloys within each sub-cluster to achieve an optimal balance of properties and composition, as described in **Section 3.7**. Table 7.7 lists the alloys in each sub-cluster along with their corresponding mixing ratios, highlighting the key improvements in the overall properties of the sub-clusters.

Table 7.7. Mixing ratios of 3xxx series sub-clusters

Cluster	Sub-cluster	Optimum alloy (OA)	Other mixed alloys	Resultant recycled alloy properties
Cluster 1	Sub-cluster 11	3307: 0.9608	3001: 0.0085, 3007: 0.0081, 3011: 0.0051, 3017: 0.007, 3021: 0.0059, 3065: 0.0046	Medium strength, fatigue strength, medium modulus of toughness, excellent weldability, corrosion resistance
	Sub-cluster 12	3020: 0.8754	3012: 0.1246	Low strength, fatigue strength, low modulus of toughness, excellent weldability, corrosion resistance
Cluster 0	Sub-cluster 01	3103: 0.9287	3003: 0.0373, 3203: 0.0341	Medium strength, fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance
	Sub-cluster 02	3009: 0.9369	3303: 0.0315, 3107: 0.0315	Low-medium strength, fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance
Cluster 2	Sub-cluster 21	3102: 1	N/A	Low strength, fatigue strength, high modulus of toughness, good weldability, excellent corrosion resistance
	Sub-cluster 22	3130: 0.8833	3002: 0.0176, 3207: 0.0188, 3010: 0.0234, 3110: 0.0246, 3026: 0.0237, 3030: 0.0085	Medium strength, fatigue strength, high modulus of toughness, fair weldability, good corrosion resistance
Cluster 3	Sub-cluster 31	3403: 0.9215	3105: 0.0208, 3013: 0.034, 3014: 0.0237	Moderate strength, high fatigue strength, low modulus of toughness, good weldability, excellent corrosion resistance
	Sub-cluster 32	3104: 0.9682	3004: 0.0133, 3204: 0.0091, 3304: 0.0094	Superior strength, medium fatigue strength, low modulus of toughness, good weldability, excellent corrosion resistance
	Sub-cluster 33	3016: 0.8979	3005: 0.0186, 3015: 0.0289, 3019: 0.0192, 3025: 0.0354	High strength, moderate fatigue strength, modulus of toughness, good weldability, excellent corrosion resistance

7.4.3 Metallurgical Reasoning Behind Clustering of 3xxx Series Alloys

The 3xxx series alloys are non-age-hardenable, meaning their strength cannot be significantly enhanced through heat treatment. The primary strengthening mechanisms in these alloys are

work hardening and solid-solution hardening. Mn is the principal alloying element, while Fe, Si, Mg, Cu, and Zn are present as impurities. The presence of these elements influences both the microstructure and the properties of the alloys.

Mn plays a key role in solid-solution strengthening, while Cu and Mg impurities in solution within the matrix also contribute to solid-solution hardening [261].

Two major second-phase particles are typically observed in these alloys: $\text{Al}_6(\text{Mn,Fe})$ and $\alpha\text{-Al}_{12}(\text{Fe,Mn})_3\text{Si}$. $\text{Al}_6(\text{Mn,Fe})$ is commonly found in commercial alloys; however, in the presence of Si impurities, it can partially transform into $\alpha\text{-Al}_{12}(\text{Fe,Mn})_3\text{Si}$. The ratio of Mn to Fe, as well as the combined (Fe + Mn) to Si ratio, also affects the properties of these alloys [262].

Si impurities promote the precipitation of dispersoids such as $\alpha\text{-AlMnSi}$ and $\alpha\text{-Al}_{12}(\text{Fe,Mn})_3\text{Si}$, which form more readily in alloys with higher Si content. The most prevalent dispersoid, $\alpha\text{-Al}_{12}(\text{Fe,Mn})_3\text{Si}$, is partially coherent with the aluminium matrix. A higher volume fraction of these dispersoids, influenced by the levels of Si and Mn, enhances mechanical properties, including strength and fatigue strength. However, excessive impurity content may lead to the formation of coarse intermetallic phases, which can reduce ductility and the modulus of toughness [263]. The clustering results align well with the underlying metallurgical characteristics, as discussed in detail below.

Cluster 3 (Highest Strength Alloys, see Appendix Figure 11.9 and Table 11.3): This cluster exhibits the highest strength among the 3xxx series alloys, primarily due to its high Mn content, which contributes to strengthening via work hardening and dispersoid formation. Mg, Cu, Si, and Fe are also present as impurities, influencing both the microstructure and mechanical properties.

The elevated Mn content increases the volume fraction of dispersoids such as $\alpha\text{-AlMnSi}$ and $\text{Al}_6(\text{Mn,Fe})\text{Si}$, which act as effective barriers to dislocation motion, thereby enhancing alloy strength [264], [265]. Fe impurities reduce Mn solubility in the aluminium matrix and accelerate the precipitation of these dispersoids, further strengthening the alloys [263], [266], [267].

Although present only as an impurity, Cu influences dispersoid formation by promoting the development of smaller dispersoids with higher number density, which in turn hinders dislocation motion [259]. Mg promotes the precipitation of Mg_2Si during annealing, which enhances

nucleation of α -Al₆(Mn,Fe)Si dispersoids. Si and Cu do not directly contribute to strengthening, as these alloys are non-age-hardenable [268].

Corrosion resistance was high in these alloys due to the elevated Mn content, which reduced the potential difference between the aluminium matrix and intermetallic phases. This effect arose from Mn supersaturation within the matrix, along with Mn enrichment in the intermetallic phases, thereby lowering the overall corrosion susceptibility [268].

Weldability is moderate, as impurities such as Cu, Fe, Mg, and Si can lead to hot cracking. These elements tend to segregate at grain boundaries, forming low-melting-point eutectic phases that create continuous films along the grain boundaries, making them prone to cracking.

Cluster 1 (High Strength Alloys): This group of alloys attains high strength through the combined effects of dispersoids and solid-solution hardening. High Mn content, along with impurities such as Si, Zn, Cr, Fe, Mg, and Cu, significantly influences their microstructure and properties.

The relatively large Si content, although an impurity, promotes the formation of α -Al(Mn,Fe)Si dispersoids, resulting in a high dispersoid volume fraction and size, which strengthens these alloys [266], [269], [270], [271]. Moderate Cu and Mg impurities assist in dispersoid formation, with Mg facilitating the precipitation of Mg₂Si, which aids dispersoid nucleation. However, as these alloys are non-age-hardenable, Mg₂Si does not contribute significantly to strengthening [263], [266].

High Cr content leads to the formation of Al₆(Mn,Fe,Cr) dispersoids, which enhance solid-solution strengthening and act as dislocation-pinning sites, further improving mechanical properties [272]. Medium Ti levels contribute to grain refinement by constraining grain growth during solidification and promoting the formation of Al₃Ti particles, which serve as low-energy nucleation sites [273].

Corrosion resistance is generally good, primarily due to the high Mn/Fe ratio, which reduces the potential difference between the aluminium matrix and intermetallic phases, thereby lowering corrosion susceptibility [268]. Ti supports the formation of a compact Al₂O₃ passivation layer, enhancing corrosion resistance [273], while high Cr content further increases corrosion resistance by reducing the likelihood of metastable pit initiation [274], [272].

Weldability is moderate, as the presence of Cu, Mn, Fe, Mg, and Si can promote hot cracking. This behaviour is attributed to impurity segregation at grain boundaries and the formation of low-melting-point eutectic phases, which create brittle films along the grain boundaries and increase crack susceptibility during welding [275].

Cluster 0 (Medium Strength Alloys): The moderate strength of these alloys is primarily due to their intermediate Mn content, which supports work hardening and the formation of dispersoids.

The Mn content promotes the development of dispersoids such as α -Al(Mn,Fe)Si and Al₆(Mn,Fe,Cr); however, the overall dispersoid volume fraction remains relatively low due to the limited Mn levels [269], [270], [271]. Fe impurities slightly increase Mn solubility but have minimal impact on the precipitation of α -Al(Mn,Fe)Si dispersoids [272], [271]. Additionally, Mg₂Si precipitation is inhibited upon heating, which in turn limits the nucleation of α -Al(Mn,Fe)Si dispersoids [263], [267], [276].

Corrosion resistance is moderate, as the medium Mn content increases the potential difference between intermetallic phases and the aluminium matrix, thereby raising susceptibility to localised corrosion. Cu impurities further elevate the potential of intermetallic particles, exacerbating this effect [268]. Fe impurities contribute to the formation of Al₃Fe and Al₆Fe, which act as cathodic sites and accelerate corrosion [263], [271].

Weldability is excellent owing to the moderate levels of Si, Fe, Mn, Cu, and Mg. These conditions result in a low content of residual eutectic phases, which are discontinuously distributed along grain boundaries as small, dispersed particles, thereby minimising the likelihood of hot cracking during welding [275].

Cluster 2 (Low Strength Alloys): The low content of solute atoms in these alloys results in minimal dispersoid formation, leading to poor mechanical properties. The relatively high Ti content, however, promotes the formation of a dense Al₂O₃ passivation layer, which effectively inhibits oxidation of the aluminium matrix and enhances corrosion resistance [273]. Despite this improvement in corrosion resistance, the overall mechanical performance of these alloys remains limited.

7.4.4 Thermo-Calc Analysis as a Basis for Cluster Classification

Thermo-Calc single-point equilibrium analysis was conducted to support the literature-based metallurgical interpretation of the 3xxx series alloy clusters under H32 temper conditions, and atmospheric pressure. The analysis focused on evaluating the phase volume fractions to elucidate their influence on the mechanical properties across the different clusters.

Since the analysis was conducted across all clusters, the phase formation and yield strength values were averaged to provide a clearer overview of the trends. As illustrated in Figure 7.3, Cluster 2 exhibits relatively low yield strength, which can be attributed to the limited formation of Mn-containing phases such as Al_6Mn , Al_{12}Mn , and $\text{Al}_{15}\text{Si}_2\text{Mn}_4$. By contrast, Cluster 3 shows a higher yield strength, reflecting the substantial formation of the Al_6Mn phase, which contributes significantly to its superior mechanical performance.

Overall, the observations from the Thermo-Calc analysis aligned with the metallurgical reasoning derived from the literature and provided additional support for the distinctions observed between clusters, confirming that the phase distributions correlated with the mechanical behaviour of the 3xxx series alloys.

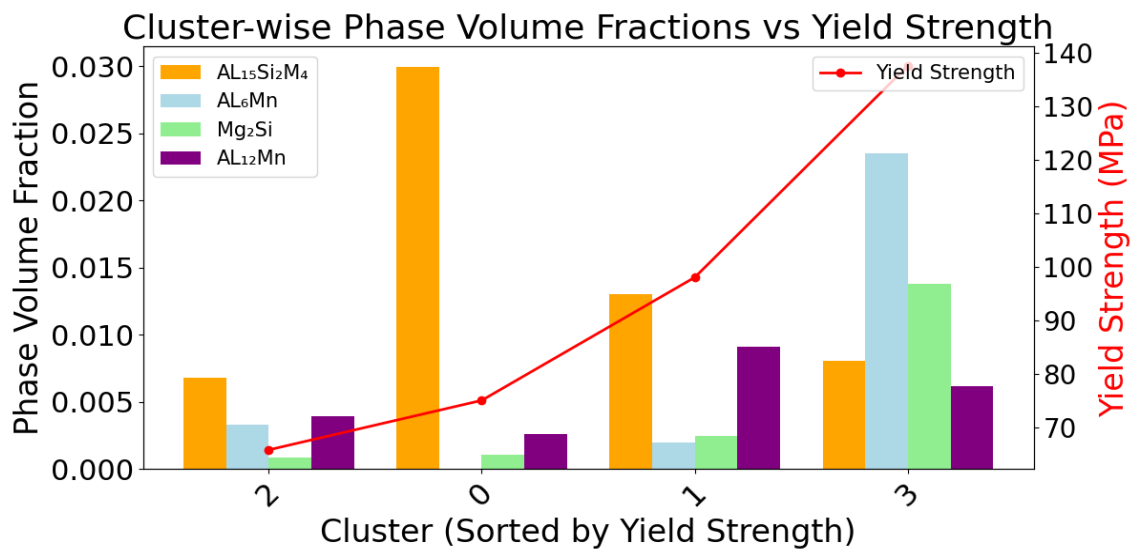


Figure 7.3. Thermo-Calc analysis of clustering of 3xxx series

7.5 Comparative Study – Internal Industrial Scrap Alloys vs Optimised Alloy Recycling

In this section, a comparative analysis is presented between the recycling of Set 1: Mixed internal industrial scrap from the 3xxx, 5xxx, and 6xxx aluminium alloy series and the recycling of Set 2: Optimum alloys for the same series selected by our design framework, as defined in **Chapter 3: Methodology of Study**. The same design framework is applied to both sets to enable a direct comparison of their recycling performance.

7.5.1 Data Collection

The data collection was carried out on two distinct sets. Set 1 (6 alloys of 3xxx, 13 alloys of 5xxx, 14 alloys 6xxx) comprised mixed internal industrial scrap, consisting of 33 alloys commonly found in industrial recycling streams. These alloys, drawn from the 3xxx, 5xxx, and 6xxx series, are widely utilised in cans, utensils, construction, building materials, and automotive applications [1], [138], [140], [277] [278] [279], [280], [281][282], [283] [284] [285] [286], [287], [278], [288], [289], [290], [291]. Their compositional and property ranges are summarised in Table 7.8. Set 2 (9 alloys of 3xxx, 12 alloys of 5xxx, 11 alloys 6xxx) consisted of 32 optimised alloys, identified through our Design Framework outlined in **Sections 3.2**, and designed to encompass a wider range of applications. The optimised alloys are presented in **Table 5.3** for the 6xxx series and in **Tables 7.4** and **7.6** for the 5xxx and 3xxx series, respectively, with their overall compositional and property ranges summarised in **Table 7.9**.

Table 7.8. Composition and property ranges for Set-1 alloys

Composition/ Property	3xxx Series	5xxx Series	6xxx Series
Si (%)	0.3–0.6	0.08–0.4	0.35–1.35
Fe (%)	≤0.4	≤0.35	0.075–0.35
Cu (%)	0.1–0.3	≤0.5	0.1–0.95
Mn (%)	0.55–1.25	0.03–1.1	0.03–0.8
Mg (%)	0–1.3	0.45–5.0	0.35–1.4
Yield Strength (MPa)	90–170	100–294	150–350
Fatigue Strength (MPa)	55–120	82–169	62–140
Modulus of Toughness (MPa %)	3.45–12.65	10.8–30.55	15.4–49.5
Weldability	Good–Excellent	Good–Excellent	Good–Excellent
Corrosion Resistance	Good–Excellent	Good–Excellent	Good–Excellent

Table 7.9. Composition and property ranges for Set-2 alloys

Composition/ Property	3xxx Series	5xxx Series	6xxx Series
Si (%)	0.15–1.8	0.08–0.4	0.35–1.35
Fe (%)	0.1–0.4	0.05–0.35	0.075–0.35

Cu (%)	0.05–0.5	0.05–0.2	0.1–0.85
Mn (%)	0.225–1.5	0.03–1.1	0.05–0.7
Mg (%)	0–1.3	0.4–5.05	0.5–1.0
Yield Strength (MPa)	34–170	160–310	170–430
Fatigue Strength (MPa)	34–120	90–180	68.9–180
Modulus of Toughness (MPa %)	9.8–19.6	12.9–28.1	13.1–49.5
Weldability	Good–Excellent	Good–Excellent	Good–Excellent
Corrosion Resistance	Good–Excellent	Good–Excellent	Good–Excellent

7.5.2 Alloys Sorting for Both Sets 1 and 2

7.5.2.1 Alloy Sorting of Set 1: Mixed Internal Industrial Scrap Alloys

The recycling analysis of Set-1 alloys commenced with the sorting of mixed scrap alloys using the combined PCA and K-means clustering approach as described in **Section 3.4**. Cluster validation, carried out using the elbow method ($K = 5$) and a silhouette score of 0.34 (see **Section 2.2.3.7.1** for methodological details on the elbow plot and silhouette scores), confirmed the presence of five well-defined clusters as shown in Figure 7.4, each exhibiting similar ranges of compositions and properties, with the figure also highlighting the specific applications associated with each cluster. This structured organisation of the mixed scrap enabled a more systematic and reliable recycling analysis.

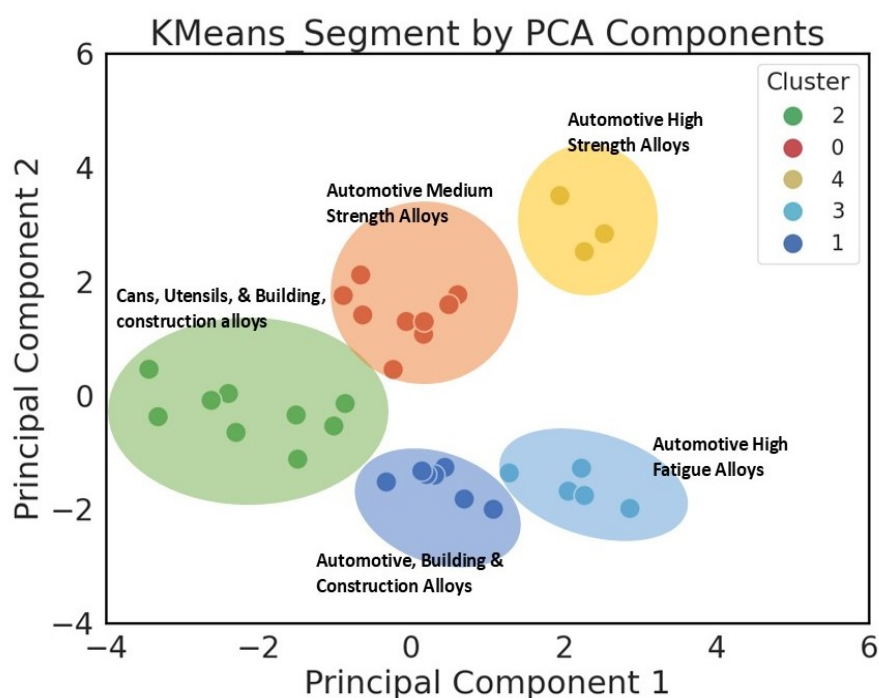


Figure 7.4. Clustering of mixed scrap aluminium alloys

As presented in Table 7.10, the alloys within each cluster exhibited closely aligned compositions and properties, confirming that the combined PCA and K-means clustering approach effectively sorted the mixed scrap alloys into meaningful clusters.

Table 7.10. Composition and property ranges within each cluster in Figure 7.4

Property	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Si, wt%	0.45 – 1.25	0.08 – 0.4	0.3 – 0.6	0.25 – 0.4	0.8 – 0.85
Fe, wt%	0.1 – 0.35	0.04 – 0.2	0.35 – 0.4	0.2	0.2 – 0.25
Cu, wt%	0.1 – 0.4	0 – 0.15	0.2 – 0.3	0.1 – 0.5	0.85 – 0.9
Mn, wt%	0.1 – 0.7	0.03 – 1.1	0.55 – 1.25	0.1 – 1	0.45 – 0.8
Mg, wt%	0.425 – 1	0.45 – 3.2	0 – 1.3	3.5 – 5	0.75 – 1
Yield strength, MPa	150 – 276	186 – 219	90 – 170	200 – 294	310 – 350
Modulus of Toughness (MoT) MPa%	23.21 – 42	17.54 – 21.75	3.45 – 21.45	23.13 – 30.55	35.04 – 49.5
Fatigue strength, MPa	62 – 110	117 – 128	55 – 120	130 – 169	97 – 140
Corrosion	Good – Excellent	Good – Excellent	Good – Excellent	Excellent	Good – Excellent
Weldability	Good – Excellent	Good – Excellent	Good – Excellent	Excellent	Good – Excellent

For example, Cluster 2 primarily comprises 3xxx series alloys, with some 5xxx and 6xxx series, characterised by low strength, moderate modulus of toughness, medium fatigue strength, excellent corrosion resistance, and good weldability. They are suitable for cans and utensils, with some use in building, construction, and automotive sectors [278], [288], [289], [289], [292].

Cluster 4 includes exclusively 6xxx series automotive alloys with high strength, medium fatigue strength, high modulus of toughness, excellent weldability, and moderate corrosion resistance [293].

Other clusters were similarly organised by properties and applications: Cluster 0 covers automotive and some building/construction alloys [286], [292], [294]; Cluster 3 includes medium yield strength and high fatigue strength automotive alloys [280]; and Cluster 1 comprises automotive and building/construction alloys with excellent corrosion resistance and medium weldability [15], [281], [294].

This analysis demonstrated that clustering effectively sorted the Set-1 mixed industrial scrap alloys.

7.5.2.2 Alloy Sorting of Set-2: Optimum Alloys

A similar alloy sorting approach of PCA and K-means clustering (See Section 3.4) was applied to Set-2, comprising optimum alloys of 3xxx, 5xxx and 6xxx series selected to cover a broader range of applications, including those represented in Set-1 mixed-scrap alloys. The analysis resulted in three clusters, in agreement with the elbow method ($K = 3$) and a silhouette score of 0.24 (see Section 2.2.3.7.1 for details). These clusters represent alloys with similar property and composition ranges, as illustrated in Figure 7.5. Table 7.11 summarises the composition and property ranges for each cluster.

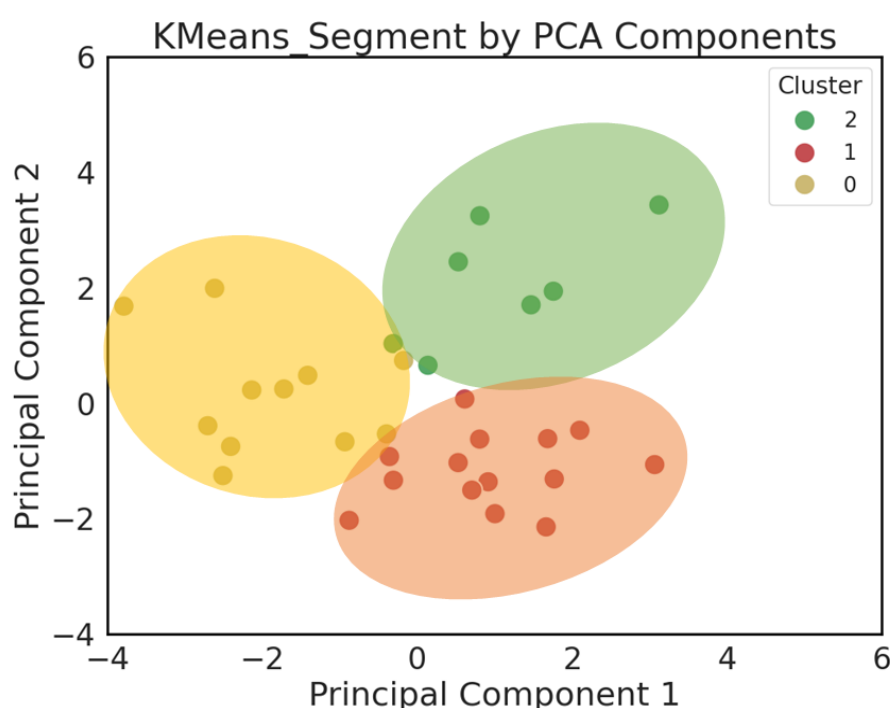


Figure 7.5. Clustering of optimum alloys of 3xxx, 5xxx, and 6xxx series

Table 7.11. Composition and property ranges of alloys within each cluster in Figure 7.5

Property	Group 0	Group 1	Group 2
Si, wt%	0.34 – 1.8	0.08 – 0.4	0.6 – 1.35
Fe, wt%	0.25 – 0.4	0.05 – 0.35	0.15 – 0.35
Cu, wt%	0.1 – 0.5	0.05 – 0.25	0.1 – 0.85
Mn, wt%	0 – 1.5	0.03 – 1.1	0.15 – 0.7
Mg, wt%	0 – 1.45	0.4 – 5.05	0.6 – 1
Yield strength, MPa	34 – 270	170 – 310	220 – 430
Modulus of Toughness (MOT), MPa%	9.84 – 29.4	10.12 – 28.05	13.14 – 49.5
Fatigue strength, MPa	34 – 120	59 – 180	88 – 180
Corrosion	Excellent	Good – Excellent	Fair – Good / Good – Excellent
Weldability	Good – Excellent	Fair – Excellent	Good – Excellent

For example, Cluster 1 comprised medium strength 5xxx and 6xxx series alloys, with some 3xxx series, exhibiting high fatigue strength, excellent corrosion resistance and weldability, used in transportation, marine, construction, packaging, consumer goods, and electrical applications. Cluster 2 mainly included high strength 6xxx series alloys with excellent weldability for automotive, construction, electronics, aerospace, pipelines, and industrial wear applications [156], [164], [1], [243], [152], [153], [154], [159]. Cluster 0 consisted mostly of 3xxx series, with some 5xxx and 6xxx series alloys, showing excellent corrosion resistance, good weldability, low strength, and moderate modulus of toughness, applied in transportation, construction, packaging, HVAC, marine machinery, and consumer goods [160], [258], [260], [1].

7.5.3 Recycling Comparison between Set-1 and Set-2 Alloys

After performing the alloy sorting of Set-1 and Set-2 alloys into well-defined clusters, the optimisation and recycling algorithms described in **Sections 3.6** and **3.7** were applied within each cluster to generate the top-performing recycled alloys. These alloys, which possess the optimum combination of properties and represent the entire cluster, are summarised in Table 7.12.

A comparative analysis was then conducted between the two sets to evaluate the effectiveness of the alloy sorting, top-performing alloy selection, and recycling algorithms. The Set-2 alloys were carefully pre-selected for high performance and broad application coverage. As anticipated, their recycling yielded alloys with superior properties and versatility, demonstrating that pre-selection prior to usage enhances the performance and application potential of recycled alloys.

In contrast, Set-1 alloys comprise commonly available industrial scrap, characterised by moderate properties tailored to specific applications. Despite these limitations, the same methodology successfully optimised and recycled the Set-1 alloys within their clusters, producing five top-performing alloys with balanced properties, as presented in Table 7.12.

The comparison emphasises two key points. First, carefully pre-selected optimum alloys (Set-2) inherently generate higher-performing recycled alloys. Second, even standard industrial scrap (Set-1) can be systematically upgraded using clustering, optimisation, and recycling algorithms, yielding top performing alloys with broad potential applications. Overall, the methodology demonstrates flexibility and effectiveness across both commonly available and carefully selected alloy sets.

Table 7.12. Comparison of optimised recycled alloys from Set-1 and Set-2

Alloy Set	Optimum recycled Alloy	Representative Applications	Key Property Highlights
(Set-1 Recycled Industrial Scrap)	6060	Building and Construction [292]	Low strength, excellent corrosion resistance, good weldability, medium modulus of toughness, low fatigue strength
Set-1			
Set-1	6061	Automotive [294], Building and Construction [292], utensils [1]	Medium-high strength, good corrosion resistance, excellent weldability, high modulus of toughness, medium fatigue strength
Set-1	6010	Automotive [293], [295]	High strength, excellent corrosion resistance, good weldability, high modulus of toughness, high fatigue strength
Set-1	5056	Automotive [296]	High strength, excellent corrosion resistance, excellent weldability, medium modulus of toughness, very high fatigue strength
Set-1	5110	Automotive [296]	Moderate strength, good corrosion resistance, excellent weldability, medium modulus of toughness, high fatigue strength
(Set-2 Recycled Optimum Alloys)	6040	Automotive, small engine parts, sports products, as well as potentially other machined parts requiring good performance [153], [154]	High strength, excellent corrosion resistance, excellent weldability, high modulus of toughness, high fatigue strength
Set-2			
Set-2	5954	Road transport bodywork, such as trailers and caravans, chemical and process plant equipment, and components for marine and offshore structures [1], [243]	High strength, excellent corrosion resistance, good weldability, medium modulus of toughness, very high fatigue strength
Set-2	6082	High stress applications such as bridges, trusses, and cranes [160]	Medium-high strength, excellent corrosion resistance, good weldability, medium modulus of toughness, medium fatigue strength

7.6 Conclusion on Optimisation of 2xxx, 3xxx, and 5xxx Series Alloys and Mixed Scrap Recycling Comparison

This study presents a comprehensive framework for the optimisation and recycling of aluminium alloys, spanning both single-series (2xxx, 3xxx, 5xxx) and mixed-series (3xxx, 5xxx, 6xxx) optimisation.

Single-Series Optimisation: The 2xxx, 3xxx, and 5xxx series were optimised using PCA, K-means clustering, an optimisation algorithm, and a recycling algorithm. Clustering identified distinct groups of alloys based on compositional, mechanical, corrosion-related, weldability,

and high-temperature properties: six clusters for the 2xxx series, four clusters for the 3xxx series, and three clusters for the 5xxx series alloys. Subsequent sub-clustering provided greater granularity, yielding 15 sub-clusters for the 2xxx series, 9 sub-clusters for the 3xxx series, and 11 sub-clusters for the 5xxx series. The optimisation algorithm reduced the solution space while retaining high-performing alloys representative of each sub-cluster, decreasing the number of candidates from 59 to 17 alloys in the 2xxx series, 36 to 9 alloys in the 3xxx series, and 71 to 12 alloys in the 5xxx series.

The recycling algorithm enabled alloy recycling within sub-clusters, producing alloys with minimal deviations from the optimised alloy properties and compositions: variations in yield strength, creep strength, and rupture properties were below 1 MPa, and composition errors were under 0.001%. Some alloys were incompatible with mixing due to unique property sets.

A metallurgical analysis was conducted for each cluster of the 2xxx, 3xxx, and 5xxx series alloys. The analysis indicated that performance differences were primarily influenced by key elements: Cu, Ni, and Ag in the 2xxx series (forming Al_2Cu and S phases); Mg and Mn in the 3xxx series (forming Al_6Mn , Al_{12}Mn , and $\text{Al}_{15}\text{Si}_2\text{Mn}_4$ intermetallic phases); and Mg in the 5xxx series (forming the T phase, Al_2Mg_3). Thermo-Calc equilibrium analysis further validated these findings.

Mixed-Series Optimisation: The framework was extended to the 3xxx, 5xxx, and 6xxx series alloys to compare the recycling of mixed internal industrial scrap (Set-1: 33 alloys) with pre-selected optimum alloys (Set-2: 32 alloys). The sorting of each set resulted in well-organised clusters: 5 clusters for set 1 and 3 clusters for set 2 alloys. Clustering of Set-1 revealed that Cluster 2 comprised alloys commonly used in cans, utensils, building, and automotive applications, while Clusters 3 and 4 consisted predominantly of automotive alloys. In contrast, clustering of Set-2: optimum alloys, produced three clusters with broader sectoral applicability. The optimisation and recycling algorithms were then applied within each cluster, yielding recycled top-performing alloys: five alloys from Set-1 and three alloys from Set-2.

The pre-selection of optimum alloys in set 2 ensured that the resultant recycled top-performing alloys had wider applicability, whereas the set 1: mixed industrial scrap recycled alloys were more application-specific and exhibited moderate properties. This demonstrates that pre-selection of optimum alloys prior to usage results in the production of high-performing recycled alloys with wider applicability.

Overall, this integrated framework successfully reduced the aluminium alloy solution space for various alloy series and mixed series; enabled systematic scrap sorting and recycling; and provided a robust, industrially relevant strategy for developing high-performance alloys with broad or specialised application potential.

CHAPTER 8: PRACTICAL APPLICATION: ALUMINIUM ALLOY RECYCLING SOFTWARE

8.1 Overview of Recycling Software Application

After successfully optimising the individual 2xxx, 3xxx, 5xxx, 6xxx, and 7xxx series, as well as performing mixed-series comparative optimisation of the 3xxx, 5xxx, and 6xxx series aluminium alloys using our Design Framework as discussed in **Section 3.2**, the next logical step was the development of a practical application for aluminium recycling. This application was developed in two stages. In the first stage, an individual-series recycling application was created, initially focused on the 6xxx series, with the flexibility to accommodate other series as needed. In the second stage, a comprehensive mixed-series recycling application was developed, capable of processing combinations of alloys across the 1xxx–8xxx series, with enhanced functionality and broader applicability. Both applications build upon the optimisation and recycling algorithms detailed in **Sections 3.6** and **3.7**. The objective of this stage is to provide researchers and industry practitioners with a cost-effective, time-efficient, and resource-efficient methodology for recycling aluminium alloys while achieving the desired compositions and material properties.

To address this need, the Alloy Recycling Software Application has been developed to predict the optimal properties and compositions of recycled aluminium alloys, as well as the appropriate mixing ratios for recycling. The application offers significant advantages by enabling precise recycling processes that ensure the resulting alloys achieve optimal mechanical, corrosion, and technological performance.

Recycling aluminium alloys necessitates stringent control of chemical composition to guarantee that recycled materials maintain, or indeed enhance, their performance for subsequent applications. This facilitates the production of high-performance materials. Moreover, the application incorporates an intuitive user interface, rendering it a valuable tool for materials scientists and engineers engaged in sustainable materials engineering.

Within the broader context of sustainability, aluminium recycling is pivotal in environmental protection and the reduction of carbon footprint. By integrating computational methods into the recycling process, this software application supports the development of efficient,

optimised, and sustainable alloy recycling practices. This chapter provides a comprehensive account of the design, operation, and impact of the Alloy Recycling Software Application.

8.2 Design of Recycling Software Application

The Alloy Recycling Software Application was developed using Python and deployed in the Google Colaboratory environment. To improve accessibility and user experience, the application was integrated with the Streamlit framework, enabling interactive visualisation through an intuitive and user-friendly interface [297], [298].

The software application was designed with a range of user interface elements to facilitate data input and selection processes, as illustrated in Figure 8.1. Users can input alloy compositions, specify mechanical properties depending on the software stage, and select desired tempering conditions. Currently, the software application supports only selected artificial ageing conditions for 6xxx series alloys.

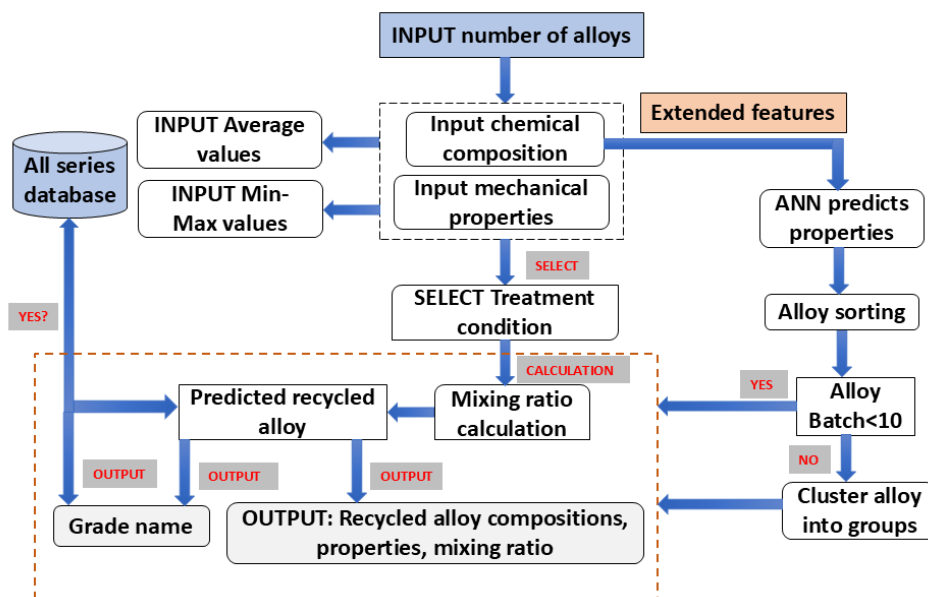


Figure 8.1. Design of recycling software application

8.3 Interface of the 6xxx Series-Based Alloy Recycling Software Application

8.3.1 User Input Parameters

User-defined inputs initiate the alloy recycling process by specifying the parameters of the recyclable alloys within the Software Application:

Step 1: Specify the Number of Alloys – Users indicate the number of 6xxx series alloy grades to be recycled.

Step 2: Provide Chemical Compositions and Mechanical Properties – For each alloy, users enter the chemical compositions (Fe, Si, Mg, Cu, and Mn) and mechanical properties, including yield strength (MPa), modulus of toughness (MPa %), and fatigue strength (MPa).

Input options for chemical compositions:

Typical Average Values: Users provide a single average value for each alloying element in the alloy.

Minimum–Maximum Range: Users specify the minimum and maximum values for each alloying element, allowing the software application to perform calculations across the defined range.

Input options for mechanical properties:

Typical Average Values: Users enter a single average value for each property.

Minimum–Maximum Range: Users input individual minimum and maximum values for each property, which the software uses to perform property calculations.

Step 3: Select Tempering Condition – Users select the tempering condition appropriate for the intended application. The current options include T5, T6, and T7 (as defined in **Section 2.1.5**).

8.3.2 Calculation of Alloy Mixing Ratios, Composition, and Properties

The Alloy Recycling Software Application calculates alloy mixing ratios while simultaneously predicting the resulting composition and mechanical properties of the recycled alloys in the following manner:

Step 1: Normalisation – User-input alloy properties were normalised to a common scale using min–max normalisation, as defined in Equation (14).

Step 2: TOPSIS Score Calculation – The optimisation algorithm was applied to determine the closeness scores of the user-input alloys to ideal solutions, as outlined in **Section 3.6** and summarised below:

- The ideal best and ideal worst solutions were determined using Equations (15) and (16).
- The distances of each alloy from the ideal best and ideal worst solutions were computed.
- TOPSIS scores for each alloy were subsequently calculated using Equation (17).

Step 3: Mixing Ratio Calculation – The alloy exhibiting the optimal range of properties (i.e., the maximum TOPSIS score) was assigned a minimum mixing ratio of 50%. The remaining 50% was distributed proportionally among the other alloys according to their respective TOPSIS closeness scores, following the procedure described in **Section 3.7**.

Step 4: Final Recycled Alloy Composition and Properties Calculation – Weighted averages of compositions and mechanical properties were computed using the determined mixing ratios to obtain the final recycled alloy's composition and properties using Equation (27).

8.3.3 Similar Alloy Grade Search

Once the final recycled alloy composition and properties were determined, the software application searched for similar alloys within a pre-established database of 6xxx series alloys, which includes standard grades along with their compositions, properties, and tempering conditions (T5, T6, and T7) as follows:

Step 1: Filter by Tempering Condition – The software application filtered database records to include only those alloys that share the same tempering condition as the one selected by the USER in **Step 3** in **Section 8.3.1**.

Step 2: Similarity Analysis – The similarity of each remaining alloy to the recycled alloy was assessed as follows:

Each user-defined alloy was represented as a composition vector, $V_i = [\text{Cu}, \text{Mg}, \text{Mn}, \text{Si}, \text{Fe}]$, while the recycled alloy was represented similarly as $W_i = [\text{Cu}, \text{Mg}, \text{Mn}, \text{Si}, \text{Fe}]$

For each alloy in the filtered dataset, a numerical distance to the recycled alloy was calculated using the Euclidean distance [299]:

$$d_s(\mathbf{V}_l, \mathbf{W}_l) = \sqrt{(\text{Cu}_{V_l} - \text{Cu}_{W_l})^2 + (\text{Mg}_{V_l} - \text{Mg}_{W_l})^2 + (\text{Mn}_{V_l} - \text{Mn}_{W_l})^2 + (\text{Si}_{V_l} - \text{Si}_{W_l})^2 + (\text{Fe}_{V_l} - \text{Fe}_{W_l})^2}$$

(28)

This distance represents how similar the elemental composition is: Smaller d indicates greater similarity and larger d_s indicates greater difference.

Step 3: Output – The alloy with the minimum distance (d_s) was considered the closest match and was displayed by the application.

This approach enhances the efficiency and effectiveness of the recycling process, enabling researchers and engineers to produce optimised, high-performance recycled aluminium alloys.

8.4 Demonstration of the Alloy Recycling Software Application for 6xxx Series

8.4.1 User Input Compositions, Properties, and Processing Conditions

The Alloy Recycling Software Application begins with the user entering the chemical compositions and mechanical properties of the 6xxx series aluminium alloys selected for recycling. In this demonstration, the user chose three alloys, designated A1, A2, and A3, and opted to provide average values for both composition and mechanical properties. The input data are summarised in Table 8.1, with normalised property values indicated in parentheses. Figures 8.2(a)–(d) present the software interface employed for data entry.

Table 8.1. USER defined 6xxx series alloys

Alloy	Si (%)	Cu (%)	Mg (%)	Mn (%)	Fe (%)	Yield Strength (MPa)	Toughness (MPa %)	Fatigue Strength (MPa)
A1	0.45	0.25	0.6	0.125	0.175	220 (1)	23.205 (0.3606)	88 (1)
A2	0.45	0.25	0.6	0.125	0.175	200 (0)	20.93 (0)	79 (0.5263)
A3	0.4	0.2	0.675	0.05	0.075	214 (0.7)	27.3 (1)	68.9 (0)

Alloy Recycling Software

Welcome to the Alloy Recycling Application. Recycle your alloys and create new compositions!

Recycle 6XXX Series Alloys

Enter the number of 6XXX series alloys for recycling:

3

Alloy 6XXX 1

☒ Use average typical values for Alloy 6XXX 1

Average Typical Yield Strength (MPa) of Alloy 6XXX 1:

220.00

Average Typical Toughness (MPaH) of Alloy 6XXX 1:

23.20

(a)

Alloy 6XXX 1

☒ Use average typical values for Alloy 6XXX 1

Average Typical Yield Strength (MPa) of Alloy 6XXX 1:

220.00

Average Typical Toughness (MPaH) of Alloy 6XXX 1:

23.20

Average Typical Fatigue Strength (MPa) of Alloy 6XXX 1:

88.00

☒ Use average composition for Si:

Avg Composition of Si (%)

0.45

☒ Use average composition for Cu:

Avg Composition of Cu (%)

(b)

Alloy 6XXX 2

☒ Use average typical values for Alloy 6XXX 2

Average Typical Yield Strength (MPa) of Alloy 6XXX 2:

200.00

Average Typical Toughness (MPaH) of Alloy 6XXX 2:

20.93

Average Typical Fatigue Strength (MPa) of Alloy 6XXX 2:

79.00

☒ Use average composition for Si:

Avg Composition of Si (%)

0.45

☒ Use average composition for Cu:

Avg Composition of Cu (%)

(c)

Alloy 6XXX 3

☒ Use average typical values for Alloy 6XXX 3

Average Typical Yield Strength (MPa) of Alloy 6XXX 3:

214.00

Average Typical Toughness (MPaH) of Alloy 6XXX 3:

27.30

Average Typical Fatigue Strength (MPa) of Alloy 6XXX 3:

68.90

☒ Use average composition for Si:

Avg Composition of Si (%)

0.40

☒ Use average composition for Cu:

Avg Composition of Cu (%)

(d)

Figure 8.2. Interface for input of (a) Number of alloys, (b) Chemical composition and Properties of alloy 1, (c) Chemical composition and Properties of alloy 2 (d) Chemical composition and Properties of alloy 3

Figure 8.3. Tempering condition selection: T6

Following the entry of compositions and properties, the user selected the T6 tempering condition, as illustrated in Figure 8.3. This condition corresponds to the standard heat treatment applied to 6xxx series aluminium alloys.

8.4.2 Optimum Alloy Calculation

Step 1: User Data Normalisation – To enable comparison of mechanical properties on a common scale, all properties were normalised using min–max normalisation. This procedure maps data from any range to a standard range between 0 and 1, allowing for direct comparison. The normalised alloy properties are indicated in brackets in Table 8.1.

Step 2: Identification of Ideal Best and Worst Solutions – The TOPSIS optimisation algorithm (see Section 3.6) uses two reference points:

- **Ideal Best Solution:** The maximum normalised values for all properties, i.e., (1.0, 1.0, 1.0).
- **Ideal Worst Solution:** The minimum normalised values for all properties, i.e., (0.0, 0.0, 0.0).

Step 3: TOPSIS Score Calculation – The closeness scores were then calculated using the distances from the ideal best (T^*) and ideal worst (T^-) solutions, according to Equation (17). The resulting closeness scores for the alloys were: A1 = 0.6941, A2 = 0.2619, and A3 = 0.5390.

8.4.3 Mixing Ratio Calculation

Following the calculation of the TOPSIS closeness scores, the alloy with the highest closeness score (A1) was allocated the largest portion of the mixing ratio, set at 50%. The remaining 50% was distributed between alloys A2 and A3 proportionally, based on their respective TOPSIS scores, as determined using Equation (25) and Equation (26) (see **Section 3.7** for details on recycling algorithm).

The resulting proportions in the mixture were A1 – 50%, A2 – 16.4%, and A3 – 33.7%. The composition of the final recycled alloy, together with its mechanical properties, was calculated as the weighted average of the selected alloys and their corresponding mixing ratios, in accordance with Equation (27).

The resulting recycled alloy had the following composition and properties: 0.43% Si, 0.23% Cu, 0.63% Mg, 0.10% Mn, and 0.14% Fe; with a yield strength of 215 MPa, Modulus of toughness of 24.2 MPa %, and fatigue strength of 80 MPa as shown in Figure 8.4 (a).

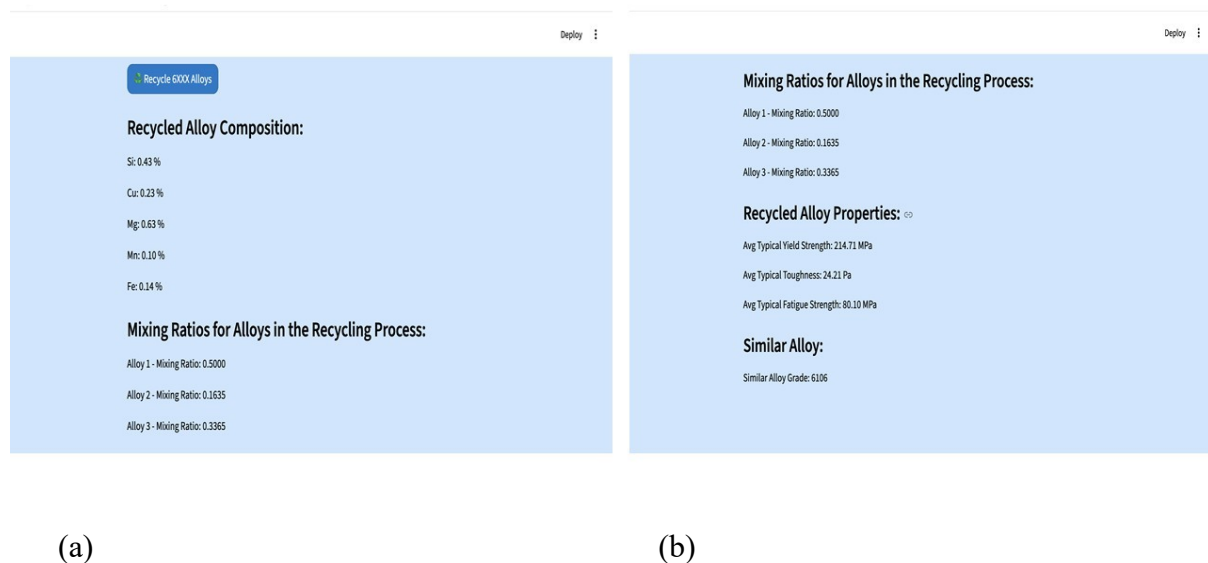


Figure 8.4. Recommended recycled alloy: (a) Composition & (b) Mixing ratios and properties and similar alloy grade

8.4.4 Finding the Closest Standard Alloy Grade

The composition, properties, and tempering condition (T6) of the final recycled alloy were subsequently compared with those of the standard 6xxx series alloys available in the database. The software iterated through the database of standard alloys and calculated the Euclidean

distance between the final recycled alloy's composition and properties and each standard alloy composition and properties using, Equation (28). The closest match was identified based on the smallest Euclidean distance.

In this demonstration, the most similar alloy was found to be Aluminium Alloy 6106-T6, as illustrated in Figure 8.4(b) within the software interface output.

This stepwise process demonstrated the Alloy Recycling Software Application's ability to determine optimal recycled-alloy compositions and predict their mechanical properties. By optimising the reuse of materials, the application improves recycling efficiency and reduces waste, energy consumption and carbon emissions. It also identifies the closest standard alloy grades, allowing verification that the recycled alloy achieves comparable performance. By enabling the use of scrap alloys to produce high-quality materials, the software reduces dependence on primary aluminium and supports circular-economy principles. Overall, the approach promotes sustainable aluminium recycling and facilitates greener industrial applications.

Although the software was demonstrated in detail for the 6xxx series, the same computational framework was subsequently extended to other aluminium alloy series, as discussed in the following section.

8.5 Extension of Recycling Software Application to Mixed-Series Recycling

Following the successful development of the 6xxx series Alloy Recycling Software Application, the application was extended to handle mixed-series aluminium recycling. This enables the combination of alloys from multiple series, including 1xxx, 2xxx, 3xxx, 5xxx, 6xxx, 7xxx, and 8xxx, addressing practical industrial scenarios where scrap materials from different alloy series are often combined.

8.5.1 Model Enhancements and Performance

The mixed-series software application extension introduces several key enhancements compared to the 6xxx series implementation, while the other features remain the same as 6xxx series software application:

Feature 1: Expanded Element Library: The user input composition in **Step 2** of **Section 8.3.1** now accommodates additional alloying elements beyond those in the 6xxx series, such

as Cr, Zn, Ti, Ag, Li, Ni, Zr, Ga, and V. This allows the software to process a wide variety of alloys simultaneously.

Feature 2: Advanced Property Prediction via ANN: To accommodate a broader range of elements and cross-series variations, an artificial neural network (ANN) was implemented to predict mechanical (yield strength, fatigue strength, modulus of toughness), technological and corrosion properties, as well as thermal and electrical properties of aluminium alloys (see **Section 2.2.3.8.2**). In industrial practice, the chemical composition of scrap alloys is often known, whereas their properties remain uncertain. This extended software application requires the user to input only the alloy composition. The ANN, trained on a dataset encompassing 1xxx–8xxx series alloys, learns the relationship between composition and material properties and predicts the properties of user-input compositions internally. These predicted values are then used for subsequent optimisation and alloy recycling steps.

The ANN architecture and training details are as follows:

Input Layer: 14 nodes corresponding to elemental features (Si, Fe, Cu, Mn, Mg, Cr, Zn, Ti, Ag, Li, Ni, Zr, Ga, V)

Hidden Layers: Four hidden layers with the following configuration:

- First hidden layer: 256 neurons, ReLU activation, followed by Batch Normalization and 15% Dropout
- Second hidden layer: 128 neurons, ReLU activation, followed by Batch Normalization and 15% Dropout
- Third hidden layer: 64 neurons, ReLU activation
- Fourth hidden layer: 32 neurons, ReLU activation

Dropout: 15% applied after the first two hidden layers to prevent overfitting

Output Layer: 7 nodes corresponding to predicted properties (Yield Strength, Fatigue Strength, Toughness, Weldability, Corrosion, Electrical, Thermal conductivity) with linear activation

Optimiser: Adam with a learning rate of 0.0005

Training Parameters: Up to 500 epochs with early stopping (patience 40) and batch size of 16

Performance Metrics:

- **Training Set:** MAE = 6.74, $R^2 = 0.841$
- **Test Set:** MAE = 9.81, $R^2 = 0.732$

The lower R^2 on the test set is primarily due to the relatively small size of the dataset (364 samples) and inherent variability in alloy properties across different series and compositions. Despite this, the ANN demonstrates reasonably strong predictive capability on the training data and provides sufficiently accurate property predictions for recycling optimisation, making it a reliable tool for alloy design and sustainable material recycling. As more data becomes available under different tempering, and hardening conditions, the ANN can be retrained on this expanded dataset to further improve prediction accuracy and generalisation across a wider range of alloys and corrosion conditions.

Feature 3: Enhanced Decision Matrix: The software has been extended to manage both small and large batches of alloys. For small batches (fewer than 6 alloys), the workflow proceeds directly to the full procedures outlined in **Section 8.3.2** and **Section 8.3.3**. For larger scrap batches (≥ 6 alloys), PCA and K-means clustering as detailed in **Section 3.4**, to sort the alloys into groups of similar compositions and properties. These well sorted groups were treated separately to predict the properties and composition of the recycled alloys as performed in **Sections 8.3.2** and **8.3.3**. The clustering and decision thresholds can be adjusted according to specific industrial requirements.

8.5.2 User Interface and Demonstration

The extended software application user interface integrates the new ANN-based property prediction and clustering functionalities within the same structured workflow as the 6xxx series version.

Figure 8.5 illustrates the interface developed for mixed-series alloy recycling. Users can input the alloy compositions (Figure 8.5(a), (b), (c)), after which the software automatically calculates the optimal mixing ratios for the recycled alloy (Figure 8.5(d)), its resulting compositions

(Figure 8.5(e)), and the predicted properties (Figure 8.5(f)), following the computational workflow outlined in **Sections 8.3.2, 8.3.3, and 8.5.1.**

Deploy

Alloy Recycling Extended Software

Professional Alloy Recycling Software - Input compositions, predict properties, and optimize batches.

Input Scrap Batch

Number of alloys in batch:

3

Alloy 1 Composition Range

Min Si (%)	Min Fe (%)	Min Cu (%)	Min Mn (%)	Min Mg (%)	Min Cr (%)	Min Zn (%)
0.57	0.10	0.15	0.15	0.35	0.00	0.00
Max Si (%)	Max Fe (%)	Max Cu (%)	Max Mn (%)	Max Mg (%)	Max Cr (%)	Max Zn (%)
0.57	0.10	0.15	0.15	0.35	0.00	0.00
Min Ti (%)	Min Ag (%)	Min Li (%)	Min Ni (%)	Min Zr (%)	Min Ga (%)	Min V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max Ti (%)	Max Ag (%)	Max Li (%)	Max Ni (%)	Max Zr (%)	Max Ga (%)	Max V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00

(a)

Deploy

Alloy 2 Composition Range

Min Si (%)	Min Fe (%)	Min Cu (%)	Min Mn (%)	Min Mg (%)	Min Cr (%)	Min Zn (%)
0.60	0.11	0.10	0.12	0.45	0.00	0.00
Max Si (%)	Max Fe (%)	Max Cu (%)	Max Mn (%)	Max Mg (%)	Max Cr (%)	Max Zn (%)
0.60	0.11	0.10	0.12	0.45	0.00	0.00
Min Ti (%)	Min Ag (%)	Min Li (%)	Min Ni (%)	Min Zr (%)	Min Ga (%)	Min V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max Ti (%)	Max Ag (%)	Max Li (%)	Max Ni (%)	Max Zr (%)	Max Ga (%)	Max V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00

Alloy 3 Composition Range

Min Si (%)	Min Fe (%)	Min Cu (%)	Min Mn (%)	Min Mg (%)	Min Cr (%)	Min Zn (%)
0.40	0.17	0.10	0.10	0.68	0.00	0.00
Max Si (%)	Max Fe (%)	Max Cu (%)	Max Mn (%)	Max Mg (%)	Max Cr (%)	Max Zn (%)
0.40	0.17	0.10	0.10	0.68	0.00	0.00
Min Ti (%)	Min Ag (%)	Min Li (%)	Min Ni (%)	Min Zr (%)	Min Ga (%)	Min V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max Ti (%)	Max Ag (%)	Max Li (%)	Max Ni (%)	Max Zr (%)	Max Ga (%)	Max V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00

(b)

Deploy

Alloy 3 Composition Range

Min Si (%)	Min Fe (%)	Min Cu (%)	Min Mn (%)	Min Mg (%)	Min Cr (%)	Min Zn (%)
0.40	0.17	0.10	0.10	0.68	0.00	0.00
Max Si (%)	Max Fe (%)	Max Cu (%)	Max Mn (%)	Max Mg (%)	Max Cr (%)	Max Zn (%)
0.40	0.17	0.10	0.10	0.68	0.00	0.00
Min Ti (%)	Min Ag (%)	Min Li (%)	Min Ni (%)	Min Zr (%)	Min Ga (%)	Min V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max Ti (%)	Max Ag (%)	Max Li (%)	Max Ni (%)	Max Zr (%)	Max Ga (%)	Max V (%)
0.00	0.00	0.00	0.00	0.00	0.00	0.00

Recycle Batch

Small batch (<=6 alloys): TOPSIS-based recycling.

(c)

Deploy

Recycle Batch

Small batch (<=6 alloys): TOPSIS-based recycling.

Mixing Ratios

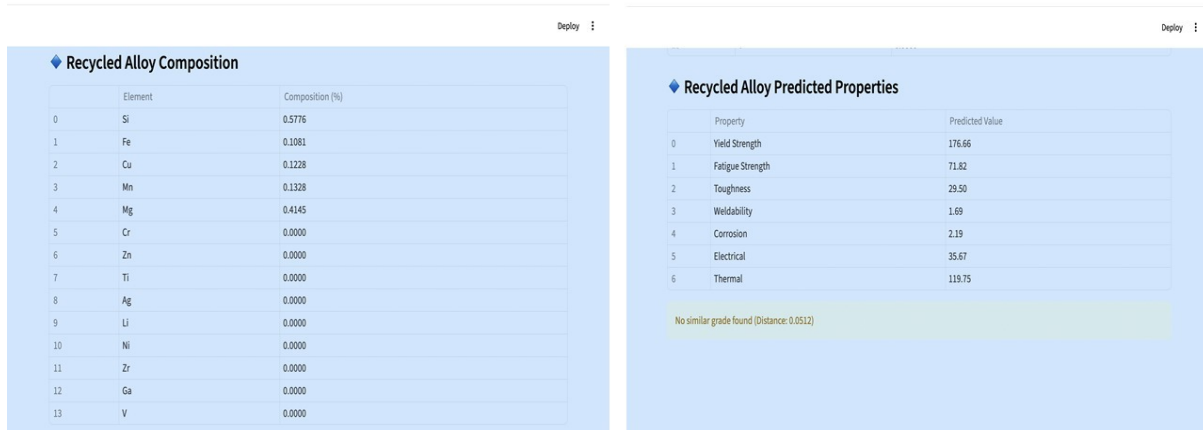
	Alloy	Mixing Ratio
0	Alloy 1	0.4562
1	Alloy 2	0.5000
2	Alloy 3	0.0438

Recycled Alloy Composition

	Element	Composition (%)
0	Si	0.5776
1	Fe	0.1081
2	Cu	0.1228

(d)

217



(e) (f)

Figure 8.5. Figure 8.57: Demonstration of the extended alloy recycling software application interface: (a), (b) & (c) Composition of the USER input alloys, (d) Mixing ratios of each USER input alloys, (e) Recycled alloy composition, and (f) Recycled alloy properties

The redesigned interface highlights the mixed-series functionality and demonstrates the software’s intuitive layout, stepwise logic, and practical relevance for industrial aluminium recycling applications.

8.6. Conclusion on Alloy Recycling Software Application

The Alloy Recycling Software Application provides a streamlined and efficient approach for sorting and recycling user-defined scrap alloys, enabling the prediction of high-quality secondary aluminium without depending on primary aluminium sources. It employs TOPSIS-based optimisation and recycling algorithms to determine the optimal mixing ratios, as previously implemented in both individual series optimisation and recycling as well as in selected mixed-series recycling.

The application consists of two main components. In the first part, users can input either average values or minimum–maximum ranges for the compositions and mechanical properties of 6xxx series alloys. Users may also specify tempering conditions (T5, T6, T7) to define the alloy treatment parameters. The software application then calculates the optimal mixing ratios using optimisation and recycling algorithms, assigning the largest proportion (atleast 50%) to the alloy with the highest TOPSIS closeness score, and distributing the remaining ratios among other alloys proportionally according to their respective TOPSIS closeness scores. This

approach enhances the overall performance and reliability of the final recycled alloy. The application outputs the recycled alloy's composition, predicted mechanical properties, and corresponding mixing ratios.

A notable feature of the software is its ability to recommend a standard alloy grade closely matching the recycled alloy, thereby ensuring compliance with industry standards.

The software application has been further extended to handle mixed-series aluminium recycling across the full compositional range of 1xxx–8xxx alloys, thereby expanding its applicability beyond the 6xxx series. An artificial neural network (ANN) has been integrated to predict the mechanical, technological, and physical properties of user-defined compositions. These predicted properties are subsequently utilised in the TOPSIS-based optimisation and recycling algorithm to determine the recycled alloy compositions and properties.

For large scrap batches, the software incorporates a combined Principal Component Analysis (PCA) and K-means clustering approach to perform sorting and group alloys with similar compositions before performing optimisation and recycling calculations. This ensures more efficient processing and accurate recycling across well sorted scrap groups.

These enhancements enable the software to efficiently manage mixed-series scrap while maintaining computational robustness and prediction accuracy. By adopting a data-driven methodology, the application provides a powerful tool for sustainable aluminium recycling, supporting circular economy initiatives and informed decision-making in industrial alloy recovery and reuse without relying on primary aluminium.

CHAPTER 9: SUMMARY, CONCLUSIONS AND FUTURE WORK

9.1 Conclusions

This thesis has systematically demonstrated how advanced data-driven techniques and practical tools can enhance recyclability and streamline aluminium alloy grade systems. The key findings from the comprehensive analyses are as follows:

9.1.1 It is Feasible to Reduce Standard Alloy Grades Across Several Alloy Series

The widespread use of aluminium alloy grades, often distinguished only by minute compositional variations, has resulted in many alloys sharing similar mechanical, corrosion, and weldability properties. Such redundancy diminishes the sustainability benefits of recycling by complicating alloy selection and significantly impeding recycling efficiency, particularly for wrought alloys, where strict tolerances on trace elements necessitate dilution with primary aluminium.

Our analysis demonstrates that, within each alloy series, up to approximately 2 wt.% of major alloying elements (Mn, Cu, Si, Zn, Mg) are typically present in aluminium alloys, with certain specialised grades containing higher levels of these elements. Additionally, minor alloying elements are often incorporated in trace amounts below 0.5 wt.%. The variability in mechanical properties within each series is similarly narrow. Despite this, numerous distinct alloy grades have historically been developed within these limited composition–property ranges. The proliferation of such near-identical alloys is not practically justified and introduces unnecessary complexity into the overall alloy system.

By employing machine learning–based optimisation in conjunction with recycling strategies, it is possible to consolidate alloys with similar characteristics into a small set of “optimum alloys”. These representative alloys are capable of encompassing the performance envelope of several existing grades, allowing alloy standards to be simplified without compromising functionality. This data-driven approach provides a clear pathway towards a more sustainable and streamlined aluminium alloy system, enabling the consolidation of alloy grades, reducing reliance on primary aluminium, and enhancing recycling efficiency.

9.1.2 Machine Learning Algorithms are Effective for Optimising Alloy Grade Systems

A major outcome of this work is the development of a machine learning-based design framework for optimising the aluminium alloy series, which reduces grade proliferation while maintaining material performance. For the initial analysis, a 6xxx series alloy dataset at the T6 condition was used as a case study and examined using multiple algorithms to develop a design framework suitable for both optimisation and recycling.

Several unsupervised learning algorithms were evaluated based on their silhouette scores and their ability to form distinct alloy clusters. K-means clustering was applied directly to the data (0.31), while PCA was combined with other algorithms, including Mean Shift (0.25), Gaussian Mixture Models (0.19), Hierarchical Clustering (0.22), and DBSCAN (-1). Additionally, a PCA combined with K-means clustering approach (0.20) was tested. Although the PCA and K-means method had a relatively lower silhouette score, it was identified as the most effective for achieving high-quality clustering, producing tighter separation of alloy groups with similar property ranges. This PCA and K-means clustering approach together with a defined mathematical model was also applied to the parent clusters to generate more compact sub-clusters with narrower property ranges.

Given the unlabelled nature of the dataset and the objective of uncovering intrinsic groupings, supervised classification algorithms were not considered. Within the resulting clusters and sub-clusters, optimisation algorithms, including multi-criteria decision making algorithms (TOPSIS, AHP, VIKOR), Bayesian optimisation, and artificial neural networks (ANN), were applied to assess their capability in selecting optimal alloys that cover the broad property ranges and potential applications of alloys within the cluster and sub-cluster. While ANN and Bayesian optimisation were less suitable, ANN demonstrated some effectiveness on the small dataset but had limitations in terms of overfitting. It can be useful when using large datasets in future. Bayesian optimisation is more appropriate for the design of novel alloys, which was beyond the scope of this study. Among all the algorithms, TOPSIS was the most effective in identifying optimal alloys within the sub-clusters across multiple performance criteria, including mechanical, and corrosion properties.

Leveraging this optimisation process, a recycling algorithm was designed using the TOPSIS closeness scores to perform alloy recycling within sub-clusters. The algorithm performs recycling by generating alloy mixing ratios, assigning at least 50% to the optimum alloys identified

by the TOPSIS optimisation algorithm, while distributing the remaining 50% among other alloys within the sub-cluster according to their TOPSIS closeness scores. This approach achieves optimal compositions and properties with minimal deviation and successfully reproduces existing standard optimum alloys identified by the TOPSIS optimisation algorithm, thereby confirming the practical relevance and accuracy of the methodology.

Following these analyses, a design framework for the optimisation of all aluminium alloy series was established. The framework comprises:

Data Collection → Missing Value Imputation (using Random Forest Regression and Classification) → PCA & K-Means Clustering → Sub-Clustering → Optimisation Algorithm → Alloy Mixing (Recycling Algorithm) → Validation Steps (Results Validation using an unseen dataset) → Metallurgical Interpretation (LIME Explainable AI (XAI) → Literature-based metallurgical reasoning behind ML findings → Thermo-Calc Validation).

This systematic, interpretable, and scientifically grounded approach enables alloy optimisation, recycling, and the simplification of the alloy grading system. Random forest regression and classification models were employed in the initial data processing step to impute missing property values, ensuring a complete and reliable dataset for subsequent analysis.

To elucidate the reasoning behind the ‘black box’ clustering of aluminium alloys, a combination of LIME XAI, metallurgical analysis based on literature, and Thermo-Calc single-point equilibrium calculations was employed. This approach provides insights into compositional and underlying property variations, as well as phase formation within clusters that influence alloy properties. This validation confirms the robustness of the framework. Furthermore, the results of the optimisation and recycling algorithms, used for selecting optimal alloys, were evaluated using unseen datasets that were not part of the data collection process.

Overall, this research demonstrates the transformative potential of machine learning in simplifying aluminium alloy classification systems, facilitating the selection of high-performance, sustainable, and resource-efficient optimum alloys.

9.1.3 Optimisation Framework Yields Robust and Representative Alloy Selections and Recycling

Prior to undertaking series-specific optimisation, a hybridisation analysis was carried out to investigate the potential for combining different alloy series and to examine cross-series groupings arising from compositional and property similarities. This step also served to validate the PCA–K-means clustering approach employed in this study.

For age-hardenable alloys (2xxx, 6xxx, and 7xxx), a dataset of 211 alloys was compiled. PCA and K-means clustering revealed natural groupings of alloys according to similarities in composition and properties, ultimately identifying five distinct clusters. Notably, Cluster 0 comprised corrosion-resistant 2xxx and 8xxx alloys containing Ag and Li, which are particularly suited to applications in corrosion-prone environments. Cluster 1 brought together 2xxx and some 7xxx alloys, which are limited by poor weldability and reduced corrosion resistance. The remaining three clusters reflected different balances between strength, corrosion resistance, and other properties, thereby capturing the inherent diversity of performance among these alloys.

For non-age-hardenable alloys (1xxx, 3xxx, 5xxx, and selected 8xxx), 152 alloys were analysed. Three clusters emerged: Cluster 1 included low-Mg alloys with high thermal and electrical conductivity; Cluster 2 consisted of medium-Mg alloys exhibiting moderate strength; and Cluster 0 comprised high-Mg or otherwise strengthened alloys, which balanced enhanced mechanical performance with corrosion resistance.

All together, these findings revealed overlapping property domains across different series. This observation is of particular importance, as it indicates the possibility of cross-series alloy mixing during recycling, thereby enhancing material circularity. Furthermore, these insights provide a strong foundation for the subsequent series-specific optimisation conducted in this study, confirming that the clustering and optimisation framework is both robust and scientifically meaningful.

In the **6xxx series**, 41 alloys encompassing 279 datasets across various T5, T6, and T7 tempering and processing conditions, including extrusions, forgings, rolling, sheets, and plates, were analysed based on their chemical compositions, mechanical properties, weldability, corrosion resistance, and anodising response using PCA and K-means clustering. This analysis consistently identified five parent clusters across the different tempering and processing conditions. Using the same PCA and K-means clustering approach combined with the defined mathematical model, these parent clusters were further subdivided into 11 sub-clusters, each representing a narrower range of properties and compositions.

The TOPSIS optimisation algorithm was then applied within each sub-cluster to identify the optimum alloy that best matched the ideal best solution, which was defined as the most favourable combination of properties. All alloys within each sub-cluster were assigned closeness scores based on their proximity to this ideal best solution, with the alloy achieving the highest closeness score designated as the optimum alloy. Through this process, the initial 41 standard alloys were reduced to 11 optimum standard alloys, each representing the performance characteristics and wider range of applications of its respective sub-cluster.

A recycling algorithm was subsequently employed for alloy mixing, guided by the TOPSIS closeness scores to generate the mixing ratios. This recycling successfully reproduced the optimum alloys identified by the optimisation algorithm with only minimal deviations, for example, the predictions achieved an error of less than or equal to 1 MPa for yield strength and fatigue strength, less than 1 MPa % for modulus of toughness, less than 0.001% deviation in composition, and exact ratings for corrosion resistance, weldability, and anodising response. These findings validate the reproducibility and precision of the framework for alloy optimisation and recycling within sub-clusters. The selected optimum alloys were further validated using an unseen dataset from the MatWeb database, which revealed only negligible variations in properties, thereby confirming the robustness and reliability of the methodology for 6xxx series optimisation.

A similar approach was applied to the **7xxx series** at T5, T6 and T7 tempering conditions, using a dataset of 78 alloys divided into two subsets: Cu-containing 54 alloys and low Cu 24 alloys. PCA and K-means clustering identified four primary clusters for the Cu-containing alloys and three for the low Cu alloys, which were further subdivided into 11 and 6 sub-clusters, respectively, each containing alloys with a more compact range of properties. The TOPSIS optimisation algorithm was then applied, resulting in the selection of 11 optimum alloys for the Cu-containing subset and 6 for the low Cu subset, each representing the performance characteristics of its respective sub-cluster. This confirmed that the optimised alloys captured a broad spectrum of properties and potential applications. These selections were validated using the recycling algorithm to perform alloy mixing within sub-clusters and were further confirmed with external data from MatWeb. The validation revealed only negligible deviations, for example, less than 1 MPa in yield strength and near-identical compositions compared to the optimum alloys.

The **2xxx series**, comprising 59 alloys at T6 tempering condition at 210 °C, was clustered into 6 parent groups and 15 sub-clusters. Application of the TOPSIS optimisation algorithm successfully identified 17 optimum alloys, each representing the performance characteristics of its respective sub-cluster and covering a wide range of potential applications. These results were verified using recycling algorithm, which generated mixing ratios for all alloys within each sub-cluster, ensuring that the resulting mixtures reproduced the optimum alloys selected by the optimisation algorithm. This optimisation reduced the solution space of the 2xxx series from 59 alloys to 17 optimum alloys, covering a broad range of properties and potential applications.

In the **5xxx series** alloys at H32 hardening condition, 71 alloys were grouped into 3 clusters and 11 sub-clusters. Optimisation algorithm reduced these to 12 optimum alloys, which were validated through alloy recycling algorithm, showing only negligible variations in properties. This optimisation reduced the solution space of the 5xxx series from 71 alloys to 12 optimal alloys.

Similarly, in the **3xxx series**, 36 alloys collected under H32 hardening conditions were grouped into 4 clusters and 9 sub-clusters, resulting in 9 optimum alloys. Recycling analysis confirmed that these alloys could be consistently reproduced within the sub-clusters, with only minor deviations in properties, effectively reducing the number of alloys from 36 to 9 optimal compositions.

Some alloys resisted recycling and were retained as optimum alloys due to their unique combination of properties. The optimisation framework proved highly effective in condensing large and complex datasets into compact, manageable sets of representative optimum alloys that retain, and in some cases exceed, the original performance criteria while covering a broader range of applications. This systematic reduction of alloy proliferation supports greater standardisation and enhances recyclability, enabling the selection of fewer, more versatile “optimum alloys” for industrial use.

9.1.4 Metallurgical Interpretation Validates Data-Driven Results

The results of the PCA and K-means clustering analysis were critically assessed using metallurgical reasoning to validate their scientific soundness in addition to the algorithmic evaluation. This ensured that the outcomes of our design framework were consistent with established physical principles governing alloy behaviour, as well as being statistically coherent.

Metallurgical interpretation of the **6xxx series** was first performed within each cluster. Literature-based analysis indicated that compositional variations in Mg, Si, the Mg:Si ratio, Cu, Mn, and Fe were the primary factors influencing key properties, particularly yield strength, modulus of toughness, corrosion resistance, weldability, and anodising response, which were further affected by phase formation.

For example, in Cluster 4, elevated levels of Mg and Cu, together with a high total solute content (Mg + Si + Cu), promoted the formation of strengthening phases such as β'' and Q' , resulting in significantly enhanced strength. This was further supported by LIME XAI analysis, which showed that Fe had a negative effect on yield strength, although its proportion was low, while Mg and Cu were the most significant positive contributors. A high-volume fraction of strengthening phases (Mg_2Si and Q) in Cluster 4 was also confirmed by Thermo-Calc single-point simulations at T6 condition and atmospheric pressure, closely matching the observed high mechanical performance.

Similarly, Cluster 2 exhibited a lower strength range due to reduced formation of β'' and Q' phases resulting from lower solute content. This observation was corroborated by LIME analysis, literature-based metallurgical reasoning, and Thermo-Calc simulations. Other clusters in the 6xxx series displayed comparable metallurgical trends, providing a clear rationale for the clustering outcomes and confirming the physical basis underlying the data-driven classifications.

In the **7xxx series**, LIME analysis and literature-based metallurgical reasoning consistently explained the rationale behind the clustering of the two subsets of 7xxx series alloys: high-Cu and low-Cu alloys

For the high-Cu subset, Cluster 2 exhibited elevated strength, primarily due to increased levels of Zn, Mg, Cu, and Zr, resulting in a high total solute content (Mg + Zn + Cu), which LIME analysis identified as a strong positive contributors to yield strength. Literature-based analysis indicated that this composition promoted the formation of GP zones, η' , and η phases, contributing to Orowan strengthening mechanisms. High Zr levels facilitated grain refinement through the formation of Al_3Zr dispersoids, enhancing the modulus of toughness and resistance to recrystallisation, while Cu accelerated precipitation kinetics. However, the high Zn content increased susceptibility to stress corrosion cracking (SCC), reflecting a well-documented trade-off in the literature. These microstructure-related phenomena explain the observed

compromises between strength, SCC resistance, and weldability in these clusters. Similar reasoning applies to the other clusters within this subset.

In contrast, for the low-Cu subset, Cluster 1 exhibited medium-high strength, primarily due to high (Zn + Mg) contents promoting η' phase formation. The absence of Cu led to slower precipitation kinetics and lower phase density compared with the Cu-bearing subset, resulting in relatively lower strength performance. LIME explanations further supported these observations. Other clusters in the low-Cu subset also displayed consistent metallurgical trends, validating the clustering results.

In the **2xxx series**, Clusters 0 and 4, comprising alloys with high yield and creep strength, were primarily strengthened by the precipitation of Ω , S' , S'' , and θ' phases, with Cu and Ag serving as the principal drivers of precipitation. Mg–Ag co-clustering accelerated Ω -phase formation, while Cu further enhanced phase stability and precipitation density. Also, the lower-performing Cluster 3 exhibited reduced performance, which was attributed to grain boundary phase segregation (e.g., T-phase, Al_2Cu , $\text{Al}_{11}\text{Cu}_5\text{Mn}_3$). These interpretations were supported by Thermo-Calc single-point phase calculations at T6 condition, which predicted similar S-phase and Al_2Cu -phase volume fractions for these clusters.

For the **5xxx series**, Cluster 1 comprised high-strength alloys, primarily due to elevated Mg content combined with moderate Mn and Zn, which promoted work hardening and the formation of $\beta\text{-Al}_8\text{Mg}_5$. Mn contributed through the formation of Al_6Mn dispersoids, refining grains and further enhancing strength. Cluster 0, characterised by higher Fe-bearing intermetallics, exhibited reduced strength. These literature-based findings were supported by Thermo-Calc volume phase predictions at the H32 condition and atmospheric pressure, which indicated similar phase formation across clusters and confirmed that the observed strengthening was attributable to these phases.

In the **3xxx series**, the high-strength Cluster 3 was characterised by elevated Mn content, which promoted the formation of Al_6Mn and $\text{Al}_6(\text{Mn,Fe})\text{Si}$ dispersoids. These dispersoids inhibited dislocation motion, thereby increasing strength and enhancing corrosion resistance. Weldability was moderately affected by impurity segregation at grain boundaries, consistent with microcrack formation observed in lower-performing clusters. Thermo-Calc analysis further validated these observations, showing a high-volume fraction of Al_6Mn and Al_{12}Mn formation at

H32 condition and atmospheric pressure in Cluster 3. Similar findings were obtained for the other clusters, confirming consistent metallurgical trends.

Collectively, the metallurgical analyses validated the data-driven findings across all selected aluminium alloy series. The observed clustering and optimisation outcomes were not arbitrary but reflect intrinsic physical and microstructural variations in alloy behaviour. This dual validation, encompassing both algorithmic and metallurgical evidence, demonstrates the robustness, reliability, and scientific validity of the proposed methodology for alloy system simplification and recycling.

9.1.5 Summary of Selected Optimal-Performing Alloys Across Aluminium Series

The optimisation framework successfully reduced the number of alloys across all studied series while retaining the full range of compositional and property diversity. Specifically, the 6xxx series was reduced from 41 to 11 optimum alloys, the high-Cu 7xxx subset from 54 to 11, and the low-Cu 7xxx subset from 24 to 6 alloys. In the 2xxx series, 59 alloys were condensed to 17 optimum alloys, while 71 alloys in the 5xxx series were reduced to 12, and 36 alloys in the 3xxx series to 9 optimum alloys. These reductions capture the representative performance of each series, enabling more efficient selection, improved recyclability, and system standardisation, as summarised in Table 9.1.

Table 9.1. Reduced Set of Optimum Aluminium Alloys Across All Series

Series	Cluster	Selected Optimum Alloy
2xxx Series	0	2117, 2013, 2006, 2010
	1	2124, 2027, 2021
	2	2218, 2018, 2618
	3	2012, 2011
	4	2040, 2029
	5	2122, 2001, 2004
3xxx Series	0	3307, 3020
	1	3103, 3009
	2	3102, 3130

	3	3403, 3104, 3016
5xxx Series	0	5043, 5140, 5351, 5449, 5151
	1	5352, 5954, 5483, 5083
	2	5556, 5052, 5051
6xxx Series	0	6064, 6040
	1	6082, 6009, 6005
	2	6063, 6463
	3	6055, 6070
	4	6092, 6011A
7xxx Series (Low Cu Sub-set)	0	7028, 7039
	1	7278, 7035
	2	7005, 7024
7xxx Series (High Cu Sub-set)	0	7023, 7014
	1	7041, 7033
	2	7136, 7249, 7085, 7034
	3	7149, 7081, 7001

9.1.6 Validation of Optimisation Framework and Mixed-Series Recycling

To validate the optimisation framework across multiple series and assess its industrial relevance, a comparative analysis was conducted on two sets of 3xxx, 5xxx, and 6xxx series alloys. Set-1 comprised 33 commonly available mixed industrial scrap alloys, while Set-2 included 32 pre-selected optimum alloys using our design framework. Both sets were clustered separately using PCA and K-means clustering, revealing application-specific groupings. In Set-1, five well-defined clusters emerged: Cluster 2 contained alloys used in cans, utensils, building, and automotive applications, whereas Clusters 3 and 4 were predominantly automotive alloys. Set-2 clustering resulted in three clusters with broader sectoral applicability.

Within each cluster, the TOPSIS optimisation and recycling algorithms were applied to generate top-performing recycled alloys. Set-2, pre-selected optimum alloys for high performance and broad application coverage, yielded 3 top-performing recycled alloys with superior properties and versatility. Set-1, mixed industrial scrap alloys, although more application-specific

and moderate in properties, were successfully optimised and recycled within their clusters, producing five top-performing alloys with limited applicability.

This comparison confirms two key points: first, pre-selection of optimum alloys (set 2) prior to recycling enhances the performance and application potential of the resulting recycled alloys; second, even standard mixed industrial scrap (set 1) can be systematically upgraded using the clustering and recycling methodology. Overall, this demonstrates the robustness, flexibility, and industrial applicability of the optimisation and recycling framework across both standard industrial scrap and optimised alloy sets.

9.1.7 User-Friendly Recycling Software Application Enables Optimisation for Practical Use

Building upon the comprehensive optimisation and recycling frameworks developed in this research, a user-friendly software application was developed to facilitate practical alloy recycling, specifically focusing on **6xxx series alloys** at all stages of artificial ageing. The application allows users to input 6xxx series compositions (Si, Mg, Cu, Mn, Fe) and key mechanical properties, including yield strength, fatigue strength, and modulus of toughness. Users can also select tempering conditions (T5, T6, or T7) to tailor alloy recommendations according to specific manufacturing requirements.

Utilising advanced data-driven techniques, incorporating the TOPSIS optimisation and the designed recycling algorithms, the software generates optimised recycled alloy compositions and properties with precise mixing ratios based on TOPSIS closeness scores. Internally, the program performs recycling by assigning a minimum mixing ratio of 50% to the alloy with the highest closeness scores, while the remaining 50% is distributed among other user-input alloys, ensuring metallurgical consistency and maximising the utilisation of scrap. This approach produces recycled alloys that achieve optimal combinations of properties. The software application also identifies standard alloys from the database whose compositions closely match those of the recycled alloys, thereby aligning its recommendations with established industrial standards. Validation indicates that the predicted recycled alloys match the specified mechanical properties, with errors generally below 5–10 MPa for yield and fatigue strength and compositional deviations under 0.001%. It should be noted that these errors are not experimental errors, but rather the differences observed when comparing the predicted properties and compositions

of the recycled alloys against the dataset of standard alloys used in this research, reflecting the accuracy of the predictive model rather than measurement uncertainties.

The software has been further extended to incorporate all aluminium series (1xxx–8xxx), with enhanced features such as an Artificial Neural Network (ANN) for predicting properties of user-input compositions and PCA and K-means clustering for scrap sorting. This allows accurate sorting based on scrap batch size, property prediction, and recommendation of optimum mixing ratios beyond the 6xxx series, significantly increasing the tool's versatility and industrial applicability.

This software application effectively translates the data-driven alloy design and recycling methodology developed in this study into an accessible platform, promoting sustainable alloy recycling and supporting the principles of the circular economy. Designed for implementation across the aluminium manufacturing industry, it provides a reliable tool for producing performance-driven recycled alloys through complete scrap utilisation, eliminating the need for primary material additions and optimising material reuse.

9.2 Challenges and Limitations

While the proposed methodology provides a systematic and efficient approach to alloy classification, optimisation, and recycling, several challenges and limitations must be recognised:

- **General vs. Niche Applications:** The reduction in alloy grades may not fully accommodate niche applications with highly specialised property requirements. Although clustering ensures broad coverage across typical property ranges, certain unique cases may still necessitate customised alloy selection.
- **Experimental Validation:** The study of optimum alloy selection was primarily based on established reference datasets, existing metallurgical principles, and Thermo-Calc software predictions for alloy properties and compositional distributions. Experimental validation is required to confirm the performance of optimised alloys under real-world conditions and to verify the predicted mixing ratios generated by the recycling algorithm when designing recycled optimum alloy compositions within sub-clusters.

9.3 Future Work

This research provides a foundation for further advancements in aluminium alloy classification, optimisation, and recycling. The following directions are proposed for future work:

- **Integration of Process Variables:** Future studies should incorporate additional thermomechanical processing steps, such as stamping and other forming operations, as well as heat treatment responses beyond artificial ageing, selected hardening conditions, and manufacturing constraints, into the optimisation framework. The current methodology is designed to be adaptable, enabling the integration of these additional treatments and manufacturing considerations to enhance predictive accuracy and the practical applicability of the optimised alloys.
- **Experimental Validation:** Experimental investigations will be necessary to validate the properties of recycled alloys using the mixing ratios recommended by the recycling algorithm. Collaboration with industrial partners is encouraged to implement the technique in real-world alloy recycling processes. Furthermore, Thermo-Calc simulations should be extended to more complex aluminium systems, such as the 7xxx series, to validate phase formation in each cluster for both high-Cu and low-Cu alloy subsets.
- **Extending the Alloy Recycling Software Application:** The software can be further developed into a fully integrated industrial tool by directly interfacing with process instrumentation, such as elemental analysis techniques for composition measurement. The envisioned workflow is as follows: scrap material → composition data → input to the recycling software for high-precision sorting and optimal mixing ratio determination → implementation of the determined mixing ratios in the recycling plant to produce high-performance recycled alloys without the additions of primary material.

To ensure continuous optimisation and quality control, a digital twin of the entire recycling process can be established. This would include modules for scrap collection, sorting, melting, casting, and pre-treatment, enabling real-time tracking, predictive monitoring, and process optimisation. The software application can be integrated at the scrap sorting and melting stages, where it will efficiently sort the scrap and determine the correct mixing ratios to produce high-quality secondary aluminium, a capability that is currently lacking in the industry. Together, these advancements would transform the current application into a robust, industry-ready solution for automated aluminium

alloy sorting and recycling, supporting efficient circular manufacturing and sustainable production.

9.4 Practical Implications

The findings of this research have significant implications for industrial application, particularly in the utilisation of selected optimum alloys and the practical deployment of the recycling software application for aluminium alloy recycling. The software is presented in two stages: the first stage focuses on recycling and optimisation of the 6xxx series alloys, while the second stage extends its capabilities to incorporate all aluminium series (1xxx–8xxx) with enhanced features, including an Artificial Neural Network (ANN) for property prediction and sorting.

6xxx Series Recycling Software Application: The software provides a systematic and user-friendly means of predicting the compositions, properties of recycled alloys and optimising mixing ratios through the integration of the TOPSIS optimisation algorithm and the designed recycling algorithm. Users can input alloy compositions and properties, select tempering conditions (T5, T6, or T7) and receive recommendations for optimum mixing ratios, recycled alloy compositions and properties along with the closest matching standard alloy grades. The application enables practical industrial deployment and facilitates sustainable aluminium recycling.

Extended Application for 1xxx–8xxx Series: The software has been further enhanced with an Artificial Neural Network (ANN) to predict the properties of user-input compositions across the 1xxx–8xxx series aluminium alloys. It also employs PCA and K-means algorithms for alloy sorting, depending on scrap batch size. These features enable accurate classification and recommendation of optimum mixing ratios for recycling beyond the 6xxx series.

Overall, these enhanced capabilities increase the software's versatility and industrial relevance, supporting efficient scrap recycling without relying on primary metal additions and formulation across sectors such as automotive, aerospace, and construction.

9.5 Concluding Comments

This thesis establishes and implements a machine learning-based design framework for alloy sorting, optimisation, and the prediction of recycling outcomes across a broad range of aluminium alloy series. By leveraging clustering algorithm, multi-criteria decision-making algorithms, the designed recycling algorithm, and LIME explainable AI techniques, alongside

literature-based metallurgical reasoning validated by Thermo-Calc simulations, the study presents a novel approach to simplifying alloy selection while retaining key performance characteristics. This framework has been further developed into a dedicated recycling software application. The findings demonstrate the potential of ML- and AI-based methods to transform material design, selection, and recycling, offering a more efficient and sustainable pathway for the industrial utilisation of a resources.

CHAPTER 10: REFERENCES

- [1] J. R. Davis, *Aluminum and aluminum alloys*. ASM international, 1993.
- [2] L. Granta-Design, “CES-Edupack,” 2019, *Granta Design Limited: Cambridge, UK*.
- [3] J. R. Davis, *Alloying: understanding the basics*. ASM international, 2001.
- [4] C.-L. Tai *et al.*, “The nano-structural characterization of Ω and S phases in Al-5.1 Cu-1.0 Mg-(0.4 Ag) AA2024 aluminum alloys,” *Materials Science and Engineering: A*, vol. 881, p. 145361, 2023.
- [5] T. Liu, M. Karkkainen, L. Nastac, V. Arvikar, I. Levin, and L. N. Brewer, “Iron-rich intermetallics in high pressure die cast A383 aluminum alloys,” *Intermetallics (Barking)*, vol. 126, p. 106814, 2020.
- [6] P. Rambabu, N. Eswara Prasad, V. V. Kutumbarao, and R. J. H. Wanhill, “Aluminium alloys for aerospace applications,” *Aerospace materials and material technologies: volume 1: aerospace materials*, pp. 29–52, 2016.
- [7] ASM Handbook Committee, *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials*, vol. 2. Materials Park, OH: ASM International, 1990. doi: <https://doi.org/10.31399/asm.hb.v02.9781627081627>.
- [8] M. Baruah and A. Borah, “Processing and precipitation strengthening of 6xxx series aluminium alloys: A review,” *Int. J. Mater. Sci*, vol. 1, no. 1, pp. 40–48, 2020.
- [9] X. Zhu, H. Yang, X. Dong, and S. Ji, “The effects of varying Mg and Si levels on the microstructural inhomogeneity and eutectic Mg₂Si morphology in die-cast Al–Mg–Si alloys,” *J. Mater. Sci.*, vol. 54, no. 7, pp. 5773–5787, 2019.
- [10] Righton Blackburns, “Aluminium Alloy Temper Designations,” Righton Blackburns. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.rightonblackburns.co.uk/products/metals/aluminium-alloys/aluminium-alloy-temper-designations>
- [11] X. Zhu, H. Yang, X. Dong, and S. Ji, “The effects of varying Mg and Si levels on the microstructural inhomogeneity and eutectic Mg₂Si morphology in die-cast Al–Mg–Si alloys,” *J. Mater. Sci.*, vol. 54, no. 7, pp. 5773–5787, 2019.
- [12] K. F. Karhausen and A. S. Korhonen, “Rolling of Aluminum,” in *Encyclopedia of Aluminum and Its Alloys, Two-Volume Set (Print)*, CRC Press, 2018, pp. 2407–2426.
- [13] C. Nappi, “The global aluminium industry 40 years from 1972,” *World aluminium*, vol. 69, pp. 379–387, 2013.
- [14] T. R. Beck, “A new energy-efficient and environmentally friendly process to produce aluminum,” *Jom*, vol. 65, no. 2, pp. 267–271, 2013.
- [15] J. Hirsch, “Recent development in aluminium for automotive applications,” *Transactions of Nonferrous Metals Society of China*, vol. 24, no. 7, pp. 1995–2002, 2014.
- [16] Bureau of International Recycling (BIR), “Bureau of International Recycling (BIR).” Accessed: Aug. 23, 2025. [Online]. Available: <https://www.bir.org>
- [17] Italtres, “EU aluminum demand to 2050: Primary and recycled metal ratio, parity objective.” Accessed: Aug. 23, 2025. [Online]. Available: <https://www.italpres.com/aluminum-die-casting-news/eu-aluminum-demand-2050-primary-recycled-metal-ratio>
- [18] E. Aluminum, “European aluminum,” *Belgium, Brussels*, 2019.

- [19] A. Gutschi, “End-of-life vehicle recycling in the European Union: Analysing changing material flows of end-of-life steel, aluminum, copper and plastics due to the transition toward zero-emission vehicles,” Master’s thesis, Technische Universität Wien, 2022. Accessed: Sep. 09, 2025. [Online]. Available: <http://hdl.handle.net/20.500.12708/20446>
- [20] A. Vuk, I. Szűcs, and A. Bauerné Gáthy, “Packaging waste and recycle in EU,” *International Review of Applied Sciences and Engineering*, vol. 14, no. 3, pp. 402–412, 2023.
- [21] S. Capuzzi and G. Timelli, “Preparation and melting of scrap in aluminum recycling: A review,” *Metals (Basel)*, vol. 8, no. 4, p. 249, 2018.
- [22] S. K. Das, “Emerging trends in aluminum recycling: Reasons and responses,” *Light Metals*, vol. 4, pp. 911–916, 2006.
- [23] V. Kevorkijan, “Advances in recycling of wrought aluminium alloys for added value maximisation,” *Metallurgia*, vol. 16, pp. 103–114, 2010.
- [24] D. B. Spencer, “The high-speed identification and sorting of nonferrous scrap,” *Jom*, vol. 57, no. 4, pp. 46–51, 2005.
- [25] A. Gesing and R. Wolanski, “Recycling light metals from end-of-life vehicle,” *Jom*, vol. 53, no. 11, pp. 21–23, 2001.
- [26] S. Koyanaka, K. Kobayashi, Y. Yamamoto, M. Kimura, and K. Rokuchō, “Elemental analysis of lightweight metal scraps recovered by an automatic sorting technique combining a weight meter and a laser 3D shape-detection system,” *Resour. Conserv. Recycl.*, vol. 75, pp. 63–69, 2013.
- [27] G. Gaustad, E. Olivetti, and R. Kirchain, “Improving aluminum recycling: A survey of sorting and impurity removal technologies,” *Resour. Conserv. Recycl.*, vol. 58, pp. 79–87, 2012.
- [28] S. Koyanaka and K. Kobayashi, “Automatic sorting of lightweight metal scrap by sensing apparent density and three-dimensional shape,” *Resour. Conserv. Recycl.*, vol. 54, no. 9, pp. 571–578, 2010.
- [29] K. Tsuchiya *et al.*, “Establishment of aluminum ‘Sash to Sash’ Recycling by using XRT and XRF Sorters,” in *Proceedings of the 11th International Conference on Mining, Materials and Petroleum Engineering, Chiang Mai, Thailand*, 2013, pp. 11–13.
- [30] X.-L. Huang, A. El Badawy, M. Arambewela, R. Ford, M. Barlaz, and T. Tolaymat, “Characterization of salt cake from secondary aluminum production,” *J. Hazard. Mater.*, vol. 273, pp. 192–199, 2014.
- [31] S. Capuzzi and G. Timelli, “Preparation and melting of scrap in aluminum recycling: A review,” *Metals (Basel)*, vol. 8, no. 4, p. 249, 2018.
- [32] J. Mangos and N. Birbilis, “Aluminium alloy design and discovery using machine learning,” *arXiv preprint arXiv:2105.14806*, 2021.
- [33] D. G. Sangiovanni, *Transition Metal Nitrides Alloy Design and Surface Transport Properties using Ab-initio and Classical Computational Methods*. Linköpings Universitet (Sweden), 2013.
- [34] A. M. Turing, “Computing machinery and intelligence (1950),” *Mind*, vol. 59, no. 236, pp. 33–60, 2021.
- [35] A. Newell and H. Simon, “The logic theory machine--A complex information processing system,” *IRE Transactions on information theory*, vol. 2, no. 3, pp. 61–79, 1956.
- [36] J. McCarthy and E. A. Feigenbaum, “In memoriam: Arthur samuel: Pioneer in machine learning,” *AI Mag.*, vol. 11, no. 3, p. 10, 1990.

- [37] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [38] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [39] Honavar V, “Artificial intelligence: An overview,” Pennsylvania State University. Accessed: Jul. 28, 2023. [Online]. Available: <https://faculty.ist.psu.edu/vhonavar/Courses/ai/handout1.pdf>
- [40] S. Ray, “A quick review of machine learning algorithms,” in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, IEEE, 2019, pp. 35–39.
- [41] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [42] Feigenbaum EA, *Computers and Thought*. New York: McGraw-Hill, 1963.
- [43] H. A. Shehu, A. Rabie, and M. H. Sharif, “Artificial intelligence tools and their capabilities,” *PLOMS AI*, vol. 1, no. 1, 2021.
- [44] Microsoft, “Artificial intelligence (AI) vs. machine learning (ML),” Microsoft Azure. Accessed: Sep. 14, 2025. [Online]. Available: <https://azure.microsoft.com/en-au/resources/cloud-computing-dictionary/artificial-intelligence-vs-machine-learning>
- [45] W.-M. Lee, *Python machine learning*. John Wiley & Sons, 2019.
- [46] S. Badillo *et al.*, “An introduction to machine learning,” *Clin. Pharmacol. Ther.*, vol. 107, no. 4, pp. 871–885, 2020.
- [47] J. Zupan, “Introduction to artificial neural network (ANN) methods: what they are and how to use them,” *Acta Chim. Slov.*, vol. 41, no. 3, p. 327, 1994.
- [48] S. K. Mukhiya and U. Ahmed, *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd, 2020.
- [49] G. Dong and H. Liu, *Feature engineering for machine learning and data analytics*. CRC press, 2018.
- [50] V. Barnett, “Principles and methods for handling outliers in data sets,” in *Statistical Methods and the Improvement of Data Quality*, Elsevier, 1983, pp. 131–166.
- [51] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, “Effect of data scaling methods on machine learning algorithms and model performance,” *Technologies (Basel)*, vol. 9, no. 3, p. 52, 2021.
- [52] P. J. M. Ali, R. H. Faraj, E. Koya, P. J. M. Ali, and R. H. Faraj, “Data normalization and standardization: a technical report,” *Mach Learn Tech Rep*, vol. 1, no. 1, pp. 1–6, 2014.
- [53] M. Franzese and A. Iuliano, “Correlation analysis,” in *Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics*, vol. 1, Elsevier, 2018, pp. 706–721.
- [54] A. O. Sykes, “An introduction to regression analysis,” 1993.
- [55] P. F. Smith, S. Ganesh, and P. Liu, “A comparison of random forest regression and multiple linear regression for prediction in neuroscience,” *J. Neurosci. Methods*, vol. 220, no. 1, pp. 85–91, 2013.
- [56] A. Liaw and M. Wiener, “Classification and regression by randomForest,” *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [57] P. C. Sen, M. Hajra, and M. Ghosh, “Supervised classification algorithms in machine learning: A survey and review,” in *Emerging technology in modelling and graphics: Proceedings of IEM graph 2018*, Springer, 2019, pp. 99–111.

- [58] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, 1993.
- [59] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access*, vol. 8, pp. 80716–80727, 2020.
- [60] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, 2005.
- [61] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, "DBSCAN: Past, present and future," in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, IEEE, 2014, pp. 232–238.
- [62] K. Virupakshappa and E. Oruklu, "Unsupervised machine learning for ultrasonic flaw detection using gaussian mixture modeling, k-means clustering and mean shift clustering," in *2019 IEEE international ultrasonics symposium (IUS)*, IEEE, 2019, pp. 647–649.
- [63] M. Pelikan, "Bayesian optimization algorithm," in *Hierarchical Bayesian optimization algorithm: toward a new generation of evolutionary algorithms*, Springer, 2005, pp. 31–48.
- [64] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, 2000.
- [65] E. Roszkowska, "Multi-criteria decision making models by applying the TOPSIS method to crisp and interval data," *Multiple Criteria Decision Making/University of Economics in Katowice*, vol. 6, no. 1, pp. 200–230, 2011.
- [66] J. Papathanasiou and N. Ploskas, "Topsis," in *Multiple criteria decision aid: Methods, examples and python implementations*, Springer, 2018, pp. 1–30.
- [67] B. Ceballos, M. T. Lamata, and D. A. Pelta, "A comparative analysis of multi-criteria decision-making methods," *Progress in artificial intelligence*, vol. 5, no. 4, pp. 315–322, 2016.
- [68] D. Garreau and U. Luxburg, "Explaining the explainer: A first theoretical analysis of LIME," in *International conference on artificial intelligence and statistics*, PMLR, 2020, pp. 1287–1296.
- [69] T. Mueller, A. G. Kusne, and R. Ramprasad, "Machine learning in materials science: Recent progress and emerging applications," *Reviews in computational chemistry*, vol. 29, pp. 186–273, 2016.
- [70] J. F. Durodola, "Machine learning for design, phase transformation and mechanical properties of alloys," *Prog. Mater. Sci.*, vol. 123, p. 100797, 2022.
- [71] J. Wei *et al.*, "Machine learning in materials science," *InfoMat*, vol. 1, no. 3, pp. 338–358, 2019.
- [72] Y. Kumar, K. Kaur, and G. Singh, "Machine learning aspects and its applications towards different research areas," in *2020 International conference on computation, automation and knowledge management (ICCAKM)*, IEEE, 2020, pp. 150–156.
- [73] G. Liu, L. Jia, B. Kong, S. Feng, H. Zhang, and H. Zhang, "Artificial neural network application to microstructure design of Nb-Si alloy to improve ultimate tensile strength," *Materials Science and Engineering: A*, vol. 707, pp. 452–458, 2017.
- [74] M. Buscema, "Back propagation neural networks," *Subst. Use Misuse*, vol. 33, no. 2, pp. 233–270, 1998.

- [75] Y. Xw, Z. H. U. Jc, N. Zs, L. A. I. Zh, and L. I. U. Fw, “Prediction of mechanical properties of A357 alloy using artificial neural network,” *Transactions of Nonferrous Metals Society of China*, vol. 3, no. 23, pp. 788–795, 2013.
- [76] M. V Pathan *et al.*, “Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning,” *Sci. Rep.*, vol. 9, no. 1, p. 13964, 2019.
- [77] S. A. Razavi, F. Ashrafizadeh, and S. Fooladi, “Prediction of age hardening parameters for 17-4PH stainless steel by artificial neural network and genetic algorithm,” *Materials Science and Engineering: A*, vol. 675, pp. 147–152, 2016.
- [78] C. Wang, H. Fu, L. Jiang, D. Xue, and J. Xie, “A property-oriented design strategy for high performance copper alloys via machine learning,” *NPJ Comput. Mater.*, vol. 5, no. 1, p. 87, 2019.
- [79] Y. Liu *et al.*, “Predicting creep rupture life of Ni-based single crystal superalloys using divide-and-conquer approach based machine learning,” *Acta Mater.*, vol. 195, pp. 454–467, 2020.
- [80] W. A. Saidi, W. Shadid, and I. E. Castelli, “Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network,” *NPJ Comput. Mater.*, vol. 6, no. 1, p. 36, 2020.
- [81] J. D. Evans and F.-X. Coudert, “Predicting the mechanical properties of zeolite frameworks by machine learning,” *Chemistry of Materials*, vol. 29, no. 18, pp. 7833–7839, 2017.
- [82] A. H. Mashhadzadeh *et al.*, “Machine Learning-Assisted design of boron and nitrogen doped graphene nanosheets with tailored thermomechanical properties,” *Comput. Mater. Sci.*, vol. 240, p. 112998, 2024.
- [83] R. Wang, S. Zeng, X. Wang, and J. Ni, “Machine learning for hierarchical prediction of elastic properties in Fe-Cr-Al system,” *Comput. Mater. Sci.*, vol. 166, pp. 119–123, 2019.
- [84] J. Xiong, T. Zhang, and S. Shi, “Machine learning of mechanical properties of steels,” *Sci. China Technol. Sci.*, vol. 63, no. 7, pp. 1247–1255, 2020.
- [85] I. Di Martino, J. W. Brooks, P. A. S. Reed, P. Holdway, and A. Wisbey, “Adaptive numerical modelling of high temperature strength, creep and fatigue behaviour in Ni based superalloys,” *Materials Science and Technology*, vol. 23, no. 12, pp. 1402–1407, 2007.
- [86] G. B. Goh, C. Siegel, A. Vishnu, N. O. Hodas, and N. Baker, “Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models,” *arXiv preprint arXiv:1706.06689*, 2017.
- [87] M. Kusaba, C. Liu, and R. Yoshida, “Crystal structure prediction with machine learning-based element substitution,” *Comput. Mater. Sci.*, vol. 211, p. 111496, 2022.
- [88] L. Piloizzi, F. A. Farrelly, G. Marcucci, and C. Conti, “Topological nanophotonics and artificial neural networks,” *Nanotechnology*, vol. 32, no. 14, p. 142001, 2021.
- [89] P. Raccuglia *et al.*, “Machine-learning-assisted materials discovery using failed experiments,” *Nature*, vol. 533, no. 7601, pp. 73–76, 2016.
- [90] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, “A general-purpose machine learning framework for predicting properties of inorganic materials,” *NPJ Comput. Mater.*, vol. 2, no. 1, pp. 1–7, 2016.

- [91] U. Bork and R. E. Challis, "Non-destructive evaluation of the adhesive fillet size in a T-peel joint using ultrasonic Lamb waves and a linear network for data discrimination," *Meas. Sci. Technol.*, vol. 6, no. 1, p. 72, 1995.
- [92] A. Mukherjee, S. Schmauder, and M. Ru, "Artificial neural networks for the prediction of mechanical behavior of metal matrix composites," *Acta metallurgica et materialia*, vol. 43, no. 11, pp. 4083–4091, 1995.
- [93] S. Park *et al.*, "High strength aluminum alloys design via explainable artificial intelligence," *J. Alloys Compd.*, vol. 903, p. 163828, 2022.
- [94] S. M. Golovanov, V. I. Orlov, L. A. Kazakovtsev, and A. M. Popov, "Recursive clustering algorithm based on silhouette criterion maximization for sorting semiconductor devices by homogeneous batches," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2019, p. 022035.
- [95] A. L. Kazakovtsev, A. N. Antamoshkin, and V. V Fedosov, "Greedy heuristic algorithm for solving series of EEE components classification problems," in *IOP conference series: materials science and engineering*, IOP Publishing, 2016, p. 012011.
- [96] M. Ziatdinov, A. Maksov, L. Li, A. S. Sefat, P. Maksymovych, and S. V Kalinin, "Deep data mining in a real space: separation of intertwined electronic responses in a lightly doped BaFe₂As₂," *Nanotechnology*, vol. 27, no. 47, p. 475706, 2016.
- [97] C. Yeung *et al.*, "Elucidating the behavior of nanophotonic structures through explainable machine learning algorithms," *ACS Photonics*, vol. 7, no. 8, pp. 2309–2318, 2020.
- [98] M. J. Eaton, R. Pullin, J. J. Hensman, K. M. Holford, K. Worden, and S. L. Evans, "Principal component analysis of acoustic emission signals from landing gear components: an aid to fatigue fracture detection," *Strain*, vol. 47, pp. e588–e594, 2011.
- [99] J.-C. Golinval, "Damage detection in structures based on principal component analysis of forced harmonic responses," *Procedia Eng.*, vol. 199, pp. 1912–1918, 2017.
- [100] A. Behnia, H. K. Chai, N. Ranjbar, and M. Z. Jumaat, "Damage detection of SFRC concrete beams subjected to pure torsion by integrating acoustic emission and Weibull damage function," *Struct. Control Health Monit.*, vol. 23, no. 1, pp. 51–68, 2016.
- [101] G. Manson, K. Worden, K. Holford, and R. Pullin, "Visualisation and dimension reduction of acoustic emission data for damage detection," *J. Intell. Mater. Syst. Struct.*, vol. 12, no. 8, pp. 529–536, 2001.
- [102] Rossiter JA, "Analysis and prediction of properties of concrete and its component," PhD theses, Queensland University of Technology, Brisbane (AU), 1993.
- [103] N. Godin, S. Huguet, R. Gaertner, and L. Salmon, "Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers," *Ndt & E International*, vol. 37, no. 4, pp. 253–264, 2004.
- [104] S. Tayfur, N. Alver, S. Abdi, S. Saatci, and A. Ghiami, "Characterization of concrete matrix/steel fiber de-bonding in an SFRC beam: Principal component analysis and k-mean algorithm for clustering AE data," *Eng. Fract. Mech.*, vol. 194, pp. 73–85, 2018.
- [105] R. Yuan *et al.*, "The search for BaTiO₃-based piezoelectrics with large piezoelectric coefficient using machine learning," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 2, pp. 394–401, 2018.

- [106] H. Zhang, H. Fu, S. Zhu, W. Yong, and J. Xie, “Machine learning assisted composition effective design for precipitation strengthened copper alloys,” *Acta Mater.*, vol. 215, p. 117118, 2021.
- [107] D. Xue, P. V Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, “Accelerated search for materials with targeted properties by adaptive design,” *Nat. Commun.*, vol. 7, no. 1, pp. 1–9, 2016.
- [108] T. Lookman, P. V Balachandran, D. Xue, and R. Yuan, “Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design,” *NPJ Comput. Mater.*, vol. 5, no. 1, p. 21, 2019.
- [109] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, and B. Meredig, “High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates,” *Integr. Mater. Manuf. Innov.*, vol. 6, no. 3, pp. 207–217, 2017.
- [110] R. Khorshidi and A. Hassani, “Comparative analysis between TOPSIS and PSI methods of materials selection to achieve a desirable combination of strength and workability in Al/SiC composite,” *Materials & Design (1980-2015)*, vol. 52, pp. 999–1010, 2013.
- [111] B. Singaravel and T. Selvaraj, “Optimization of machining parameters in turning operation using combined TOPSIS and AHP method,” *Tehnički vjesnik*, vol. 22, no. 6, pp. 1475–1480, 2015.
- [112] W.-C. Yang, S.-H. Chon, C.-M. Choe, and J.-Y. Yang, “Materials selection method using TOPSIS with some popular normalization methods,” *Engineering Research Express*, vol. 3, no. 1, p. 015020, 2021.
- [113] Y.-X. Zhang, S.-J. Xie, W. Guo, J. Ding, L. H. Poh, and Z.-D. Sha, “Multi-objective optimization for high-performance Fe-based metallic glasses via machine learning approach,” *J. Alloys Compd.*, vol. 960, p. 170793, 2023.
- [114] Berboucha M, “Scientists use artificial intelligence to discover new materials,” *Forbes*. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.forbes.com/sites/meriamberboucha/2018/04/22/scientists-use-artificial-intelligence-to-discover-new-materials/>
- [115] European Aluminium, “SALEMA: Substituting critical raw materials in aluminium alloys,” European Aluminium. Accessed: Jul. 28, 2025. [Online]. Available: <https://european-aluminium.eu/projets/salema-substituting-critical-raw-materials-in-aluminium-alloys-for-electrical-vehicles/>
- [116] H. Zhan *et al.*, “Unified casting (UniCast) aluminum alloy—a sustainable and low-carbon materials solution for vehicle lightweighting,” *J. Mater. Sci. Technol.*, vol. 154, pp. 251–268, 2023.
- [117] Q. Liu *et al.*, “Compositionally flexible alloy design towards recycling mixed stainless steel scraps,” *J. Mater. Sci. Technol.*, vol. 225, pp. 227–239, 2025.
- [118] B. Campanella *et al.*, “Classification of wrought aluminum alloys by Artificial Neural Networks evaluation of Laser Induced Breakdown Spectroscopy spectra from aluminum scrap samples,” *Spectrochim. Acta Part B At. Spectrosc.*, vol. 134, pp. 52–57, 2017.
- [119] D. Diaz-Romero, W. Sterkens, S. Van den Eynde, T. Goedeme, W. Dewulf, and J. Peeters, “Deep learning computer vision for the separation of Cast-and Wrought-Aluminum scrap,” *Resour. Conserv. Recycl.*, vol. 172, p. 105685, 2021.
- [120] H. Wilts, B. R. Garcia, R. G. Garlito, L. S. Gómez, and E. G. Prieto, “Artificial intelligence in the sorting of municipal waste as an enabler of the circular economy,” *Resources*, vol. 10, no. 4, p. 28, 2021.

- [121] M. Auer, K. Osswald, R. Volz, and J. Woidasky, "Artificial intelligence-based process for metal scrap sorting," *arXiv preprint arXiv:1903.09415*, 2019.
- [122] P. B. Gedam, A. Khan, N. Purohit, and V. K. Jha, "Development of AI Based Computer Vision Scrap Sorting System for Aluminium Alloy Scrap," in *2025 International Conference on Innovative Trends in Information Technology (ICITIIT)*, IEEE, 2025, pp. 1–8.
- [123] S. Adhithya and J. Kathirvelan, "Development of Computer Vision Based Aluminium Scrap Sorting Using Real Time Metal Scrap Dataset," in *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, IEEE, 2024, pp. 1522–1527.
- [124] B. S. Costa *et al.*, "Artificial intelligence in automated sorting in trash recycling," in *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, SBC, 2018, pp. 198–205.
- [125] A. Abeygunawardhana, R. Shalinda, W. Bandara, W. D. S. Anesta, D. Kasthuriathna, and L. Abeysiri, "AI-driven smart bin for waste management," in *2020 2nd International Conference on Advancements in Computing (ICAC)*, IEEE, 2020, pp. 482–487.
- [126] K. H. Yu, Y. Zhang, D. Li, C. E. Montenegro-Marin, and P. M. Kumar, "Environmental planning based on reduce, reuse, recycle and recover using artificial intelligence," *Environ. Impact Assess. Rev.*, vol. 86, p. 106492, 2021.
- [127] M. Hu *et al.*, "Prediction of mechanical properties of wrought aluminium alloys using feature engineering assisted machine learning approach," *Metallurgical and Materials Transactions A*, vol. 52, no. 7, pp. 2873–2884, 2021.
- [128] B. Li *et al.*, "Manipulation of mechanical properties of 7xxx aluminum alloy via a hybrid approach of machine learning and key experiments," *Journal of Materials Research and Technology*, vol. 19, pp. 2483–2496, 2022.
- [129] S. Dey, N. Sultana, M. S. Kaiser, P. Dey, and S. Datta, "Computational intelligence based design of age-hardenable aluminium alloys for different temperature regimes," *Mater. Des.*, vol. 92, pp. 522–534, 2016.
- [130] N. Altinkok and R. Koker, "Neural network approach to prediction of bending strength and hardening behaviour of particulate reinforced (Al–Si–Mg)-aluminium matrix composites," *Mater. Des.*, vol. 25, no. 7, pp. 595–602, 2004.
- [131] J.-L. Tang, C.-Z. Cai, S.-J. Huang, and T.-T. Xiao, "Strength Prediction for Al–Cu–Mg–Ag Alloy Based on Support Vector Regression," *Journal of Aeronautical Materials*, vol. 32, no. 5, pp. 92–96, 2012.
- [132] S. Verma, J. P. Misra, and D. Popli, "Modeling of friction stir welding of aviation grade aluminium alloy using machine learning approaches," *International Journal of Modelling and Simulation*, vol. 42, no. 1, pp. 1–8, 2022.
- [133] J. Li *et al.*, "Accelerated discovery of high-strength aluminum alloys by machine learning," *Commun. Mater.*, vol. 1, no. 1, p. 73, 2020.
- [134] M. A. Devi, C. P. S. Prakash, R. P. Chinnannavar, V. P. Joshi, R. S. Palada, and R. Dixit, "An informatic approach to predict the mechanical properties of aluminium alloys using machine learning techniques," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2020, pp. 536–541.
- [135] Q. Kuang and L. Zhao, "A practical GPU based kNN algorithm," in *Proceedings. The 2009 International Symposium on Computer Science and Computational Technology (ISCSCI 2009)*, Academy Publisher, 2009, p. 151.
- [136] Z. Fang, K. Roy, J. Xu, Y. Dai, B. Paul, and J. B. P. Lim, "A novel machine learning method to investigate the web crippling behaviour of perforated roll-

- formed aluminium alloy unlipped channels under interior-two flange loading,” *Journal of Building Engineering*, vol. 51, p. 104261, 2022.
- [137] O. Kessler, “Encyclopedia of Aluminum and Its Alloys,” 2018.
 - [138] MakeItFrom, “Aluminium Alloys – Materials – Engineering,” Aluminium Alloys – Materials – Engineering. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.makeitfrom.com/material-group/Aluminium-Alloy>
 - [139] S. T. Alexander, “The mean square error (MSE) performance criteria,” in *Adaptive Signal Processing: Theory and Applications*, Springer, 1986, pp. 8–33.
 - [140] J. Mitlöhner, S. Neumaier, J. Umbrich, and A. Polleres, “Characteristics of open data CSV files,” in *2016 2nd International Conference on Open and Big Data (OBD)*, IEEE, 2016, pp. 72–79.
 - [141] D. Rolon-Mérette, M. Ross, T. Rolon-Mérette, and K. Church, “Introduction to Anaconda and Python: Installation and setup,” *Quant. Methods Psychol*, vol. 16, no. 5, pp. S3–S11, 2016.
 - [142] E. Bisong, “Google colaboratory,” in *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, Springer, 2019, pp. 59–64.
 - [143] H. Kuhn and D. Medlin, *Mechanical testing and evaluation*. ASM international, 2000.
 - [144] M. L. Waskom, “Seaborn: statistical data visualization,” *J. Open Source Softw.*, vol. 6, no. 60, p. 3021, 2021.
 - [145] C. Ding and X. He, “K-means clustering via principal component analysis,” in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 29.
 - [146] T. Tiwari, S. Jalalian, C. Mendis, and D. Eskin, “Classification of T6 tempered 6xxx series aluminum alloys based on machine learning principles,” *JOM*, vol. 75, no. 11, pp. 4526–4537, 2023.
 - [147] T. Tiwari, C. Mendis, and D. Eskin, “Facilitating recycling of 6xxx series aluminum alloys by machine Learning-Based optimization,” *Journal of Sustainable Metallurgy*, pp. 1–12, 2025.
 - [148] E. P. Series A, “Aluminium alloys and heat treatment,” CAB Worldwide. Accessed: Jul. 28, 2025. [Online]. Available: https://www.cabwww.com/uploads/case_studies/HeatTreatAlum-wp.pdf
 - [149] D. J. Chakrabarti and D. E. Laughlin, “Phase relations and precipitation in Al–Mg–Si alloys with Cu additions,” *Prog. Mater. Sci.*, vol. 49, no. 3–4, pp. 389–410, 2004.
 - [150] J. C. Pang, S. X. Li, Z. G. Wang, and Z. F. Zhang, “General relation between tensile strength and fatigue strength of metallic materials,” *Materials Science and Engineering: A*, vol. 564, pp. 331–341, 2013.
 - [151] J. M. Harrison *et al.*, “Principal variance component analysis of crop composition data: a case study on herbicide-tolerant cotton,” *J. Agric. Food Chem.*, vol. 61, no. 26, pp. 6412–6422, 2013.
 - [152] P. Sláma and J. Nachazel, “Effect of mechanical treatment on intergranular corrosion of 6064 alloy bars,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2017, p. 012064.
 - [153] R. Zettler, A. A. M. Da Silva, S. Rodrigues, A. Blanco, and J. F. dos Santos, “Dissimilar Al to Mg alloy friction stir welds,” *Adv. Eng. Mater.*, vol. 8, no. 5, pp. 415–421, 2006.
 - [154] W. J. K. J. Anderson K, *Properties and Selection of Aluminum Alloys*, vol. 2B. Materials Park (OH): ASM International, 2019.

- [155] V. R. Sivakumar, V. Kavitha, N. S. Saravanan, T. T. Nanjundamoorthi, and C. Chanakyan, "Tribological behavior on stir casted metal matrix composites of Al 6070 and TiC reinforcement with Taguchi S/N ratios," *Mater. Today Proc.*, vol. 77, pp. 455–461, 2023.
- [156] K. Dhoska *et al.*, "Manufacturing process of the aluminum alloy AA6063 for engineering applications," *Journal of Integrated Engineering and Applied Sciences*, vol. 1, no. 1, pp. 1–13, 2023.
- [157] Y. C. Sekhar, A. Gopichand, R. S. Sukumar, and N. P. Kumar, "Optimization of Surface Roughness of 6463 Aluminium Alloy and Brass Materials in CNC Milling Operation Using Taguchi's Design," *Optimization*, vol. 6, no. 2, 2017.
- [158] S. Gatea, H. Ou, and G. McCartney, "Deformation and fracture characteristics of Al6092/SiC/17.5 p metal matrix composite sheets due to heat treatments," *Mater. Charact.*, vol. 142, pp. 365–376, 2018.
- [159] J. Klemenc, S. Glodež, M. Steinacher, and F. Zupanič, "LCF behaviour of high strength aluminium alloys AA 6110A and AA 6086," *Int. J. Fatigue*, vol. 177, p. 107971, 2023.
- [160] O. R. Myhr and Ø. Grong, "Process modelling applied to 6082-T6 aluminium weldments—II. Applications of model," *Acta Metallurgica et Materialia*, vol. 39, no. 11, pp. 2703–2708, 1991.
- [161] L. Yin *et al.*, "High-temperature mechanical properties of constructional 6082-T6 aluminum alloy extrusion," in *Structures*, Elsevier, 2023, pp. 1244–1258.
- [162] J. T. Staley and D. J. Lege, "Advances in aluminium alloy products for structural applications in transportation," *Le Journal de Physique IV*, vol. 3, no. C7, pp. C7-179, 1993.
- [163] F. E. Saldanha, S. de A. Sousa, G. L. de Gouveia, M. Dwek, and J. E. Spinelli, "Evaluation of 6000 Al alloys for application in chassis of electric vehicles," *Materials Research*, vol. 25, p. e20220275, 2022.
- [164] X. Sheng *et al.*, "Microstructure and mechanical properties of friction stir welded joint of an aluminum alloy sheet 6005A-T4," *Metals (Basel)*, vol. 9, no. 11, p. 1152, 2019.
- [165] MatWeb, "Aluminium reference," 2024. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.matweb.com/reference/aluminium>
- [166] D. G. Eskin, "Decomposition of supersaturated solid solutions in Al–Cu–Mg–Si alloys," *J. Mater. Sci.*, vol. 38, no. 2, pp. 279–290, 2003.
- [167] R. Zang *et al.*, "Study on properties and precipitation behavior of 6000 series alloys with high Mg/Si ratios and Cu contents," *Mater. Charact.*, vol. 194, p. 112402, 2022.
- [168] Y. Zou, Q. Liu, Z. Jia, Y. Xing, L. Ding, and X. Wang, "The intergranular corrosion behavior of 6000-series alloys with different Mg/Si and Cu content," *Appl. Surf. Sci.*, vol. 405, pp. 489–496, 2017.
- [169] N. Maruyama, R. Uemori, N. Hashimoto, M. Saga, and M. Kikuchi, "Effect of silicon addition on the composition and structure of fine-scale precipitates in Al–Mg–Si alloys," *Scr. Mater.*, vol. 36, no. 1, 1997.
- [170] X. Xu *et al.*, "Effect of Mn content on microstructure and properties of 6000 series aluminum alloy," *Applied Physics A*, vol. 125, no. 8, p. 490, 2019.
- [171] Y. Han, K. Ma, L. Li, W. Chen, and H. Nagaumi, "Study on microstructure and mechanical properties of Al–Mg–Si–Cu alloy with high manganese content," *Mater. Des.*, vol. 39, pp. 418–424, 2012.
- [172] C. Blanc, Y. Roques, and G. Mankowski, "Application of phase shifting interferometric microscopy to studies of the behaviour of coarse intermetallic

- particles in 6056 aluminium alloy,” *Corros. Sci.*, vol. 40, no. 6, pp. 1019–1035, 1998.
- [173] M. H. Larsen, J. C. Walmsley, O. Lunder, R. H. Mathiesen, and K. Nisancioglu, “Intergranular corrosion of copper-containing AA6xxx AlMgSi aluminum alloys,” *J. Electrochem. Soc.*, vol. 155, no. 11, p. C550, 2008.
 - [174] M. Katoh, “Investigation of heat-affected zone cracking of GTA welds of Al-Mg-Si alloys using the Varestraint test,” *Weld. J.*, vol. 66, no. 12, pp. 360s–368s, 1987.
 - [175] S. Zajac, L.-O. Gullman, A. Johansson, and B. Bengtsson, “Hot ductility of some Al-Mg-Si alloys,” in *Materials Science Forum*, Trans Tech Publ, 1996, pp. 1193–1198.
 - [176] I. V Dubova, T. R. Gilmanshina, N. V Vasyunina, G. S. Sanacheva, and O. V Chupriy, “Electrochemical behavior of 6xxx series aluminium alloys in electrolytes containing chloride ions,” in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 042095.
 - [177] M. Miyazaki, K. Nishio, M. Katoh, S. Mukae, and H. W. Kerr, “Quantitative investigation of heat-affected zone cracking in aluminum Alloy 6061,” *Weld. J.*, vol. 69, no. 9, p. 362s, 1990.
 - [178] M. Aggerbeck *et al.*, “Appearance of anodised aluminium: Effect of alloy composition and prior surface finish,” *Surf. Coat. Technol.*, vol. 254, pp. 28–41, 2014.
 - [179] C. Huang and S. Kou, “Liquation mechanisms in multicomponent aluminum alloys during welding,” *Weld. J.*, vol. 81, no. 10, pp. 211S–222S, 2002.
 - [180] Mondolfo LF, *Aluminium alloys: structure and properties*. London: Butterworths, 1979.
 - [181] C. N. A. Jaafar, G. W. Lorimer, and N. C. Parson, “The effect of composition and temperature on the ageing response of some dilute 6xxx series alloys,” in *Materials science forum*, Trans Tech Publ, 2006, pp. 227–232.
 - [182] K. U. Bhat, D. B. Panemangalore, S. B. Kuruveri, M. John, and P. L. Menezes, “Surface modification of 6xxx Series aluminum alloys,” *Coatings*, vol. 12, no. 2, p. 180, 2022.
 - [183] J.-O. Andersson, T. Helander, L. Höglund, P. Shi, and B. Sundman, “ThermoCalc & DICTRA, computational tools for materials science,” *Calphad*, vol. 26, no. 2, pp. 273–312, 2002.
 - [184] E FORU, “Aluminum Metal and Alloys – Product Catalog,” E.Foru. Accessed: Nov. 19, 2025. [Online]. Available: <https://www.eforu-rareearth.com/aluminum-metal-and-alloys/7028-aluminum-alloy%2Caluminum-7028-alloy.html>
 - [185] Y. A. Fageehi, R. Saminathan, G. Venugopal, J. Valder, H. Kumar, and K. S. Ravishankar, “Effect of thermal and surface chemical treatment on the cyclic oxidation behavior of 7039 aluminum alloy used in aerospace armor applications,” *Mater. Today Proc.*, vol. 42, pp. 343–349, 2021.
 - [186] N.-H. Hoang, R. Porcaro, M. Langseth, and A.-G. Hanssen, “Self-piercing riveting connections using aluminium rivets,” *Int. J. Solids Struct.*, vol. 47, no. 3–4, pp. 427–439, 2010.
 - [187] N.-H. Hoang, O. S. Hopperstad, M. Langseth, and I. Westermann, “Failure of aluminium self-piercing rivets: An experimental and numerical study,” *Mater. Des.*, vol. 49, pp. 323–335, 2013.
 - [188] R. Bakhshi, M. H. Farshidi, and S. A. Sajjadi, “Strengthening of aluminium alloy 7005 through imposition of severe plastic deformation supplemented by

- different ageing treatments,” *Transactions of Nonferrous Metals Society of China*, vol. 31, no. 10, pp. 2909–2921, 2021.
- [189] F. D. Zhang *et al.*, “Corrosion behaviour of laser-cleaned AA7024 aluminium alloy,” *Appl. Surf. Sci.*, vol. 435, pp. 452–461, 2018.
- [190] M. Tajbakhsh, O. Yaghobizadeh, and M. Farhadi Nia, “Investigation of the physical and mechanical properties of Ni–P and Ni–P–PTFE nanocomposite coatings deposited on aluminum alloy 7023,” *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 233, no. 1, pp. 94–103, 2019.
- [191] D. Davoodi, S. A. H. Emami, and A. Saidi, “Synthesis and Characterization of Al 7014 Nanostructure by Mechanical Alloying,” *Journal of Advanced Materials and Technologies*, vol. 3, no. 3, pp. 79–87, 2014.
- [192] K. Narender, A. S. M. Rao, K. G. K. Rao, and N. G. Krishna, “Temperature dependence of density and thermal expansion of wrought aluminum alloys 7041, 7075 and 7095 by gamma ray attenuation method,” *Journal of Modern Physics*, vol. 4, no. 3, pp. 331–336, 2013.
- [193] D. Childree, “High-strength aluminum automotive alloy,” *Advanced materials & processes*, vol. 154, no. 3, p. 27, 1998.
- [194] Z. MA, Y. DENG, J. LIU, S. LIU, and H. LIU, “Effect of quenching rate on stress corrosion cracking susceptibility of 7136 aluminum alloy,” *Acta Metall Sin*, vol. 58, no. 9, pp. 1118–1128, 2021.
- [195] T. Oppenheim *et al.*, “On the correlation of mechanical and physical properties of 6061-T6 and 7249-T76 aluminum alloys,” *Eng. Fail. Anal.*, vol. 14, no. 1, pp. 218–225, 2007.
- [196] AZoM, “Applications of Aluminium Alloys,” AZoM. Accessed: Jul. 29, 2025. [Online]. Available: <https://www.azom.com/article.aspx?ArticleID=8719>
- [197] J. Boselli, D. J. Chakrabarti, and R. T. Shuey, “Aerospace applications: metallurgical insights into the improved performance of aluminum alloy 7085 thick products,” *Al Alloys*, vol. 1, pp. 202–208, 2008.
- [198] D. Tian, R. Wang, and J. Zheng, “Research on the mechanical properties and hot deformation behaviors of spray-deposited 7034 Al alloy processed by forward extrusion,” *J. Mater. Eng. Perform.*, vol. 31, no. 1, pp. 37–52, 2022.
- [199] F. C. Campbell, “Aluminum,” 2013.
- [200] J. C. S. Fernandes *et al.*, “SVET and SIET Studies on 7081-T73511 Aluminium Alloy”.
- [201] A. A. Alhamdany, A. Y. Khenyab, Q. K. Mohammed, and H. J. M. Alalkawi, “Development mechanical and fatigue properties of AA7001 after combined SP with deep cryogenic treatment and UIP with deep cryogenic treatment,” *East-ern-European Journal of Enterprise Technologies*, vol. 5, no. 1, p. 113, 2021.
- [202] J. Z. Liu *et al.*, “Revisiting the precipitation sequence in Al–Zn–Mg-based alloys by high-resolution transmission electron microscopy,” *Scr. Mater.*, vol. 63, no. 11, pp. 1061–1064, 2010.
- [203] T. Marlaud, A. Deschamps, F. Bley, W. Lefebvre, and B. Baroux, “Influence of alloy composition and heat treatment on precipitate composition in Al–Zn–Mg–Cu alloys,” *Acta Mater.*, vol. 58, no. 1, pp. 248–260, 2010.
- [204] W. X. Shu *et al.*, “Tailored Mg and Cu contents affecting the microstructures and mechanical properties of high-strength Al–Zn–Mg–Cu alloys,” *Materials Science and Engineering: A*, vol. 657, pp. 269–283, 2016.

- [205] Y. Deng, Z. Yin, K. Zhao, J. Duan, J. Hu, and Z. He, “Effects of Sc and Zr microalloying additions and aging time at 120 C on the corrosion behaviour of an Al–Zn–Mg alloy,” *Corros. Sci.*, vol. 65, pp. 288–298, 2012.
- [206] H. T. Kim, S. W. Nam, and S. H. Hwang, “Study on the solidification cracking behaviour of high strength aluminum alloy welds: Effects of alloying elements and solidification behaviours,” *J. Mater. Sci.*, vol. 31, no. 11, pp. 2859–2864, 1996.
- [207] S. Chen, J. Li, G. Hu, K. Chen, and L. Huang, “Effect of Zn/Mg ratios on SCC, electrochemical corrosion properties and microstructure of Al-Zn-Mg alloy,” *J. Alloys Compd.*, vol. 757, pp. 259–264, 2018.
- [208] L. Wen *et al.*, “Microstructure and stress corrosion cracking resistance of Al-6.5 Zn-2Cu-1.5 Mg-0.05 Ti alloy modified by Cr addition,” *Mater. Charact.*, vol. 183, p. 111621, 2022.
- [209] T. E. Borchers, D. P. McAllister, and W. Zhang, “Macroscopic segregation and stress corrosion cracking in 7xxx series aluminum alloy arc welds,” *Metallurgical and Materials Transactions A*, vol. 46, no. 5, pp. 1827–1833, 2015.
- [210] J. A. Wagner and R. N. Shenoy, “The effect of copper, chromium, and zirconium on the microstructure and mechanical properties of Al-Zn-Mg-Cu alloys,” *Metallurgical transactions A*, vol. 22, no. 11, pp. 2809–2818, 1991.
- [211] D. S. Park and S. W. Nam, “Effects of manganese dispersoid on the mechanical properties in Al-Zn-Mg alloys,” *J. Mater. Sci.*, vol. 30, no. 5, pp. 1313–1320, 1995.
- [212] A. K. Jha, G. N. Shiresha, K. Sreekumar, M. C. Mittal, and K. N. Ninan, “Stress corrosion cracking in aluminium alloy AFNOR 7020-T6 water tank adaptor for liquid propulsion system,” *Eng. Fail. Anal.*, vol. 15, no. 6, pp. 787–795, 2008.
- [213] Y. Liao, X. Han, M. Zeng, and M. Jin, “Influence of Cu on microstructure and tensile properties of 7XXX series aluminum alloy,” *Mater. Des.*, vol. 66, pp. 581–586, 2015.
- [214] X. Z. Wu, D. H. Xiao, Z. M. Zhu, X. X. Li, and K. H. Chen, “Effects of Cu/Mg ratio on microstructure and properties of AA7085 alloys,” *Transactions of Non-ferrous Metals Society of China*, vol. 24, no. 7, pp. 2054–2060, 2014.
- [215] D. Zhang *et al.*, “Liquation Cracking Tendency of Novel Al-Mg-Zn Alloys with a Zn/Mg Ratio below 1.0 during Fusion Welding,” *Metals (Basel)*, vol. 10, no. 2, p. 222, 2020.
- [216] Z. Cvijović, M. Vratnica, and I. Cvijović-Alagić, “The influences of multiscale-sized second-phase particles on fracture behaviour of overaged 7000 alloys,” *Procedia Eng.*, vol. 1, no. 1, pp. 35–38, 2009.
- [217] Constellium, “Aerospace plates,” Constellium. Accessed: Jul. 29, 2025. [Online]. Available: <https://www.constellium.com/products/aerospace-plates>
- [218] W. Wolf *et al.*, “Enhancement of mechanical properties of aluminum and 2124 aluminum alloy by the addition of quasicrystalline phases,” *Materials Research*, vol. 19, no. Suppl 1, pp. 74–79, 2016.
- [219] D. J. Chellman, “Development of powder metallurgy 2XXX series Al alloys for high temperature aircraft structural applications,” 1984.
- [220] G. V. Ramana, B. Yelamasetti, and T. V. Vardhan, “Study on weldability and effect of post heat treatment on mechanical and metallurgical properties of dissimilar AA 2025, AA 5083 and AA7075 GTAW weld joints,” *Mater. Today Proc.*, vol. 46, pp. 878–882, 2021.
- [221] W. Yuan, M. Wan, and X. Wu, “Prediction of forming limit curves for 2021 aluminum alloy,” *Procedia Eng.*, vol. 207, pp. 544–549, 2017.

- [222] Z. Silvayeh, M. Brillinger, and J. Domitner, "Deformation behavior of aluminum alloy rivets for aerospace applications," *Journal of Materials Research and Technology*, vol. 33, pp. 3482–3491, 2024.
- [223] N. Çömez, "Wear behavior and corrosion properties of Age-hardened AA2010 aluminum alloy," *Revista de Metalurgia*, vol. 57, no. 3, 2021.
- [224] Aluminum Company of America, "ALUMINUM 2021: Cryogenic sheet and plate alloy available to purchase," *Alloy Digest*, vol. 18, no. 9, p. Al-190, Sep. 1969.
- [225] M. Suresh, M. Mohanraju, M. S. Senthilkumar, T. Jayakumar, and M. E. H. M. S. Arafath, "Wear Behavior of Aluminium Alloy 2218 Fly ash-Talc Hybrid Metal Matrix Composites," *International Journal of Innovative Research in Science and Engineering*, vol. 8, no. 1, pp. 15–22, 2022.
- [226] Oguocha IN, "Characterization of aluminium alloy 2618 and its composites containing alumina particles," PhD thesis, University of Saskatchewan, 1999.
- [227] S. Li *et al.*, "Development and applications of aluminum alloys for aerospace industry," *Journal of materials research and technology*, vol. 27, pp. 944–983, 2023.
- [228] Smith Metal, "Smith Metal," <https://www.smithmetal.com/>. Accessed: Jul. 29, 2025. [Online]. Available: <https://www.smithmetal.com/>
- [229] OEM Aluminium Profile, "2040 Aluminium Extrusion Profile," OEM Aluminium Profile. Accessed: Jul. 29, 2025. [Online]. Available: <https://my.oemaluminiumprofile.com/aluminium-profile/aluminium-extrusion-profile/2040-aluminium-extrusion.html>
- [230] R. Kant, K. Adhikari, A. Mehmood, R. Chaudhary, and M. Kashif, "Mechanical and microstructure characterization of al 2029 by friction stir welding," *structure*, vol. 4, p. 5.
- [231] C. S. Jawalkar and S. Kant, "A review on use of aluminium alloys in aircraft components," *i-manager's Journal on Material Science*, vol. 3, no. 3, p. 33, 2015.
- [232] M. Gazizov and R. Kaibyshev, "Precipitation structure and strengthening mechanisms in an Al-Cu-Mg-Ag alloy," *Materials Science and Engineering: A*, vol. 702, pp. 29–40, 2017.
- [233] I. N. A. Oguocha and S. Yannacopoulos, "Precipitation and dissolution kinetics in Al-Cu-Mg-Fe-Ni alloy 2618 and Al-alumina particle metal matrix composite," *Materials Science and Engineering: A*, vol. 231, no. 1–2, pp. 25–33, 1997.
- [234] D. H. Xiao, J. N. Wang, D. Y. Ding, and S. P. Chen, "Effect of Cu content on the mechanical properties of an Al-Cu-Mg-Ag alloy," *J. Alloys Compd.*, vol. 343, no. 1–2, pp. 77–81, 2002.
- [235] J. Chen *et al.*, "Effect of Mg content on microstructure, mechanical properties and intergranular corrosion properties of Al-Cu-Mg-Ag alloys," *Mater. Today Commun.*, vol. 34, p. 105363, 2023.
- [236] J. Zhang *et al.*, "Study on the effect of Mg and Ag trace elements on the properties of Al-Cu-Mn alloy," in *Journal of Physics: Conference Series*, IOP Publishing, 2024, p. 012011.
- [237] D. Bakavos, P. B. Prangnell, B. Bes, and F. Eberl, "The effect of silver on microstructural evolution in two 2xxx series Al-alloys with a high Cu: Mg ratio during ageing to a T8 temper," *Materials Science and Engineering: A*, vol. 491, no. 1–2, pp. 214–223, 2008.
- [238] F. Wang *et al.*, "Effect of Si addition on the microstructure and mechanical properties of Al-Cu-Mg alloy," *J. Alloys Compd.*, vol. 585, pp. 474–478, 2014.

- [239] M. Natan and R. A. Chihoski, "Relationship between microstructure, hardness and electrical conductivity of 2219 aluminium," *J. Mater. Sci.*, vol. 18, no. 11, pp. 3288–3298, 1983.
- [240] M. Kok, "Production and mechanical properties of Al₂O₃ particle-reinforced 2024 aluminium alloy composites," *J. Mater. Process. Technol.*, vol. 161, no. 3, pp. 381–387, 2005.
- [241] I. S. Zuiko, M. R. Gazizov, and R. O. Kaibyshev, "Effect of thermomechanical treatment on the microstructure, phase composition, and mechanical properties of Al–Cu–Mn–Mg–Zr alloy," *The Physics of Metals and Metallography*, vol. 117, no. 9, pp. 906–919, 2016.
- [242] K. R. Pai, V. Vijayan, and K. N. Prabhu, "Investigation of the effect of process parameters on porosity, microstructure and mechanical properties of Al–5 Mg alloy test samples fabricated by wire arc additive manufacturing," *Progress in Additive Manufacturing*, pp. 1–14, 2024.
- [243] Alibaba Group, "Alibaba.com: Global trade starts here," Alibaba. Accessed: Jul. 29, 2025. [Online]. Available: <https://www.alibaba.com/>
- [244] J. K. McBride, R. E. Sanders Jr, and H. G. Reavis, "The development of 5XXX aluminum alloys for RCS applications," *JOM*, vol. 48, no. 6, pp. 18–21, 1996.
- [245] M. Lipińska, P. Bazarnik, and M. Lewandowska, "The influence of severe plastic deformation processes on electrical conductivity of commercially pure aluminium and 5483 aluminium alloy," *Archives of Civil and Mechanical Engineering*, vol. 16, no. 4, pp. 717–723, 2016.
- [246] A. V. Jebaraj, K. V. V Aditya, T. S. Kumar, L. Ajaykumar, and C. R. Deepak, "Mechanical and corrosion behaviour of aluminum alloy 5083 and its weldment for marine applications," *Mater. Today Proc.*, vol. 22, pp. 1470–1478, 2020.
- [247] A. V Chumaevskii *et al.*, "Instabilities of friction stir processing of aluminum-magnesium alloy with a square section pin tool," in *AIP Conference Proceedings*, AIP Publishing LLC, 2022, p. 020046.
- [248] T. E. Abioye, H. Zuhailawati, S. Aizad, and A. S. Anasyida, "Geometrical, microstructural and mechanical characterization of pulse laser welded thin sheet 5052-H32 aluminium alloy for aerospace applications," *Trans Nonferrous Met Soc China*, vol. 29, pp. 667–679, 2019.
- [249] P. C. Theja, R. Suresh, and M. C. S. Reddy, "Experimental investigation and mechanical characterization of Al 5051 with nickel/titanium powder metal matrix composites using stir casting," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 8, no. 1, p. 77, 2025.
- [250] O. Engler, K. Kuhnke, and J. Hasenclever, "Development of intermetallic particles during solidification and homogenization of two AA 5xxx series Al-Mg alloys with different Mg contents," *J. Alloys Compd.*, vol. 728, pp. 669–681, 2017.
- [251] H. A. Razazi, M. Paidar, and O. O. Ojo, "Effect of Mn and Cr on structure and mechanical properties of Al-10% Mg-0.1% Ti alloy," *Vacuum*, vol. 155, pp. 619–630, 2018.
- [252] O. Engler and S. Miller-Jupp, "Control of second-phase particles in the Al-Mg-Mn alloy AA 5083," *J. Alloys Compd.*, vol. 689, pp. 998–1010, 2016.
- [253] Q. Li, M. Li, G. Lu, S. Guan, E. Zhang, and C. Xu, "Effect of trace elements on the crystallization temperature interval and properties of 5xxx series aluminum alloys," *Metals (Basel)*, vol. 10, no. 4, p. 483, 2020.
- [254] X. Liu, J. Xu, F. Zhao, X. Liu, and Y. Tian, "Effect of homogenization on second phases and mechanical properties of AA 5052 aluminum alloy tube billets

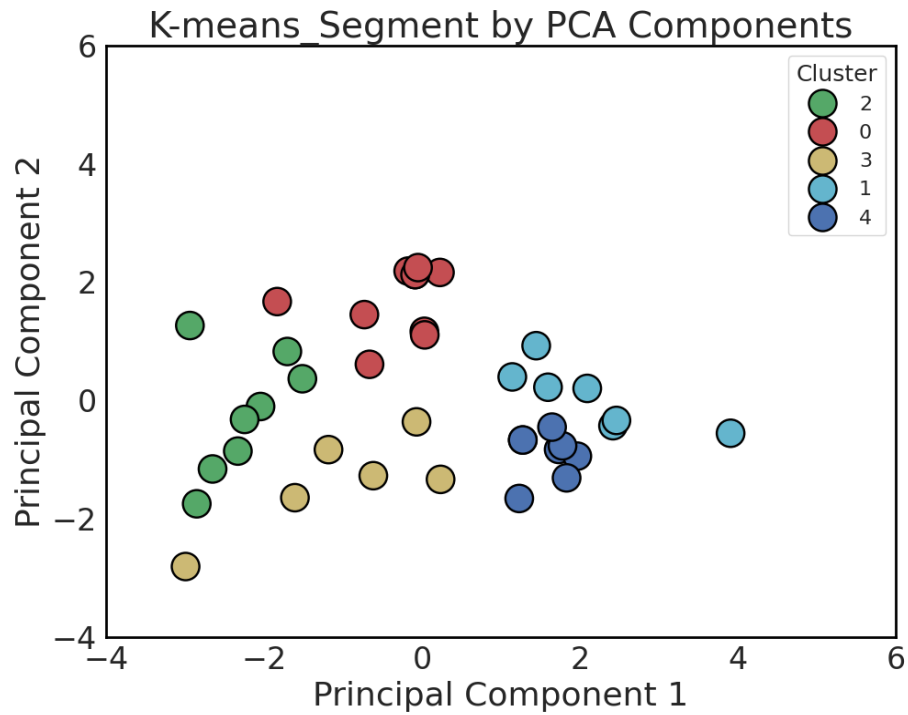
- fabricated by HCCM vertical continuous casting,” *J. Alloys Compd.*, vol. 901, p. 163645, 2022.
- [255] A. Amerioon, M. Emamy, and G. Ashuri, “Investigation the effect of Al-5Ti-1B grain refiner and T6 heat treatment on tensile properties of Al-8% Mg,” *Procedia Materials Science*, vol. 11, pp. 32–37, 2015.
 - [256] D. Su, J. Zhang, and B. Wang, “The microstructure and weldability in welded joints for AA 5356 aluminum alloy after adding modified trace amounts of Sc and Zr,” *J. Manuf. Process.*, vol. 57, pp. 488–498, 2020.
 - [257] Z.-B. He, Y.-Y. Peng, Z.-M. Yin, and X.-F. Lei, “Comparison of FSW and TIG welded joints in Al-Mg-Mn-Sc-Zr alloy plates,” *Transactions of Nonferrous Metals Society of China*, vol. 21, no. 8, pp. 1685–1691, 2011.
 - [258] H. E. Vatne, O. Engler, and E. Nes, “Influence of particles on recrystallisation textures and microstructures of aluminium alloy 3103,” *Materials science and technology*, vol. 13, no. 2, pp. 93–102, 1997.
 - [259] J. T. B. Gundersen, A. Aytac, S. Ono, J. H. Nordlien, and K. Nişancıoğlu, “Effect of trace elements on electrochemical properties and corrosion of aluminium alloy AA3102,” *Corros. Sci.*, vol. 46, no. 2, pp. 265–283, 2004.
 - [260] A. L. T. Martins, A. A. Couto, N. B. de Lima, and G. F. C. Almeida, “Crystallographic texture evolution of aluminum alloy 3104 in the drawn and wall ironing (DWI) process,” *Materials Research*, vol. 22, p. e20190037, 2020.
 - [261] J. Zander, R. Sandström, and L. Vitos, “Modelling mechanical properties for non-hardenable aluminium alloys,” *Comput. Mater. Sci.*, vol. 41, no. 1, pp. 86–95, 2007.
 - [262] H. Mraied and W. Cai, “The effects of Mn concentration on the tribocorrosion resistance of Al–Mn alloys,” *Wear*, vol. 380, pp. 191–202, 2017.
 - [263] L. I. Zhen and Z. Zhang, “Effect of magnesium on dispersoid strengthening of Al–Mn–Mg–Si (3xxx) alloys,” *Transactions of Nonferrous Metals Society of China*, vol. 26, no. 11, pp. 2793–2799, 2016.
 - [264] W. B. Hutchinson and H.-E. Ekström, “Control of annealing texture and earing in non-hardenable aluminium alloys,” *Materials science and technology*, vol. 6, no. 11, pp. 1103–1112, 1990.
 - [265] Das SK, *Aluminum Alloys for Packaging III*, vol. 3. Warrendale, PA : TMS, 1998.
 - [266] S. Xu *et al.*, “Simultaneously improving mechanical properties and electrical conductivity of Al-1.9 Mn alloys with different Mg and Si addition,” *J. Alloys Compd.*, vol. 999, p. 175072, 2024.
 - [267] L. Lodgaard and N. Ryum, “Precipitation of dispersoids containing Mn and/or Cr in Al–Mg–Si alloys,” *Materials Science and Engineering: A*, vol. 283, no. 1–2, pp. 144–152, 2000.
 - [268] C. Cao *et al.*, “Effects of Cu addition on the microstructure and properties of the Al–Mn–Fe–Si alloy,” *J. Alloys Compd.*, vol. 834, p. 155175, 2020.
 - [269] A. M. F. Muggerud, E. A. Mørtzell, Y. Li, and R. Holmestad, “Dispersoid strengthening in AA3xxx alloys with varying Mn and Si content during annealing at low temperatures,” *Materials Science and Engineering: A*, vol. 567, pp. 21–28, 2013.
 - [270] K. Huang, N. Wang, Y. Li, and K. Marthinsen, “The influence of microchemistry on the softening behaviour of two cold-rolled Al–Mn–Fe–Si alloys,” *Materials Science and Engineering: A*, vol. 601, pp. 86–96, 2014.
 - [271] K. Huang, O. Engler, Y. J. Li, and K. Marthinsen, “Evolution in microstructure and properties during non-isothermal annealing of a cold-rolled Al–Mn–Fe–Si

- alloy with different microchemistry states,” *Materials Science and Engineering: A*, vol. 628, pp. 216–229, 2015.
- [272] Z. Li, J. Qin, B. Zhang, and H. Nagaumi, “Effect of Cr Addition on Microstructure and Mechanical Properties at Elevated Temperature of Al–Mn–Mg–Si Alloys,” *Mater. Trans.*, vol. 61, no. 11, pp. 2095–2100, 2020.
 - [273] F. Liu, Q. Wang, W. Yang, J. Feng, and Y. Tu, “Simultaneous Improvement of the Mechanical and Corrosion Properties of a 3102 Aluminum Alloy via Ti Addition,” *J. Mater. Eng. Perform.*, pp. 1–9, 2024.
 - [274] Z. Li, Z. Zhang, and X.-G. Chen, “Improvement in the mechanical properties and creep resistance of Al–Mn–Mg 3004 alloy with Sc and Zr addition,” *Materials Science and Engineering: A*, vol. 729, pp. 196–207, 2018.
 - [275] P. Z. Zhao, Y. F. Pan, S. J. Shi, T. Jiang, X. X. Wu, and Y. J. Feng, “Effect of Cu addition on laser welding performance of Al–Mn–Mg alloy sheets,” *Materials Research Innovations*, vol. 18, no. sup4, pp. S4-239, 2014.
 - [276] H. Hirasawa, “Precipitation process of Al□ Mn and Al□ Cr supersaturated solid solution in presence of age hardening phases,” *Scripta Metallurgica*, vol. 9, no. 9, pp. 955–958, 1975.
 - [277] G. E. Totten, M. Tiryakioglu, and O. Kessler, *Encyclopedia of aluminum and its alloys*. CRC Press, 2018.
 - [278] R. E. Sanders Jr, P. A. Hollinshead, and E. A. Simielli, “Industrial development of non-heat treatable aluminum alloys,” in *Materials Forum*, 2004, pp. 53–64.
 - [279] UACJ Corporation, “Types and applications of aluminum alloys for vehicles,” Aluminum Automobile Technology (UACJ Aluminum Automobile Technology). Accessed: Sep. 09, 2025. [Online]. Available: https://uacj-automobile.com/types_and_applications.html
 - [280] J. A. Van Der Hoeven, L. Zhuang, B. Schepers, P. De Smet, and J. P. Baekelandt, “A new 5xxx series alloy developed for automotive applications,” SAE Technical Paper, 2002.
 - [281] W. S. Miller *et al.*, “Recent development in aluminium alloys for the automotive industry,” *Materials Science and Engineering: A*, vol. 280, no. 1, pp. 37–49, 2000.
 - [282] G. Ziobro, M. Richert, and M. Wiewióra, “Influence of corrosion on mechanical properties and microstructure of 3xxx, 5xxx, and 6xxx series aluminum alloys,” *Metallurgy and Foundry Engineering*, vol. 43, no. 4, 2017.
 - [283] F. Tolun, “Use of aluminum alloys in automotive industry,” in *Congress Proceedings Series*, 2019, p. 63.
 - [284] M. Tisza and Z. Lukács, “High strength aluminum alloys in car manufacturing,” in *IOP conference series: materials science and engineering*, IOP Publishing, 2018, p. 012033.
 - [285] O. Engler and J. Hirsch, “Texture control by thermomechanical processing of AA6xxx Al–Mg–Si sheet alloys for automotive applications—a review,” *Materials Science and Engineering: A*, vol. 336, no. 1–2, pp. 249–262, 2002.
 - [286] N. Khademian and Y. Peimaei, “Lightweight materials (LWM) in transportation especially application of aluminum in light weight automobiles (LWA),” in *International conference on interdisciplinary studies in nanotechnology*, 2020, pp. 1–22.
 - [287] J. Hirsch, “Automotive trends in aluminium-The European perspective,” in *Materials Forum*, 2004, pp. 15–23.

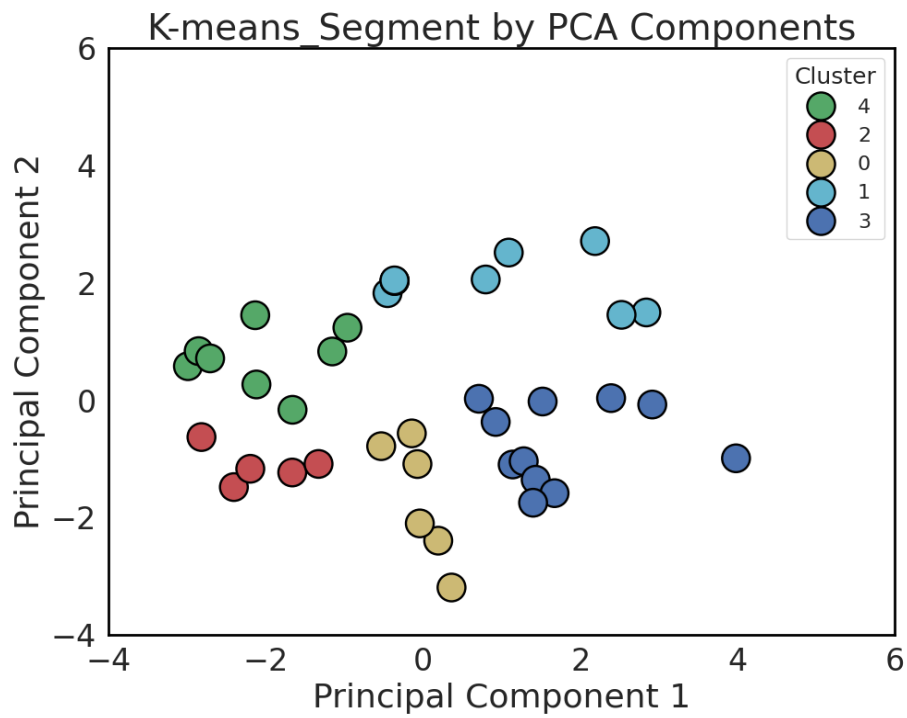
- [288] Chalco Aluminum, “3004 3104 H19 aluminum coil strip for can body,” Chalco Aluminum. Accessed: Sep. 09, 2025. [Online]. Available: <https://www.chalco-aluminum.com/knowledge/can-body-coil-strip/>
- [289] J. R. Davis, *Corrosion of aluminum and aluminum alloys*. Asm International, 1999.
- [290] R. Gitter, “Design of aluminium structures: selection of structural alloys,” *EU-ROCODES-Background and Applications, Brussels, Belgium*, 2008.
- [291] S. Majumdar, A. Sinha, A. Das, P. Datta, and D. Nag, “An Insight View of evolution of advanced aluminum alloy for aerospace and automotive industry: current status and future prospects,” *Journal of The Institution of Engineers (India): Series D*, pp. 1–18, 2024.
- [292] R. Gitter, “Design of aluminium structures: selection of structural alloys,” *EU-ROCODES-Background and Applications, Brussels, Belgium*, 2008.
- [293] O. Engler and J. Hirsch, “Texture control by thermomechanical processing of AA6xxx Al–Mg–Si sheet alloys for automotive applications—a review,” *Materials Science and Engineering: A*, vol. 336, no. 1–2, pp. 249–262, 2002.
- [294] S. Majumdar, A. Sinha, A. Das, P. Datta, and D. Nag, “An Insight View of evolution of advanced aluminum alloy for aerospace and automotive industry: current status and future prospects,” *Journal of The Institution of Engineers (India): Series D*, pp. 1–18, 2024.
- [295] J. Hirsch, “Automotive trends in aluminium-The European perspective,” in *Materials Forum*, 2004, pp. 15–23.
- [296] UACJ Aluminum Automobile Technology, “UACJ Aluminum Automobile Technology.” Accessed: Aug. 26, 2025. [Online]. Available: https://uacj-automobile.com/types_and_applications.html
- [297] J. Kissell, *Take Control of the Mac Command Line with Terminal*. alt concepts, 2025.
- [298] M. Khorasani, M. Abdou, and J. H. Fernández, “Web application development with streamlit,” *Software Development*, vol. 498, p. 507, 2022.
- [299] P.-E. Danielsson, “Euclidean distance mapping,” *Computer Graphics and image processing*, vol. 14, no. 3, pp. 227–248, 1980.

CHAPTER 11: APPENDIX

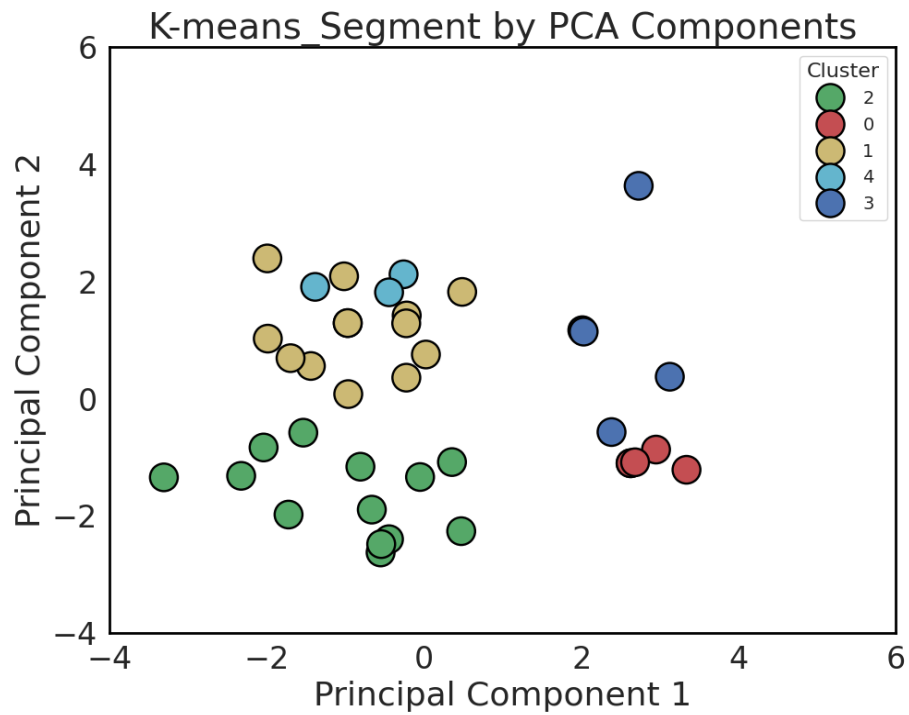
11.1 Graphs for Processing Conditions of 6xxx Series Aluminium Alloys



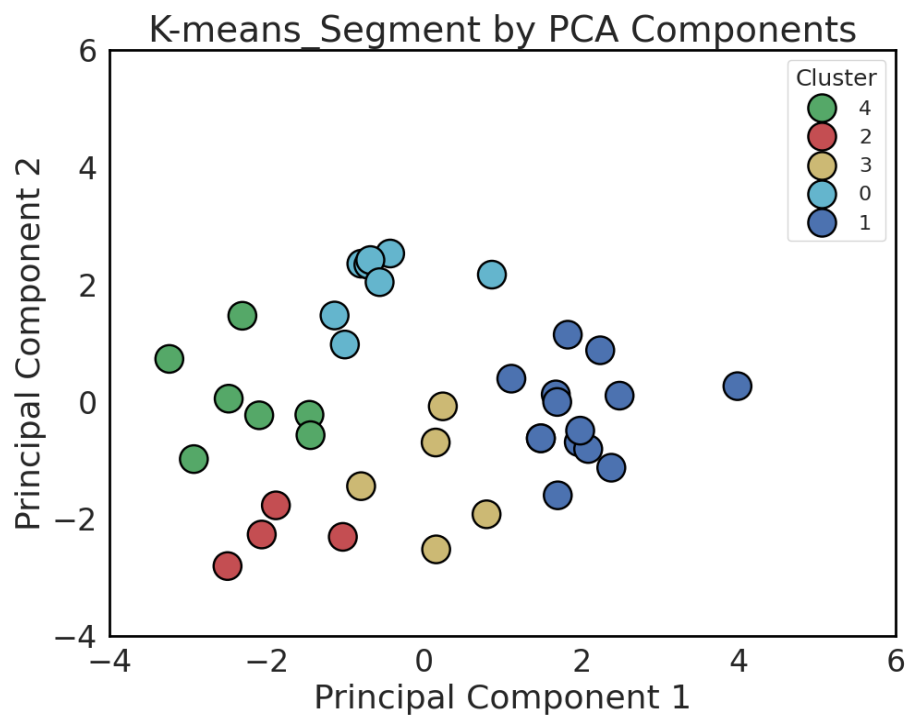
(a) Forged



(b) Rolled



(c) Sheet



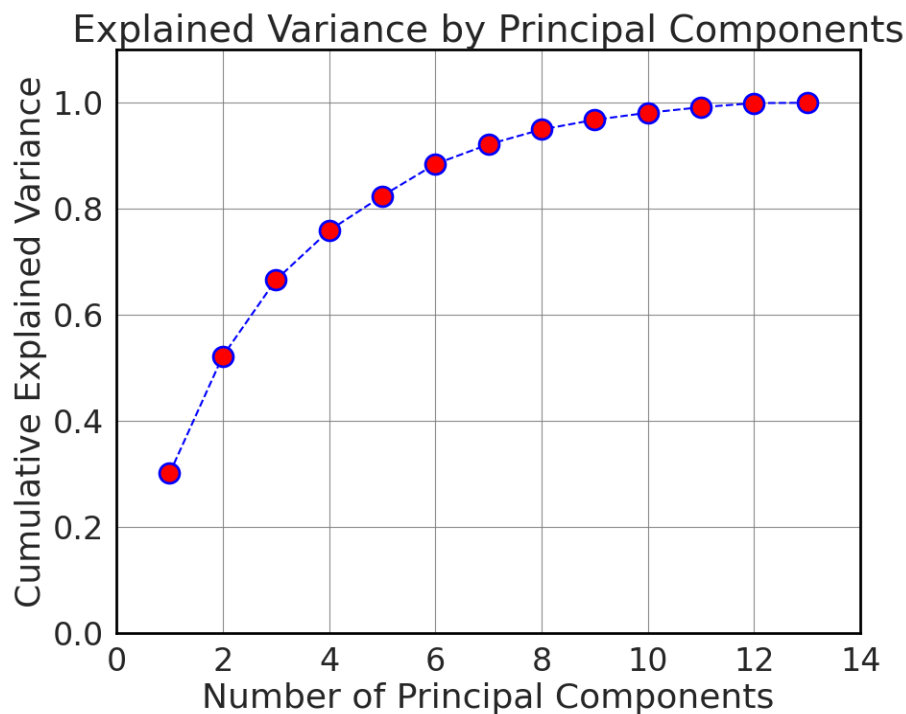
(d) Plate

Figure 11.1. Processing conditions 6xxx series: (a) Forged, (b) Rolled, (c) Sheet, and (d) Plate

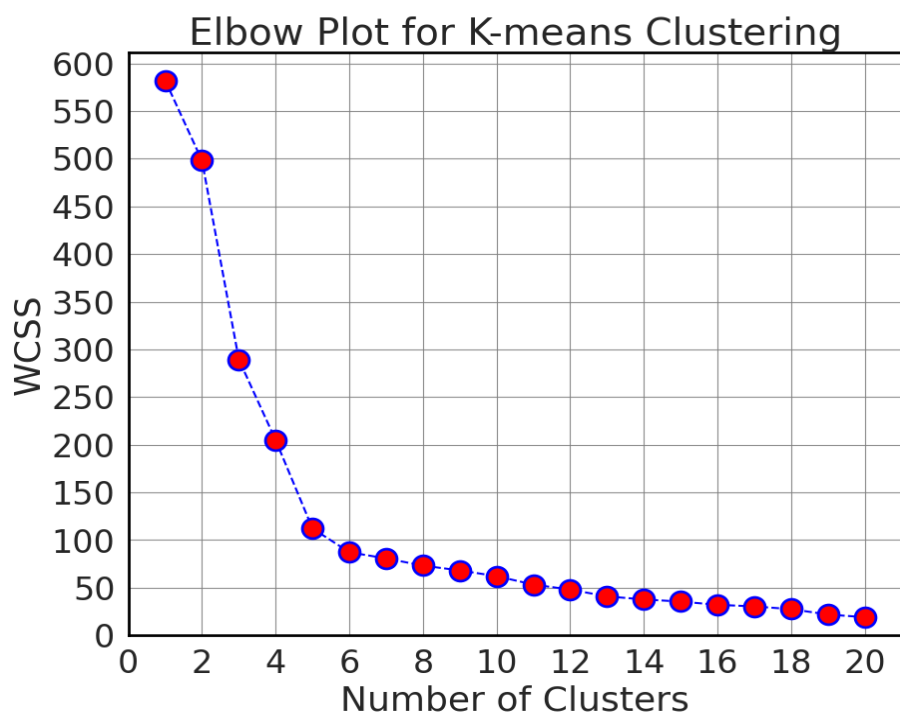
11.2 Supplementary Material: Clustering and Sub-Clustering Analysis for the Case Study Chapter

11.2.1 Clustering and Sub-clustering of 2xxx Series Alloys

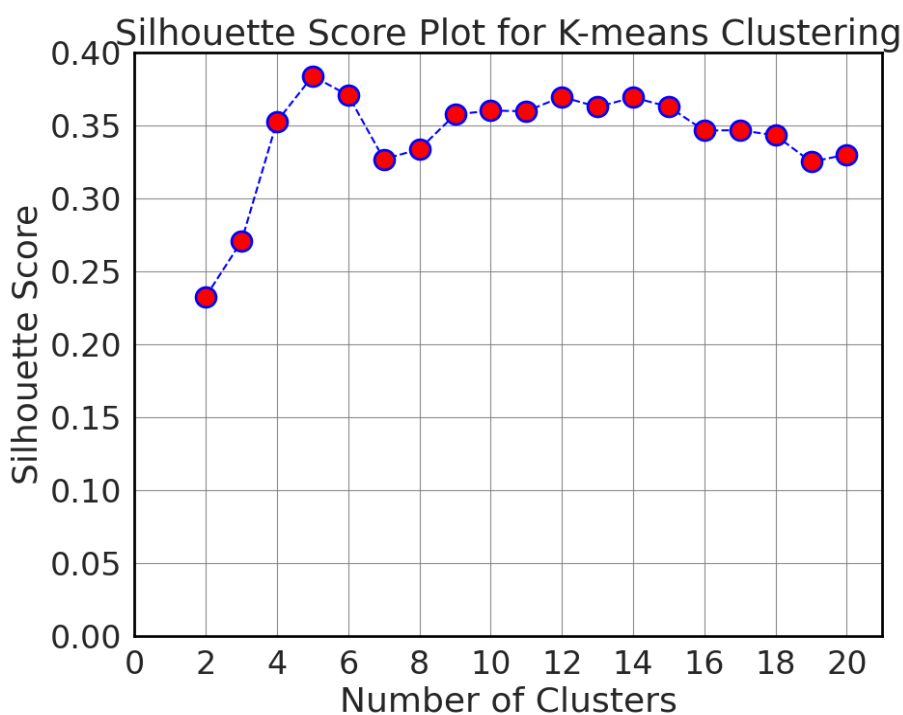
A combined approach of Principal Component Analysis (PCA) with K-means clustering, was employed to investigate the clustering behaviour of 2xxx series aluminium alloys, based on the dataset presented in **Section 7.1**. Initially, a PCA variance plot was employed to facilitate the dimensionality reduction of the dataset. Through this process, the original feature set was reduced from 11 to 4 principal components, retaining approximately 80% of the total variance within the dataset (Figure 11.2(a)). This dimensionality reduction not only enhanced the efficiency of subsequent analyses but also preserved the essential characteristics of the alloys. Furthermore, it improved the visualisation of clusters, enabling a clearer and more rigorous differentiation between the various alloy groupings.



(a)



(b)



(c)

Figure 11.2. Plots to detect the number of clusters in 2xxx-series alloys (a) PCA variance plot, (b) an Elbow plot and, (c) Silhouette score plot

After dimensionality reduction, optimal number of clusters was determined using both the elbow method and silhouette analysis. The elbow plot exhibited a pronounced inflection point at $K = 5$ and $K = 6$, as shown in the curve (Figure 11.2(b)), while the silhouette analysis indicated the highest scores at $K = 5$ and $K = 6$ (Figure 11.2(c)). Considering both criteria, $K = 6$ was selected as the optimal number of clusters, indicating that the dataset is most appropriately represented by six distinct clusters, as illustrated in Figure 11.3.

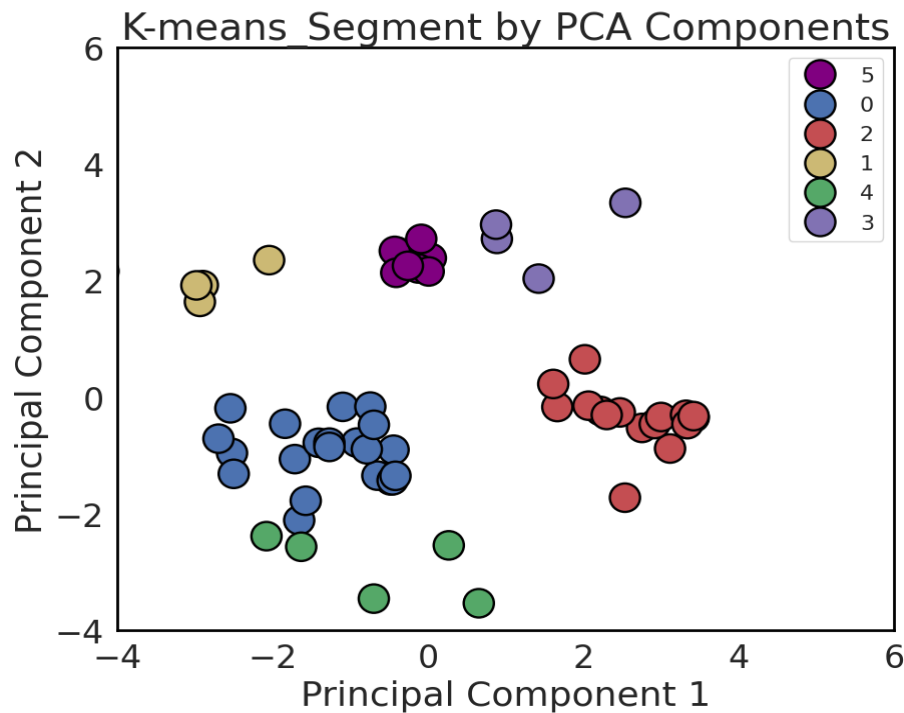


Figure 11.3. Clustering analysis of 2xxx series alloys

The initial K-means clustering yielded six distinct clusters, each exhibiting similar compositions and property ranges within the cluster, as summarised in Table 11.1. Cluster 4, for instance, predominantly consisted of high-strength alloys with elevated creep and rupture strength, characterised by relatively high Cu, Mg, and Mn contents. These alloys demonstrated superior mechanical performance, together with notable resistance to degradation at elevated temperatures. In contrast, Cluster 3, with comparatively low Mg and Mn but high Cu contents, exhibited moderate yield and creep strengths.

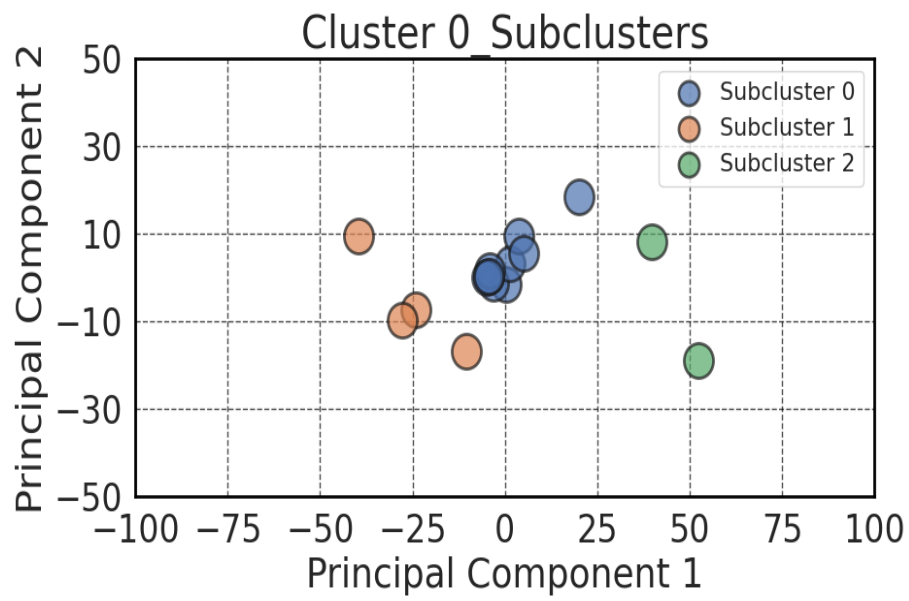
Table 11.1. Composition and properties range in 2xxx series clusters

Property/Composition	Cluster 0 (High)	Cluster 1 (Moderate)	Cluster 2 (Medium)	Cluster 3 (Moderate)	Cluster 4 (High)	Cluster 5 (Medium)
Alloying Elements						

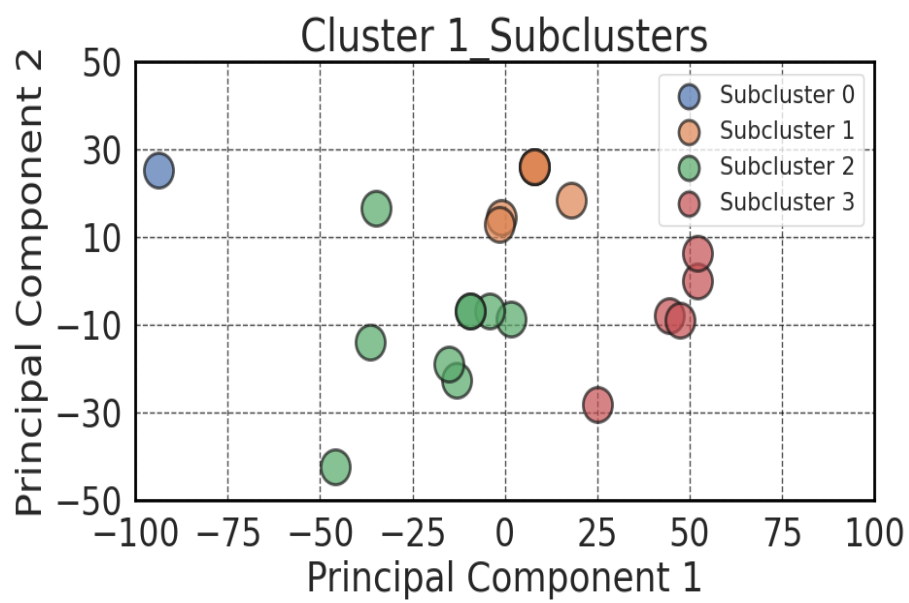
Cu (%)	3.8 – 6.3	0.9 – 4.55	2 – 4	4.75 – 6.3	3.6 – 5.1	4.95 – 6.3
Mn (%)	0 – 1.3	0.2 – 1.2	0 – 0.5	0 – 0.4	0.4 – 0.8	0.1 – 0.5
Mg (%)	0.02 – 1.6	0.05 – 1.1	0.675 – 1.55	0 – 0.02	0.5 – 0.95	0.02 – 0.5
Impurities						
Si (%)	0.05 – 0.5	0.5 – 1.3	0.25 – 1.3	0.2 – 0.4	0.08 – 0.2	0.1 – 0.25
Fe (%)	0.05 – 0.5	0.3 – 1	0.9 – 1.6	0.3 – 0.7	0.1 – 0.3	0.15 – 0.3
Ni (%)	0	0 – 0.2	1 – 2	0	0	0 – 0.05
Ag (%)	0	0	0	0	0.25 – 0.35	0
Mechanical Properties						
YS (MPa)	295 – 352	140 – 287	239 – 295	180 – 244	310 – 462	263 – 331
Rupture strength (MPa)	279 – 325	215 – 275	260 – 325	190 – 235	270 – 279	261 – 270
Creep strength (MPa)	266 – 325	205 – 259	260 – 315	170 – 203	252 – 266	245 – 252
Modulus of Toughness (MPa)	40.63 – 47.71	38.75 – 51.45	26.5 – 45.38	38.4 – 52.0	40.8 – 54.69	42.185 – 48.70

To further refine the classification, sub-clustering was performed using a combined approach of PCA, K-means clustering, and a mathematical model, as described in detail in **Section 3.5**. This approach aimed to further refine the parent clusters into sub-clusters exhibiting narrower ranges of composition and properties. The six primary clusters were consequently subdivided into 15 sub-clusters (Figure 11.4), providing a more precise classification of the alloys.

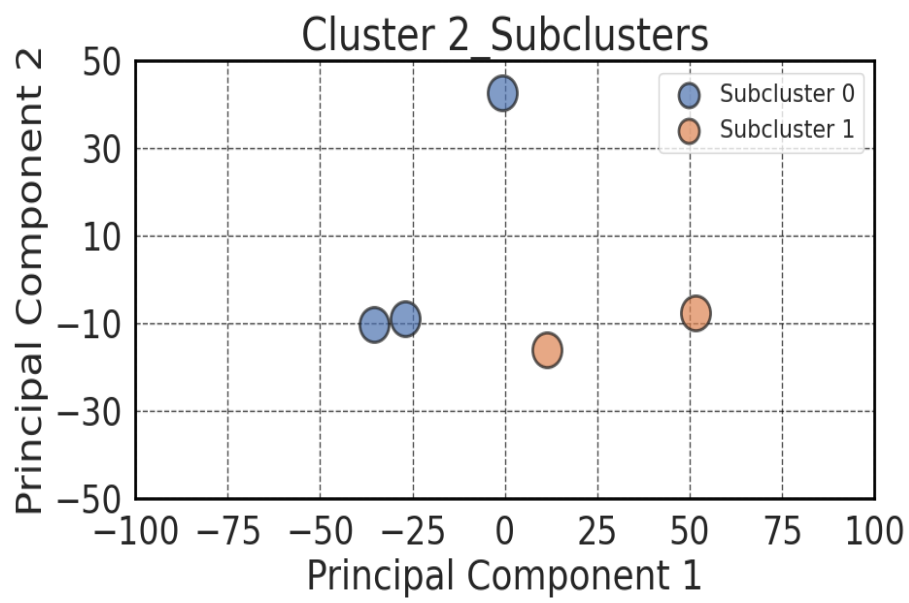
For example, sub-cluster 40 within Cluster 4 comprises high-strength alloys with elevated creep and rupture strengths, as well as a high modulus of toughness. In contrast, sub-cluster 41 of Cluster 4 contains alloys with high yield strength but moderate creep and rupture strengths, while maintaining a high modulus of toughness. This hierarchical sub-clustering methodology allows for a more detailed and nuanced distribution of alloys within the sub-clusters, thereby enhancing the overall classification and understanding of the alloys' compositional and mechanical characteristics for further optimisation in **Sections 7.2.1** and **7.2.2**.



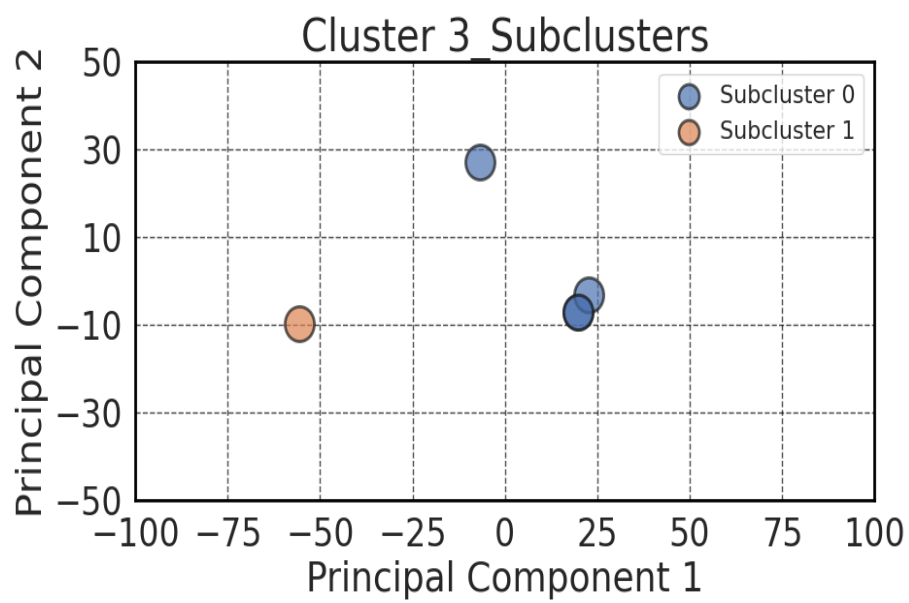
(a)



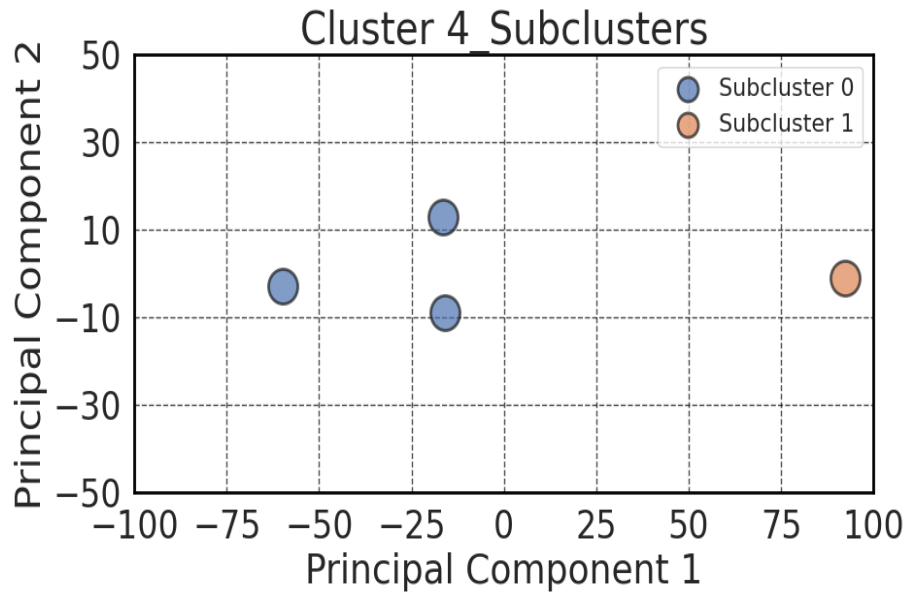
(b)



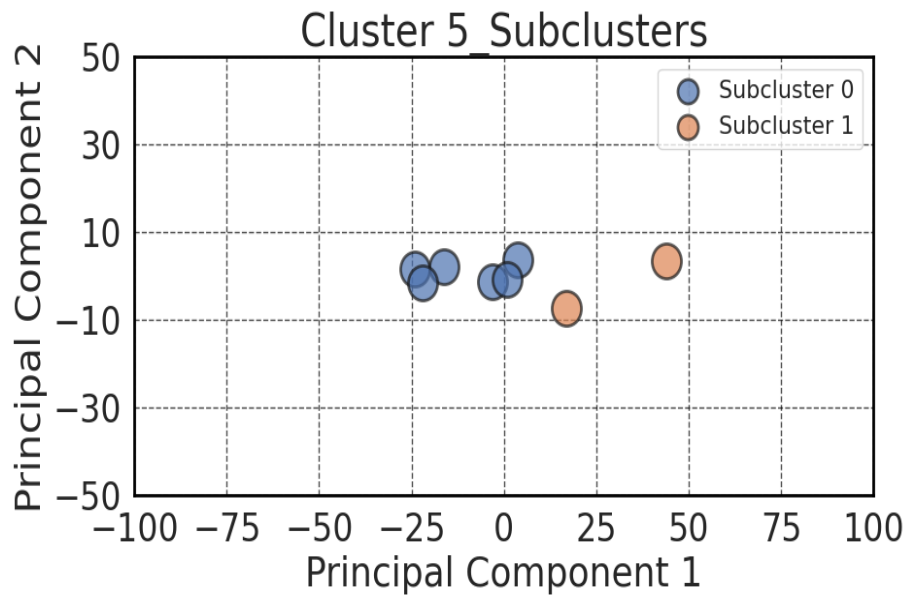
(c)



(d)



(e)



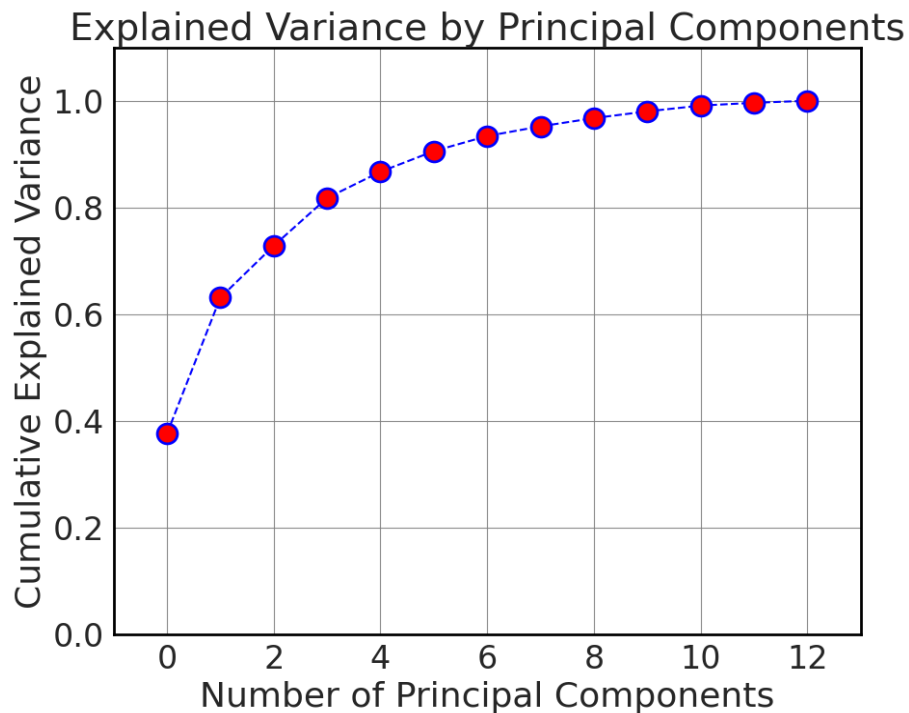
(f)

Figure 11.4. Sub-clustering analysis of 2xxx series alloys: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, (e) Cluster 4, and (f) Cluster 5

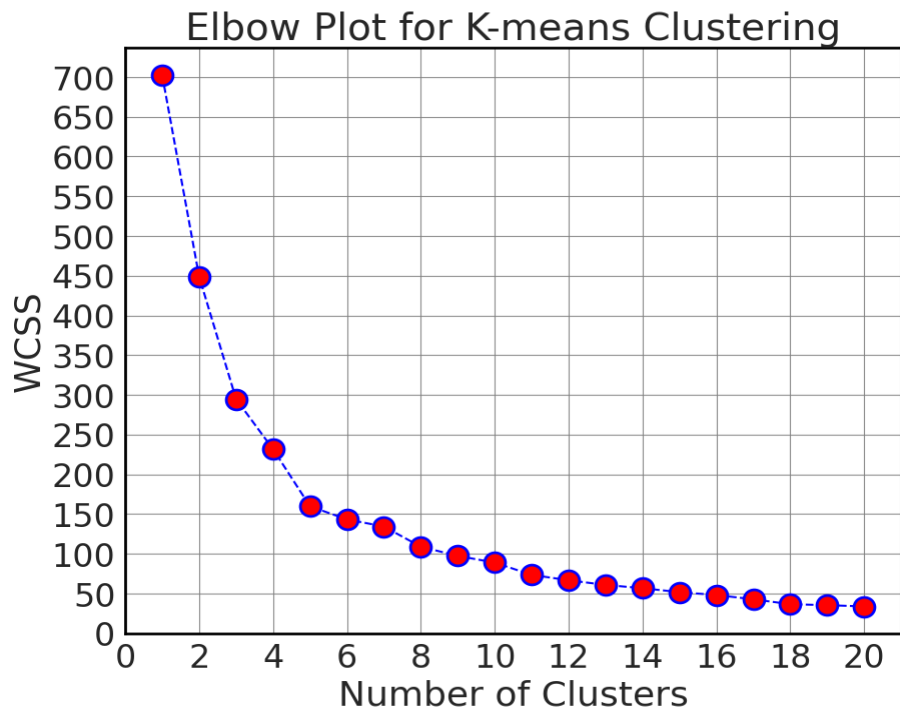
11.2.2 Clustering and Sub-clustering of 5xxx Series Alloys

The 5xxx series alloys are represented within a 13-dimensional space, comprising eight compositional features and five property features. Direct visualisation of such high-dimensional

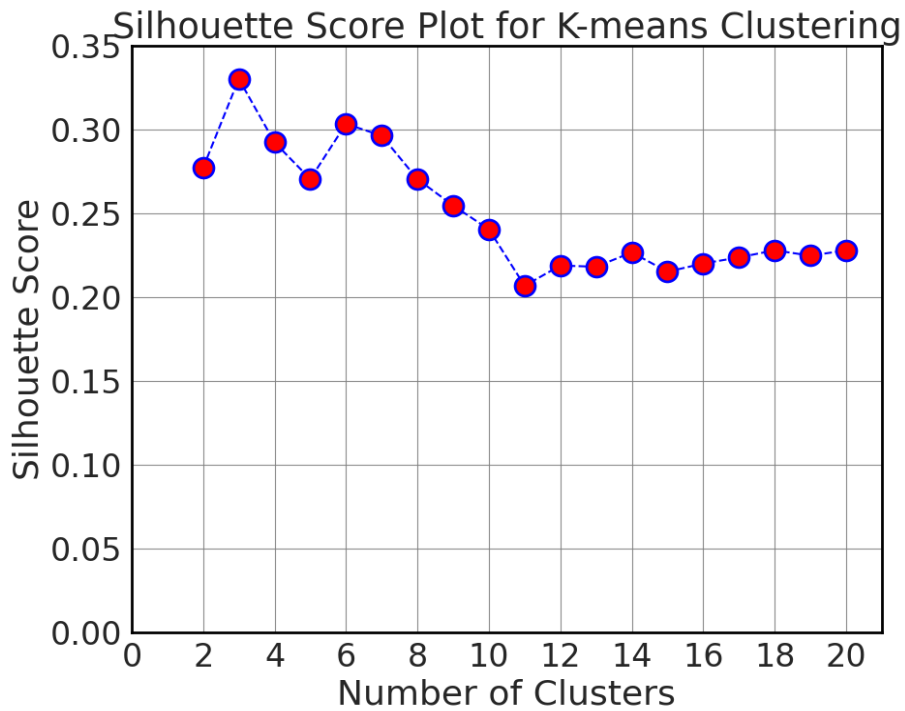
data is not feasible. To address this, principal component analysis (PCA) was employed to reduce the dimensionality of the dataset. PCA successfully compressed the dataset into four principal components, capturing approximately 80% of the total variance, as illustrated in Figure 11.5(a). This dimensionality reduction facilitated more straightforward analysis and visualisation of the data, while preserving the essential information pertaining to the alloys' compositions and properties.



(a)



(b)



(c)

Figure 11.5. Plots to detect the number of clusters in 5xxx-series alloys (a) PCA variance plot (b) an Elbow plot and, (c) Silhouette score

Following dimensionality reduction, the elbow method was employed to determine the optimal number of clusters for clustering the alloys based on their compositions and properties, as shown in Figure 11.5(b). The elbow plot indicated that $K=3$ and $K=5$ were the most appropriate choices. This finding was further supported by the silhouette analysis, which yielded a value of 0.33 for $K=3$ (Figure 11.5(c)), confirming the presence of three well-defined clusters. The clustering results are illustrated in Figure 11.6, where the three clusters are clearly separated.

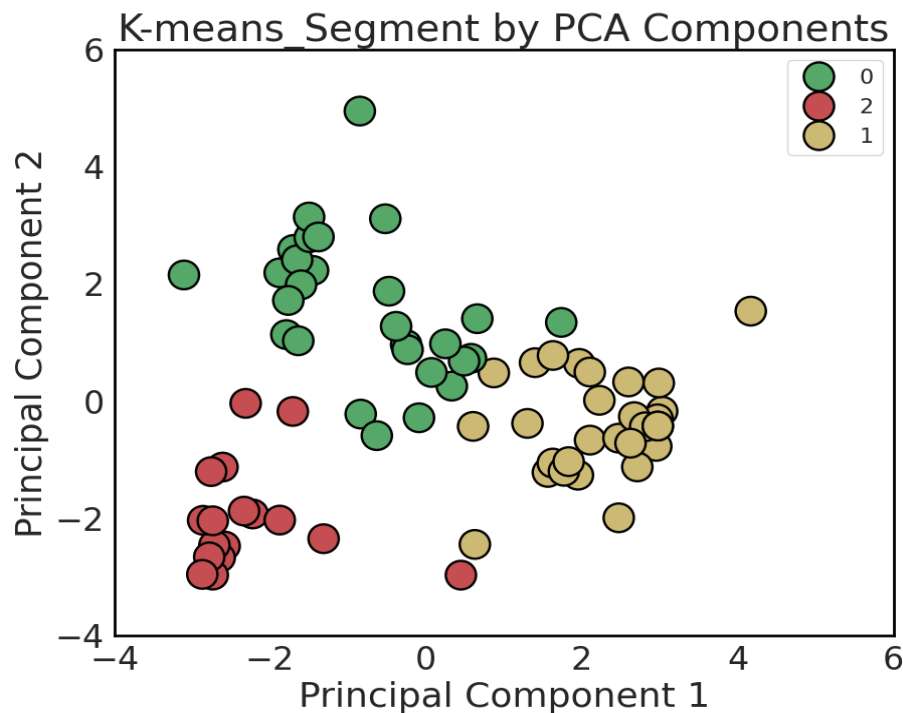


Figure 11.6. Clustering of 5xxx series alloys

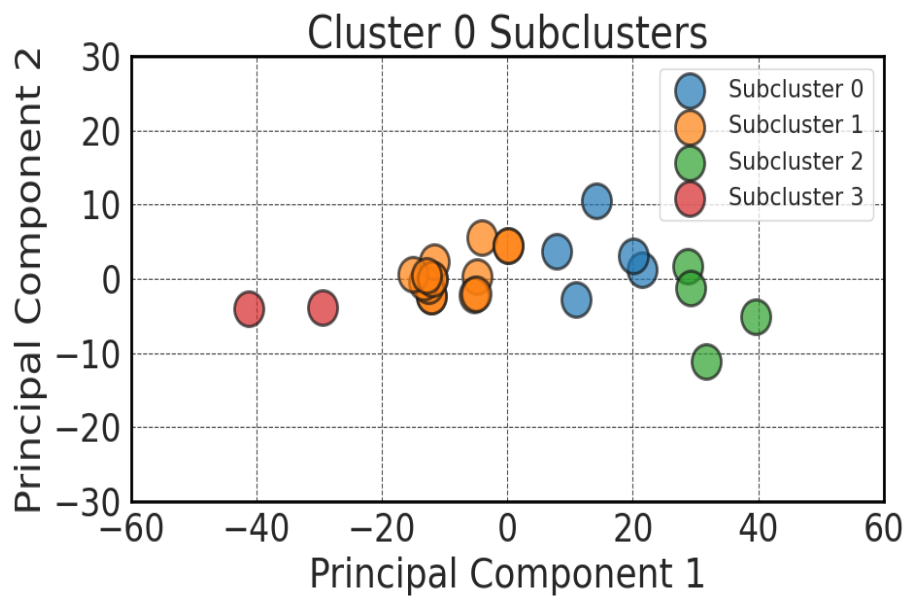
After analysing the primary clusters, categorised into low, medium, and high property ranges as shown in Table 11.2, we observed that each cluster exhibited distinct characteristics. For instance, Cluster 1, with its higher Mg content, corresponds to alloys with increased strength. In contrast, Cluster 2, which has a lower Mg content, displays comparatively reduced strength.

Table 11.2. Property and composition ranges in clusters

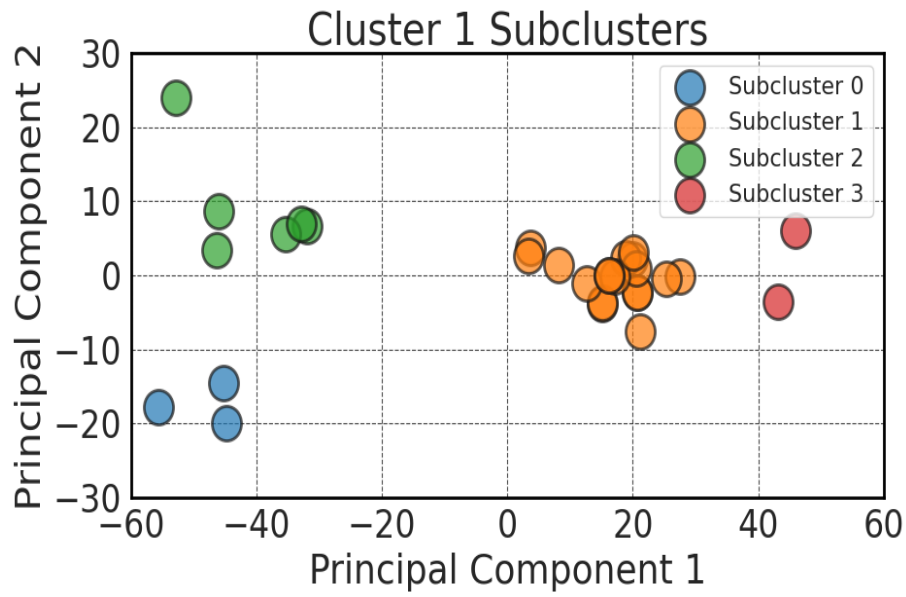
Property/Element	Cluster 0 (medium)	Cluster 1 (high)	Cluster 2 (low)
Si (%)	0.25 – 0.7	0 – 0.45	0 – 0.2
Fe (%)	0.175 – 0.4	0 – 0.25	0 – 0.35
Cu (%)	0.05 – 0.6	0.03 – 0.5	0 – 0.2
Mn (%)	0.1 – 1.4	0.01 – 1.2	0 – 0.45
Mg (%)	1 – 3.5	3.5 – 5.6	0.4 – 2.5

Cr (%)	0 – 0.25	0.05 – 0.175	0 – 0.05
Zn (%)	0.1 – 0.4	0.1 – 2.8	0.01 – 0.15
Ti (%)	0 – 0.1	0 – 0.1	0 – 0.05
YTS (MPa)	160 – 230	219 – 310	186 – 203
Fatigue strength (MPa)	90 – 130	129 – 180	119 – 128
Modulus of Toughness (MPa %)	12.95 – 24.56	22.78 – 28.62	17.13 – 19.09
Weldability	Good – Excellent	Good – Excellent	Good – Excellent
Corrosion Resistance	Good – Excellent	Good – Excellent	Good – Excellent

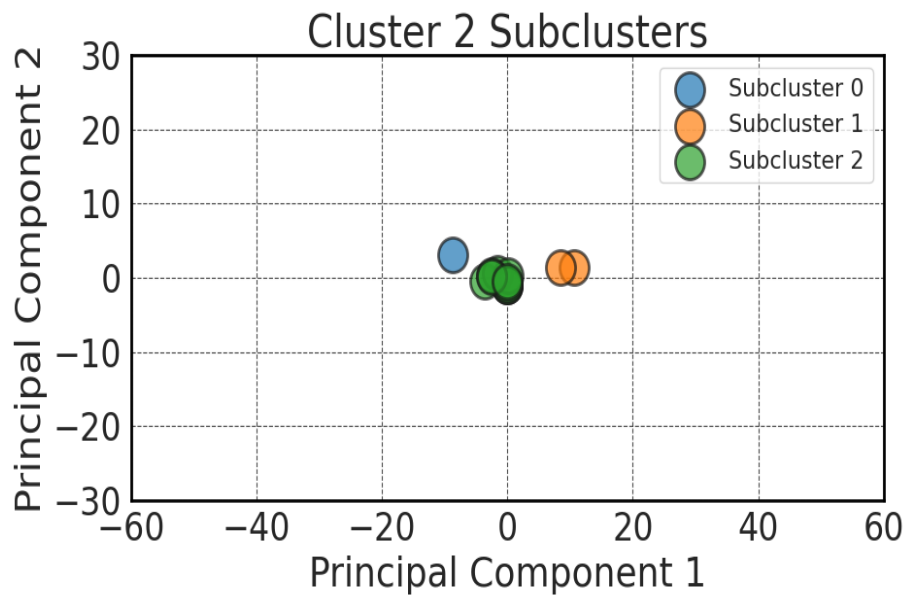
To further refine the classification, a sub-clustering process was performed, as shown in Figure 11.7, dividing the three primary clusters into 11 smaller, more specific sub-clusters. This allowed the identification of groups of alloys with more compact and distinctive combinations of properties. For example, sub-cluster 10 within Cluster 1 contains alloys with very high strength, whereas sub-cluster 11 includes alloys with medium strength but excellent corrosion resistance. This refined sub-clustering preserves critical information about the properties of alloys, which will serve as a basis for optimum alloy selection in **Sections 7.3.1** and **7.3.2**.



(a)



(b)



(c)

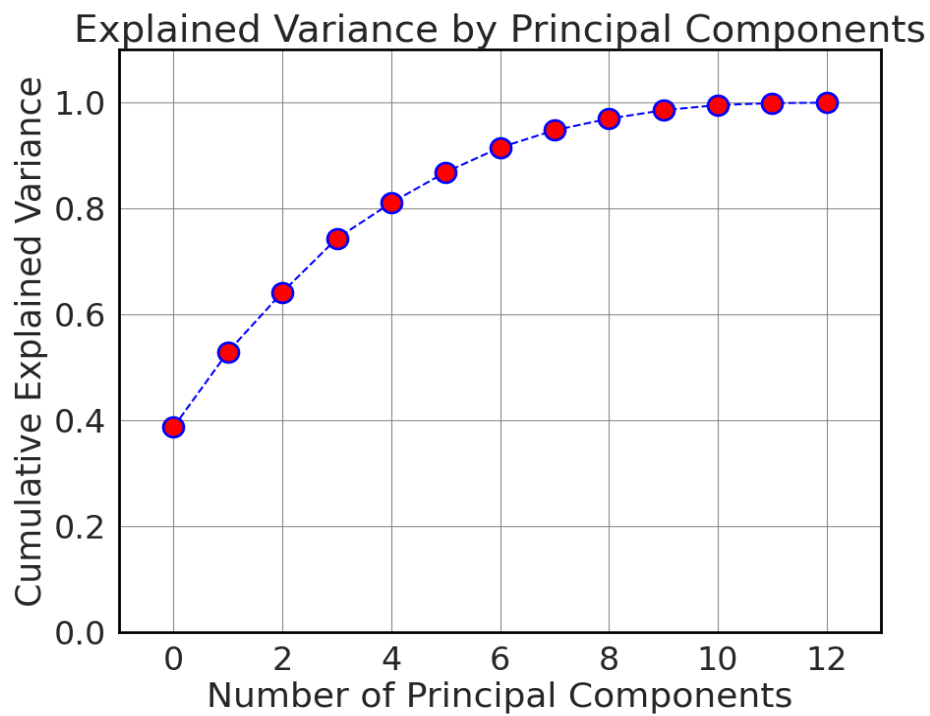
Figure 11.7. Sub-clustering of 5xxx series alloys: (a) Cluster 0, (b) Cluster 1, and (c) Cluster

2

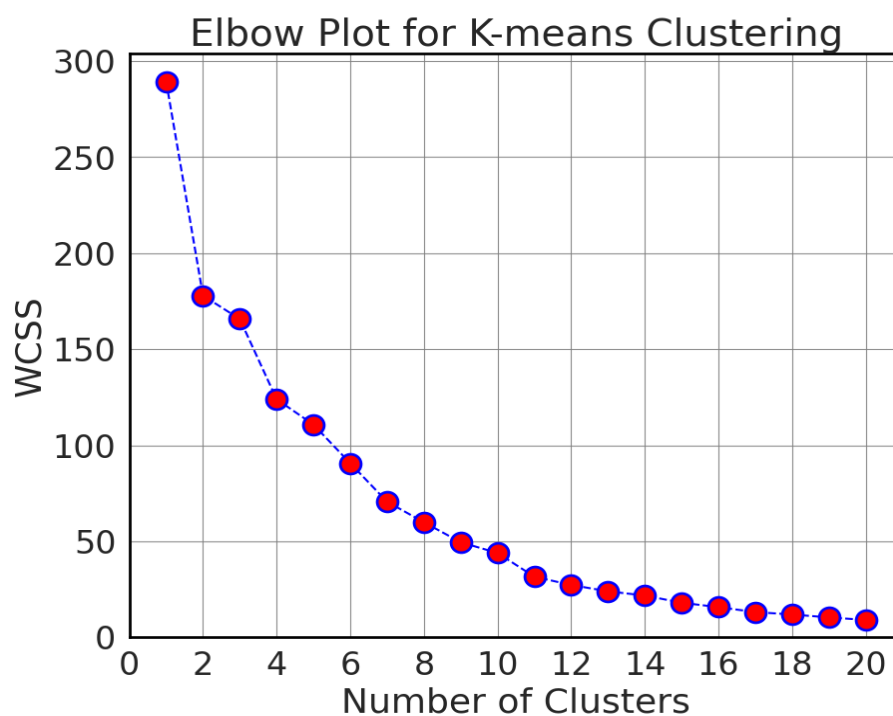
11.2.3 Clustering and Sub-clustering of 3xxx Series Alloys

We performed a clustering analysis of 36 alloys from the 3xxx series using a combined approach of PCA and K-means clustering to cluster alloys with similar composition and property ranges. Initially, PCA was applied to reduce the dimensionality of the dataset, which contained 13 attributes representing both compositions and properties. As shown in the variance plot

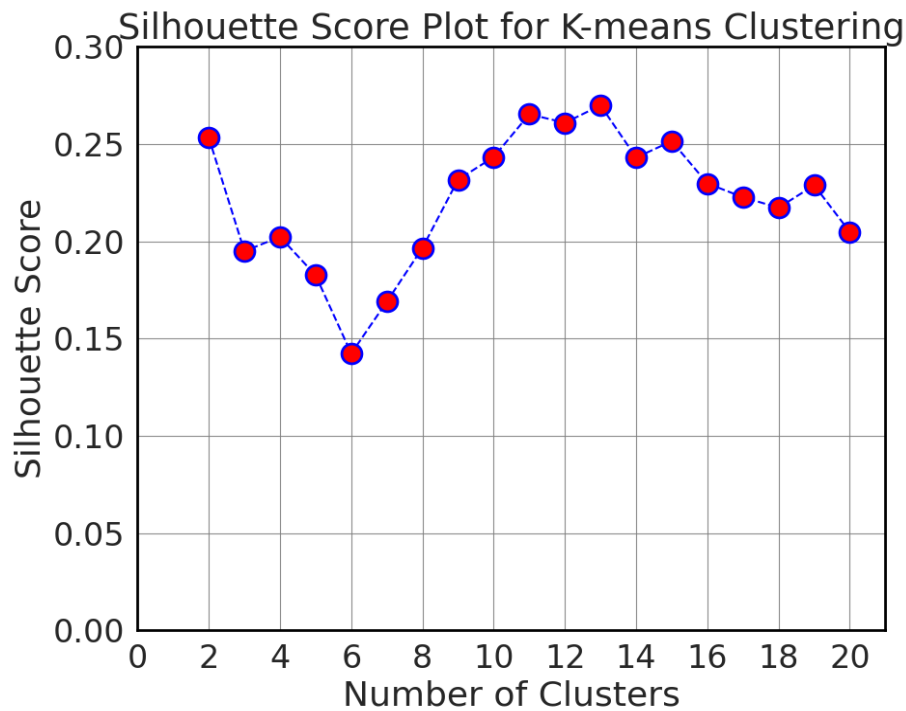
(Figure 11.8(a)), PCA reduced the dimensionality from 13 to four principal components, capturing approximately 75% of the total variance of the dataset. This reduction not only facilitated clearer visualisation of the clusters but also enhanced the effectiveness of the subsequent K-means clustering.



(a)



(b)



(c)

Figure 11.8. Plots to detect the number of clusters in 3xxx-series alloys (a) PCA variance plot, (b) an Elbow plot and, (c) Silhouette score

After reducing the dataset's dimensionality to four principal components, the optimal number of clusters was determined using the elbow method, which showed a distinct kink at $K=3$ and $K=4$ number of clusters, as shown in Figure 11.8(b). The silhouette score further supported this selection, showing high values at both two and four clusters. Considering both metrics, four clusters were selected as the most clearly defined, with a silhouette score of 0.20, as shown in Figure 11.8(c). The K-means algorithm subsequently visualised four distinct clusters, as

illustrated in Figure 11.9, providing meaningful groupings based on the alloys' compositions and properties.

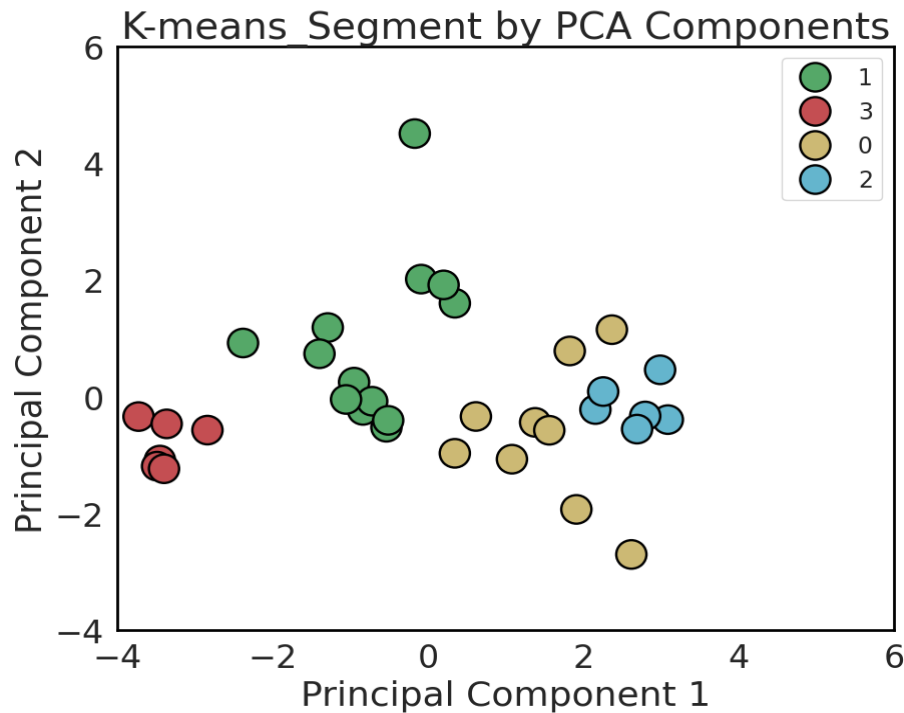


Figure 11.9. Clustering of 3xxx series alloys

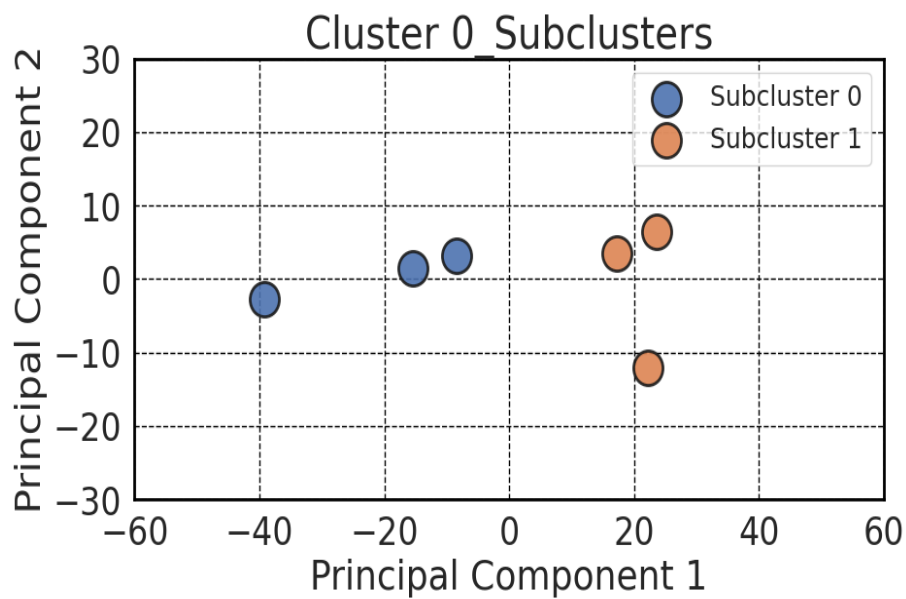
The four distinct clusters identified exhibit specific ranges of properties and compositions, largely influenced by Mn and Mg content, which play a critical role in determining the alloys' characteristics. As shown in Table 11.3, Cluster 3 contains high levels of both Mn and Mg, resulting in alloys with high strength and excellent corrosion resistance. In contrast, Cluster 2 contains lower amounts of these elements, particularly near 0% Mg, leading to alloys with reduced strength and moderate weldability.

Table 11.3. Composition and properties range's in 3xxx series clusters

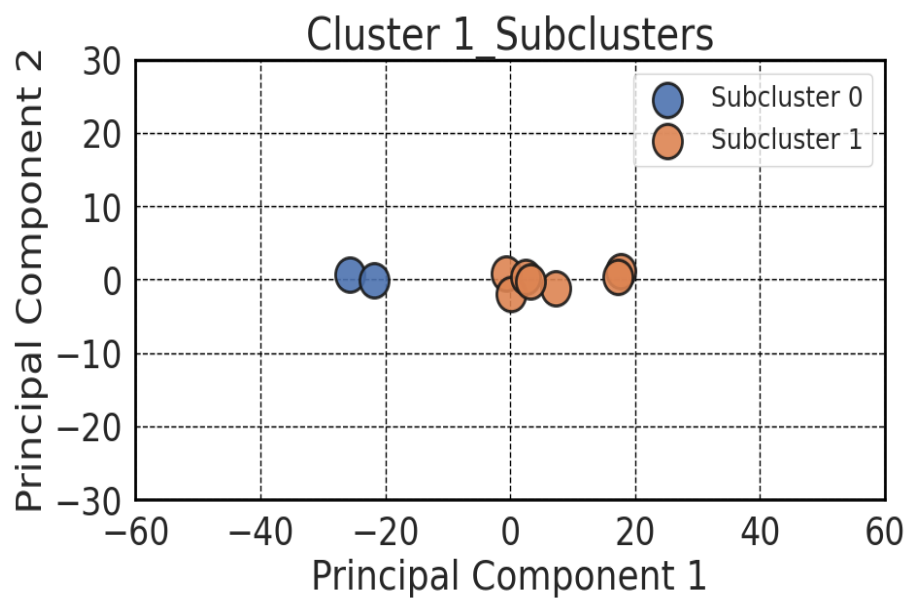
Property/Element	Cluster 0 (Medium)	Cluster 1 (High)	Cluster 2 (Low)	Cluster 3 (Highest)
Alloying Elements				
Mn (%)	0.65 – 1.5	0.425 – 1.0	0.15 – 0.65	0.55 – 1.25
Si (%)	0.5 – 1.8	0.2 – 0.6	0.08 – 0.4	0.3 – 1.3
Fe (%)	0.35	0.15 – 0.4	0.05 – 0.35	0.35 – 0.5
Cu (%)	0.05 – 0.2	0.1 – 0.8	0.03 – 0.15	0.25 – 0.9
Mg (%)	0 – 0.3	0 – 0.6	0 – 0.2	0.1 – 1.5
Cr (%)	0 – 0.05	0 – 0.2	0 – 0.2	0 – 0.1
Zn (%)	0 – 0.2	0 – 0.5	0.05 – 0.5	0.25 – 1
Ti (%)	0 – 0.05	0.025 – 0.125	0 – 0.175	0 – 0.05
Mechanical Properties				

YS (MPa)	39 – 100	75 – 113	34 – 77	107 – 170
Fatigue Strength (MPa)	43 – 73	62 – 83	34 – 61	80 – 120
Modulus of Toughness (MPa %)	10.67 – 17.14	10.52 – 16.87	17.28 – 19.57	3.45 – 15.6
Other Properties				
Weldability	Good – Excellent	Good	Fair – Good	Good
Corrosion Resistance	Good – Excellent	Good – Excellent	Good	Good – Excellent

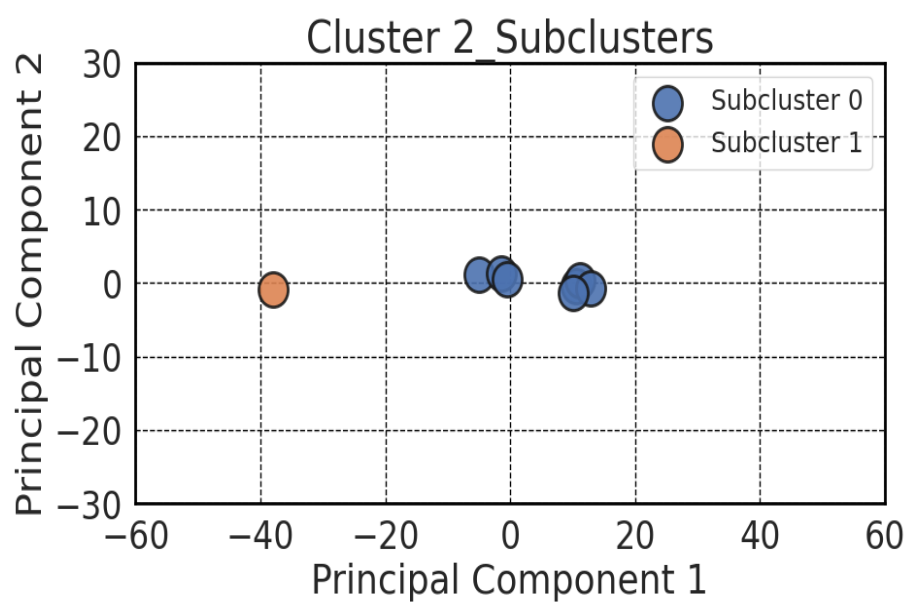
To further refine the grouping, we conducted a sub-clustering procedure, as shown in Figure 11.10, aimed at identifying more balanced and refined combination of properties within each sub-cluster. This step divided the initial clusters into finer subgroups, resulting in 9 refined sub-clusters. For instance, Cluster 1 was split into two sub-clusters: Sub-cluster 10, comprising high-strength alloys with enhanced corrosion resistance due to elevated Mn and Mg content, and Sub-cluster 11, containing alloys with lower Mn and Mg levels, resulting in medium-to-high strength and slightly reduced fatigue properties. This sub-clustering enabled a more detailed classification of the alloys, providing a clearer understanding of property variations within each sub-cluster for subsequent optimum alloy selection and recycling in **Sections 7.4.1** and **7.4.2**.



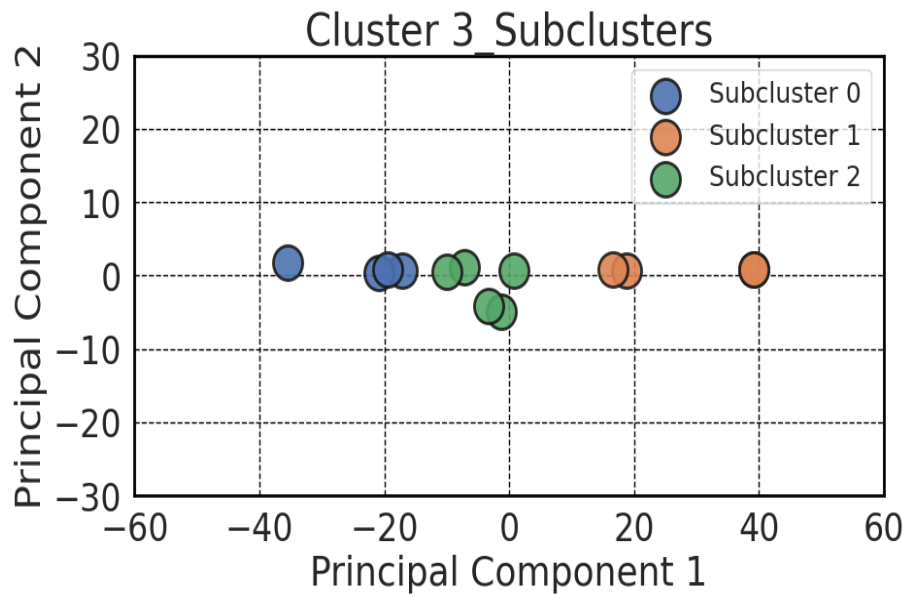
(a)



(b)



(c)



(d)

Figure 11.10. Sub-clustering of 3xxx series alloys: (a) Cluster 0, (b) Cluster 1, (c) Cluster 2, and (d) Cluster 3

11.3 Datasets and Reproducibility Workflow

11.3.1 Dataset 1: Preliminary Analysis

11.3.1.1 Raw Data

The raw data for T6 6xxx series extrusion alloys is provided in **Preliminary_Analysis_Raw_Dataset.csv** and Table 11.4 describes the dataset overview. This dataset contains all measured elemental compositions, mechanical properties, and categorical ratings of corrosion resistance, weldability, and anodising response for each alloy grade. It serves as the foundation for clustering, multi-criteria optimisation, and ANN-based analysis, and allows reproducibility of results.

Table 11.4. Preliminary Analysis Raw Dataset Overview

Column	Description
Alloy grade	Unique identifier for each alloy
Si, Fe, Cu, Mn, Mg, Mg:Si	Measured elemental compositions
Yield Strength, Modulus of Toughness, Fatigue strength	Mechanical properties
Corrosion, Weldability, Anodising	Categorical properties

Notes:

- This dataset allows reproduction of all preprocessing, exploratory analyses, and clustering subsets.
- Serves as the primary input for subsequent data processing, optimisation, and ANN modelling.

11.3.1.2 Processed Data

The processed dataset (**Preliminary_Analysis_Processed_Dataset.csv**) contains all standardized and computed features derived from the raw data. It consolidates data for clustering, PCA, multi-criteria optimisation, and ANN modelling into a single file, enabling full reproducibility of results as shown in Table 5.

Table 11.5. Preliminary Analysis Processed Dataset Overview

Component	Columns	Description
Standardized Inputs for Clustering	Si_Standardized, Fe_Standardized, Cu_Standardized, Mn_Standardized, Mg_Standardized, Mg:Si_Standardized, Yield_Strength_Standardized, Modulus_of_Toughness_Standardized, Fatigue_Standardized	Standardized elemental compositions and mechanical properties (mean=0, std=1) used for clustering algorithms
PCA Components	Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4	Principal components computed from standardized inputs for dimensionality reduction and visualization
Cluster Assignments	Kmeans_Clustering, K-means_PCA_Clustering, Hierarchical_Clustering, DBSCAN_Clustering, MeanShift_Clustering, GMM_Clustering, Final_Clusters	Cluster assignments from multiple algorithms; Final_Clusters is the consensus cluster used for further analysis and optimisation
Multi-Criteria Optimisation	TOPSIS_Score, AHP_Score, VIKOR_Score, TOPSIS_Rank, AHP_Rank, VIKOR_Rank	Multi-criteria optimisation scores and rankings of alloys within each consensus cluster
ANN Processed Data	Si_scaled_ANN, Fe_scaled_ANN, Cu_scaled_ANN, Mn_scaled_ANN, Mg_scaled_ANN, Predicted_Yield_strength_scaled_ANN, Predicted_Modulus_of_Toughness_scaled_ANN, Predicted_Fatigue_scaled_ANN,	ANN input features (scaled), predicted outputs (scaled and original units), composite score for each alloy, and gradient-based feature importance values

	Predicted_Yield_strength_ANN, Predicted_Modulus_of_Toughness_ANN, Predicted_Fatigue_ANN, ANN_Composite_Score, Importance_Si_ANN, Importance_Fe_ANN, Importance_Cu_ANN, Importance_Mn_ANN, Importance_Mg_ANN	

Notes:

- Standardization and PCA are applied globally to ensure reproducibility of clustering and visualisations.
- Multiple clustering algorithms were tested for stability; K-means (raw and PCA-assisted) was selected based on silhouette scores and consistency.
- Multi-criteria optimisation (TOPSIS, AHP, VIKOR) uses consensus clusters; a test subset (Cluster 0) validated scoring reliability.
- ANN inputs are scaled per-cluster; outputs are both scaled (0–1) and in original units.
- ANN Composite Score sums scaled predicted targets (Yield + Toughness + Fatigue) and serves as the primary metric for alloy ranking.
- ANN feature importance shows the influence of each element on predicted performance.

11.3.1.3 Reproducibility Notes for Preliminary Analysis

This section provides a complete trace of the workflow for T6 6xxx series extrusion alloys, from raw measurements to ANN-based composite scoring.

Step 1: Raw Data

- **File:** Preliminary_Analysis_Raw_Dataset.csv
- **Columns:** Alloy grade, Si, Fe, Cu, Mn, Mg, Mg:Si, Yield Strength, Modulus of Toughness, Fatigue, Corrosion, Weldability, Anodising
- **Purpose:** Primary input for clustering, PCA, multi-criteria optimisation, and ANN modelling.

Step 2: Missing Value Imputation (Random Forest)

- **Method:** Random Forest Regression for numerical features (Yield Strength, Modulus of Toughness, Fatigue, etc.) and Random Forest Classification for categorical features (Corrosion, Weldability, Anodising).
- **Purpose:** Fill missing values (if any) to produce a complete dataset for PCA, clustering, multi-criteria optimisation, and ANN modelling.
- **Processed Data:** Missing values are replaced in Preliminary_Analysis_Processed_Dataset.csv (if any).

Step 3: Standardization

- **Method:** StandardScaler (mean=0, std=1)
- **Columns:** Si_Standardized, Fe_Standardized, Cu_Standardized, Mn_Standardized, Mg_Standardized, Mg:Si_Standardized, Yield_Strength_Standardized, Modulus_of_Toughness_Standardized, Fatigue_Standardized
- **Purpose:** Makes features comparable and avoids bias in clustering.

Step 4: PCA:

- **Components:** Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4
- **Input:** Standardized elemental and mechanical features
- **Purpose:** Dimensionality reduction for visualisation and PCA-based clustering.
- Use the standardized inputs/PCA components when computing silhouette scores or elbow plots.

Step 5: Clustering

- **Algorithms Tested:** K-means (raw & with PCA), Hierarchical, DBSCAN, MeanShift, GMM
- **Final Clusters:** Final_Clusters
- **Validation:** Silhouette scores for each algorithm to assess cluster quality.

Step 6: Multi-Criteria Optimisation

- **Methods:** TOPSIS, AHP, VIKOR
- **Input:** All alloys in Final_Clusters

- **Outputs:** Scores (TOPSIS_Score, AHP_Score, VIKOR_Score) and ranks (TOPSIS_Rank, AHP_Rank, VIKOR_Rank)
- **Notes:** Ensures consistent ranking within clusters; Cluster 0 used to validate scoring reliability.

Step 7: ANN Modelling

- **Inputs:** Si_scaled_ANN, Fe_scaled_ANN, Cu_scaled_ANN, Mn_scaled_ANN, Mg_scaled_ANN (per-cluster scaled)
- **Targets:** Yield Strength, Modulus of Toughness, Fatigue
- **Architecture:** Sequential ANN, 6-neuron hidden layer (ReLU, L2 regularization), 3-neuron output layer (linear), Adam optimizer, MSE loss, 150 epochs, batch size = 1
- **Outputs:**
Scaled predictions: Predicted_Yield_strength_scaled_ANN, Predicted_Modulus_of_Toughness_scaled_ANN, Predicted_Fatigue_scaled_ANN
Original units: Predicted_Yield_strength_ANN, Predicted_Modulus_of_Toughness_ANN, Predicted_Fatigue_ANN
ANN_Composite_Score = sum of scaled predicted targets
Feature importance: Importance_Si_ANN ... Importance_Mg_ANN

Step 8: Consolidated Processed Dataset

- **File:** Preliminary_Analysis_Processed_Dataset.csv
- **Contents:** Standardized inputs, PCA components, cluster assignments (all algorithms + final), multi-criteria scores/ranks, ANN inputs, predictions, composite score, and feature importance
- **Purpose:** Full traceability from raw data to final alloy selection.

Step 9: Key Notes

- Standardization and PCA applied globally for reproducibility.
- Multiple clustering algorithms tested; K-means selected for consistency.
- Consensus clusters used for multi-criteria optimisation; all scores and ranks included.
- ANN inputs scaled per cluster; outputs stored in scaled and original units. Composite score and feature importance enable clear alloy ranking.
- Silhouette scores validate cluster quality and reproducibility.

11.3.2 Dataset 2: Non-Age-Hardenable Alloys

11.3.2.1 Raw Data

The raw data for non-age-hardenable alloys (3xxx, 5xxx, 1xxx and selected 8xxx series at T6 tempering), is provided in **Non_Age_Hardenable_Alloys_Raw.csv** (see **Section 11.3.6** for information on Dataset). As shown in Table 11.6, the dataset contains all measured elemental compositions, mechanical properties, and categorical ratings of weldability, corrosion resistance, and electrical/thermal performance for each alloy grade. It serves as the foundation for PCA, K-means clustering, and reproducibility of the analysis workflow.

Table 11.6. Non-Age-Hardenable Raw Dataset Overview

Column	Description
Alloy grade	Unique identifier for each alloy sample
Si, Fe, Cu, Mn, Mg, Cr, Zn, Ti, Ga, V, Zr	Measured elemental compositions
YTS, Fatigue, UTS, Elongation, Toughness	Mechanical properties
Weldability, Corrosion, Electrical, Thermal	Additional properties

Notes: This dataset serves as the primary input for data processing, PCA, and clustering

11.3.2.2 Processed Data

The processed dataset (**Non_Age_Hardenable_Alloys_Processed.csv**) contains all standardized and derived features from the raw data. It consolidates data for clustering, PCA, and clustering analysis, enabling full reproducibility of results.

Table 11.7. Non-Age-Hardenable Processed Dataset Overview

Component	Columns	Description
Standardized Inputs for Clustering	Si_Standardised, Fe_Standardised, Cu_Standardised, Mn_Standardised, Mg_Standardised, Cr_Standardised, Zn_Standardised, Ti_Standardised, Ga_Standardised, V_Standardised, Zr_Standardised, YTS_Standardised, Fatigue_Standardised, Toughness_Standardised, Electrical_Standardised, Thermal_Standardised	Standardized elemental compositions and mechanical, physical and categorical properties (mean=0, std=1) used for clustering algorithms

PCA Components	Component_1, Component_2, Component_3, Component_4, Component_5	Principal components computed from standardized inputs for dimensionality reduction and visualisation
Cluster Assignments	Cluster, Segment	Cluster and segment assignments obtained using K-means clustering on standardized inputs

Notes:

- Standardization ensures all features contribute equally to clustering.
- PCA components provide reduced dimensionality for visualisation and cluster validation.
- Cluster and segment assignments allow reproducibility of all PCA and K-means analyses for non-age-hardenable alloys.
- This dataset enables to trace the complete workflow from raw measurements to final cluster assignments.

11.3.2.3 Reproducibility Notes for Non-Age-Hardenable Alloys

This section provides a stepwise trace of the workflow for non-age-hardenable alloys (series 1xxx, 3xxx, 5xxx, and selected 8xxx), from raw measurements to PCA, clustering, and processed data, ensuring full reproducibility.

Step 1: Raw Data Collection: File: Non_Age_Hardenable_Alloys_Raw.csv

Contents:

- **Alloy grade:** Sample identifier
- **Elemental compositions:** Si, Fe, Cu, Mn, Mg, Cr, Zn, Ti, Ga, V, Zr
- **Mechanical properties:** YTS, Fatigue, UTS, Elongation, Modulus of toughness
- **Categorical properties:** Weldability, Corrosion
- **Physical properties:** Electrical, Thermal conductivity
- **Purpose:** Provides the measured inputs for standardization, PCA, and clustering.

Step 2: Standardization of Inputs

Method: StandardScaler (mean=0, std=1) applied globally to all alloys

Columns Standardized:

- Si_Standardised, Fe_Standardised, Cu_Standardised, Mn_Standardised, Mg_Standardised, Cr_Standardised, Zn_Standardised, Ti_Standardised, Ga_Standardised, V_Standardised, Zr_Standardised
- YTS_Standardised, Fatigue_Standardised, Toughness_Standardised, Electrical_Standardised, Thermal_Standardised

Purpose:

- Ensures all features contribute equally to clustering.
- Prevents bias due to differences in scales among elemental and mechanical properties.

Step 3: Dimensionality Reduction (PCA)

Method: Principal Component Analysis (PCA)

Components Used: Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4, Principal_Component_5

Input: Standardized elemental and mechanical properties

Purpose:

- Reduces dimensionality for visualisation and clustering validation.
- PCA-based K-means clustering uses these components to form clusters in reduced space.

Step 4: Clustering Analysis

Algorithms Tested:

- K-means combined with PCA on raw standardized inputs

Cluster Assignments:

- Cluster and Segment columns in processed dataset
- Final clusters used for subsequent analyses

Validation Metrics:

- Silhouette scores computed for each clustering variant to assess quality
- PCA assisted K-means was selected for reproducibility based on silhouette score and domain understanding

Step 5: Consolidated Processed Dataset

File: Non_Age_Hardenable_Alloys_Processed.csv

Contents:

- **Standardized inputs:** All elemental, mechanical, and categorical properties
- **PCA components:** Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4, Principal_Component_5
- **Cluster assignments:** Cluster, Segment
- **Derived features:** Enables full traceability for visualisation, cluster validation

Purpose:

- Provides a single consolidated file with all processed information for reproducibility
- Enables to replicate PCA, clustering, and visualisations exactly as performed in text

Step 6: Notes on Reproducibility

- Standardization is applied globally across all alloys.
- PCA ensures consistent dimensionality reduction for cluster comparison.
- K-means clustering validated via silhouette scores ensures stable cluster formation.
- Consolidated processed dataset allows reproduction of all plots, and analyses derived from raw measurements.

11.3.3 Dataset 3: Age-Hardenable Alloys**11.3.3.1 Raw Data**

The raw data for age-hardenable alloys (series 2xxx, 6xxx, 7xxx, and selected 8xxx) is provided in **Age_Hardenable_Alloys_Raw_Data.csv** and an overview is shown in Table 11.8. This dataset contains all measured elemental compositions, mechanical properties, and categorical ratings (weldability, corrosion), and electrical/thermal properties for each alloy grade. It serves as the foundation for PCA, K-means clustering, and subsequent analyses, allowing full reproducibility.

Table 11.8. Age-Hardenable Alloys Raw Dataset Overview

Column	Description
Alloy grade	Unique identifier for each alloy sample
Si, Fe, Cu, Mn, Mg, Cr, Zn, Ti, Ag, Li, Ni, Zr	Measured elemental compositions
YTS, Fatigue, UTS, Elongation, Toughness	Mechanical properties
Weldability, Corrosion, Electrical, Thermal	Categorical ratings and physical properties

Notes:

- This dataset allows full reproduction of preprocessing, exploratory analyses, and clustering subsets.
- Serves as primary input for standardization, PCA, and K-means clustering

11.3.3.2 Processed Data

The processed dataset **Age_Hardenable_Alloys_Processed.csv** contains all standardized, computed, and derived features as shown in Table 11.9. It consolidates data for clustering, PCA, and full analysis into a single file, enabling complete reproducibility.

Table 11.9. Age-Hardenable Alloys Processed Dataset Overview

Component	Columns	Description
Standardized Inputs for Clustering	Si_Standardised, Fe_Standardised, Cu_Standardised, Mn_Standardised, Mg_Standardised, Cr_Standardised, Zn_Standardised, Ti_Standardised, Ag_Standardised, Li_Standardised, Ni_Standardised, Zr_Standardised, YTS_Standardised, Fatigue_Standardised, Toughness_Standardised, Electrical_Standardised, Thermal_Standardised	Standardized elemental compositions and mechanical properties (mean=0, std=1) used for clustering algorithms
PCA Components	Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4, Principal_Component_5, Principal_Component_6	Principal components computed from standardized inputs for dimensionality reduction and visualization
Cluster Assignments	Cluster, Segment	Cluster and segment assignments from K-means clustering

11.3.3.3 Reproducibility Notes

Reproducibility: same as non-age-hardenable (**Section 11.3.2.3**); in Step 1 and Step 6, use **Age_Hardenable_Alloys_Processed.csv** instead of **Non_Age_Hardenable_Alloys_Processed.csv**.

11.3.4 Dataset 4: 6xxx Series Alloys

11.3.4.1 Raw dataset

The raw data for 6xxx series alloys is provided in **6xxx_Alloys_Raw_Dataset.csv** with a brief overview given in Table 11.10. This dataset contains all measured elemental compositions, mechanical properties, and processing information for each alloy grade across all temper conditions and forms. It serves as the basis for PCA, clustering, sub-clustering, and multi-criteria optimisation.

Table 11.10. 6xxx Series Raw Dataset Overview

Column	Description
Alloy grade	Unique identifier for each alloy
Si, Fe, Cu, Mn, Mg, Mg:Si	Measured elemental compositions and derived Mg:Si ratio
Yield Strength, Modulus of Toughness, Fatigue	Mechanical properties
Corrosion, Weldability, Anodising Temper (T5, T6, T7), Processing forms (Extrusion, Plate, Sheet, Rolled, Forged)	Categorical ratings, tempering, and processing conditions

Notes:

- This dataset enables reproduction of all preprocessing, PCA, clustering, and sub-clustering analyses.
- Includes all temper conditions (T5, T6, T7) and forms (extrusion, rolling, sheet, plate, forged) at T6.

11.3.4.2 Processed dataset

The processed dataset (**6xxx_Alloys_Processed_Dataset.csv**) consolidates standardized, PCA, clustering, sub-clustering, normalized, weighted, and multi-criteria computed features for the 6xxx series alloys. It allows full traceability from raw measurements to alloy ranking and evaluation.

Table 11.11. 6xxx Series Processed Dataset Overview

Component	Columns	Description
Standardized Inputs	Si_Standardized, Fe_Standardized, Cu_Standardized, Mn_Standardized, Mg_Standardized,	Standardized compositions and properties (mean=0, std=1) for PCA and clustering reproducibility

	Mg:Si_Standardized, Toughness_Standardized, Yield Strength_Standardized, Fatigue_Standardized	
PCA Components	Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4	Principal components calculated from standardized inputs for dimensionality reduction
Cluster and Sub-cluster Assignments	Cluster, Principal_Component_1_Sub-Clustering, Principal_Component_2_Sub-Clustering, Sub-Cluster	Initial clusters and sub-clusters based on PCA for detailed grouping
Normalized & Weighted Properties	Yield Strength_Normalized, Modulus of Toughness_Normalized, Fatigue_Normalized, Corrosion_Normalized, Weldability_Normalized, Anodising_Normalized; Yield Strength_Weighted, Modulus of Toughness_Weighted, Fatigue_Weighted, Corrosion_Weighted, Weldability_Weighted, Anodising_Weighted	Normalized and weighted performance metrics per alloy
Multi-Criteria Optimisation	TOPSIS_Closeness_Score	Multi-criteria ranking of alloys considering all weighted properties
Computed Columns	Mg ₂ Si, Q-Phase, Mixing Ratio	Thermo-Calc calculations (Mg ₂ Si and Q-Phase at T6) and normalized mixing ratio for recycling evaluation
LIME features contribution	YS_Si_contrib, YS_Fe_contrib, YS_Cu_contrib, YS_Mn_contrib, YS_Mg_contrib, YS_Mg:Si_contrib, MT_Si_contrib, MT_Fe_contrib, MT_Cu_contrib, MT_Mn_contrib, MT_Mg_contrib, MT_Mg:Si_contrib	LIME Feature Contributions for yield strength and modulus of toughness
Additional Info	Temper, Processing	Alloy temper condition and processing route

Notes:

- Standardization and PCA are applied globally to ensure reproducibility of clustering, sub-clustering, and all visualisation.
- Sub-clustering is performed on PCA components to refine groupings.
- Normalized and weighted properties are combined in TOPSIS to rank alloy performance across mechanical and processing attributes.
- Calculated columns (Mg₂Si, Q-Phase, Mixing Ratio) provide thermodynamic and recycling insights.

- Other conditions datasets i.e., at T5, T7, are provided for reference in the processed csv file for PCA and K-means clustering.
- LIME features contributions provide the behaviour of alloys within the clusters.
- Enables full reproduction of PCA, clustering, sub-clustering, multi-criteria evaluation workflows and LIME analysis.

11.3.4.3 Reproducibility Notes for 6xxx Series Alloys

This section provides a stepwise trace of the workflow for 6xxx series alloys, from raw measurements to PCA, clustering, sub-clustering, and multi-criteria optimisation, ensuring full reproducibility of all plots, tables, and analyses.

Step 1: Raw Data Collection

File: **6xxx_Alloys_Raw_Dataset.csv** (See Section 11.3.6 for information on dataset)

Contents:

- Alloy grade: Alloy identifier
- Elemental compositions: Si, Fe, Cu, Mn, Mg, Mg:Si
- Mechanical properties: Yield Strength, Modulus of Toughness, Fatigue
- Categorical/Processing properties: Corrosion, Weldability, Anodising Temper, Processing route

Purpose: Provides all measured inputs for standardization, PCA, clustering, sub-clustering, and multi-criteria optimisation.

Step 2: Exploratory Data Analysis (EDA): EDA was solely performed on the raw dataset to visualise the dataset distributions, relationships, and trends, no processed data generated here.

Step 3: Standardization of Inputs

Method: StandardScaler (mean=0, std=1) applied globally

Columns Standardized:

- Elemental: Si_Standardized, Fe_Standardized, Cu_Standardized, Mn_Standardized, Mg_Standardized, Mg:Si_Standardized

- Mechanical: Yield Strength_Standardized, Modulus of Toughness_Standardized, Fatigue_Standardized

Purpose:

- Ensures all features contribute equally to PCA and clustering
- Prevents bias due to differences in scale among elemental and mechanical properties

Step 3: Dimensionality Reduction (PCA)

Method: Principal Component Analysis (PCA)

Components Used: Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4

Input: Standardized elemental and mechanical properties

Purpose:

- Reduces dimensionality for visualization and cluster validation
- PCA components used for both initial clustering and sub-clustering

Step 4: Clustering and Sub-Clustering Analysis

Primary Clustering:

- K-means clustering on PCA components
- Cluster assignments recorded in Cluster column

Sub-Clustering:

- Performed on PCA components within each primary cluster
- Sub-cluster assignments recorded in Sub-Cluster column

Validation Metrics:

- Silhouette score computed to evaluate cluster quality
- Ensures stable and reproducible grouping for subsequent optimisation

Step 5: Consolidated Processed Dataset

File: 6xxx_Alloys_Processed_Dataset.csv

Contents:

- Standardized inputs: All elemental and mechanical features
- PCA components: Principal_Component_1, Principal_Component_2, Principal_Component_3, Principal_Component_4
- Cluster and Sub-Cluster assignments

- Normalized and weighted properties: Yield Strength, Modulus of Toughness, Fatigue, Corrosion, Weldability, Anodising
- Multi-criteria optimisation: TOPSIS_Closeness_Score
- Calculated features: Mg₂Si, Q-Phase, Mixing Ratio
- Additional info: Temper, Processing route

Purpose:

- Provides full traceability from raw measurements to PCA, clustering, sub-clustering, and multi-criteria evaluation, recycling, and Thermo-Calc analysis.
- Enables replication of all plots, including PCA plots, elbow plots, silhouette plots, clustering and sub-clustering visualisations, TOPSIS ranking plots, and mixing ratio calculations

Step 6: LIME Feature Contributions

- Per-alloy contributions of Si, Fe, Cu, Mn, Mg, Mg:Si to Yield Strength and Modulus of Toughness.
- Columns added: YS_Si_contrib, YS_Fe_contrib, ... MT_Mg:Si_contrib.
- Shows how each element positively or negatively affects mechanical properties according to the Random Forest models.
- Helps interpret predictions and adds transparency to the dataset without altering raw or standardized values

Step 7: Notes on Reproducibility

- Standardization and PCA applied globally ensure consistent dimensionality reduction and clustering
- Sub-clustering refines groupings for precise alloy selection
- Normalized and weighted properties combined in TOPSIS ensure reproducible multi-criteria optimisation
- Calculated columns (Mg₂Si, Q-Phase, Mixing Ratio) enable reproducible alloy mixing evaluation
- Dataset allows full reproduction of all plots, tables, analyses, and optimised alloy recommendations for 6xxx series alloys

11.3.5 Dataset 5: 7xxx Series Alloys

11.3.5.1 Raw Dataset

The raw data for 7xxx series alloys is provided in **7xxx_Alloys_Raw_Dataset.csv** and an overview is given in Table 11.12. This dataset contains all measured elemental compositions, mechanical properties, and processing information for each alloy grade, including both low-Cu and high-Cu alloys across all temper conditions (T5, T6, T7). It serves as the basis for PCA, clustering, sub-clustering, and multi-criteria optimisation.

Table 11.12. 7xxx Series Raw Dataset Overview

Column	Description
Alloy grade	Unique identifier for each alloy
Si, Fe, Cu, Mn, Mg, Cr, Zn, Ti, Zr	Measured elemental compositions
YTS, Fatigue, UTS, Elongation, Toughness	Mechanical properties
Weldability, Corrosion, Temper	Categorical ratings and processing

Notes:

- This dataset enables reproduction of all preprocessing, PCA, clustering, and sub-clustering analyses.
- Includes all temper conditions (T5, T6, T7) for both low-Cu and high-Cu containing alloys.

11.3.5.2 Processed Dataset – Low-Cu 7xxx Series

The processed dataset for low-Cu 7xxx series alloys (**7xxx_Low_Cu_Processed_All_Temper_Data.csv**) consolidates standardized, PCA, clustering, sub-clustering, normalized, weighted, and multi-criteria computed features (as shown in Table 11.13). Simplified processed datasets for T5 and T7 conditions are also provided for clustering reproducibility.

Table 11.13. 7xxx Low-Cu Processed Dataset Overview

Component	Columns	Description
Standardized Inputs	Zn_Standardized, Cu_Standardized, Mg_Standardized, Fe_Standardized, Mn_Standardized, Si_Standardized, Cr_Standardized, Ti_Standardized, Zr_Standardized, YTS_Standardized, Fatigue_Standardized, Toughness_Standardized	Standardized elemental compositions and mechanical properties (mean=0, std=1) for PCA and clustering reproducibility

PCA Components	Principal_Component_1 ... Principal_Component_4	Principal components calculated from standardized inputs for dimensionality reduction
Cluster Assignments	Cluster, Principal_Component_1_Sub-Clustering, Principal_Component_2_Sub-Clustering, Sub-cluster	Initial clusters and sub-clusters based on PCA for detailed grouping
Normalized & Weighted Properties	YTS_Normalized, Fatigue_Normalized, Toughness_Normalized, Corrosion_Normalized, Weldability_Normalized; YTS_Weighted, Fatigue_Weighted, Toughness_Weighted, Corrosion_Weighted, Weldability_Weighted	Normalized and weighted performance metrics per alloy
Multi-Criteria Optimisation	TOPSIS_Score	Multi-criteria ranking of alloys considering all weighted properties
Mixing Ratio	Mixing Ratio	Normalized mixing ratio for recycling evaluation
LIME Contributions	YS_Zn_contrib, YS_Cu_contrib, YS_Mg_contrib, YS_Fe_contrib, YS_Mn_contrib, YS_Si_contrib, YS_Cr_contrib, YS_Ti_contrib, YS_Zr_contrib, MT_Zn_contrib, MT_Cu_contrib, MT_Mg_contrib, MT_Fe_contrib, MT_Mn_contrib, MT_Si_contrib, MT_Cr_contrib, MT_Ti_contrib, MT_Zr_contrib	Estimated per-element contributions to Yield Strength (YS) and Toughness (MT) from LIME regression analysis
Additional Info	Temper	Alloy temper condition

Notes:

- Standardization and PCA are applied globally to ensure reproducibility of clustering, sub-clustering, and visualisation.
- Sub-clustering is performed on PCA components to refine groupings.
- Normalized and weighted properties are combined in TOPSIS to rank alloy performance across mechanical and processing attributes.
- Mixing ratios provide recycling and alloy mixing insights.
- Full T6 conditions for low-Cu alloys are included; T5 and T7 conditions are also provided in datasets for clustering reference.
- LIME contributions are provided for metallurgical reasoning
- Enables full reproduction of PCA, clustering, sub-clustering, and multi-criteria evaluation, recycling workflows.

11.3.5.3 Processed Dataset – High-Cu 7xxx Series

The processed dataset for high-Cu 7xxx series alloys (**7xxx_High_Cu_Processed_All_Temper_Data.csv**) consolidates standardized, PCA, clustering, sub-clustering, normalized, weighted, and multi-criteria computed features as shown in Table 11.14. Simplified datasets for T5 and T7 conditions are also provided.

Table 11.14. 7xxx High-Cu Processed Dataset Overview

Component	Columns	Description
Standardized Inputs	Zn_Standardized, Mg_Standardized, Cu_Standardized, Fe_Standardized, Mn_Standardized, Si_Standardized, Cr_Standardized, Ti_Standardized, Zr_Standardized, YTS_Standardized, Fatigue_Standardized, Toughness_Standardized	Standardized elemental compositions and mechanical properties (mean=0, std=1) for PCA and clustering reproducibility
PCA Components	Principal_Component_1 ... Principal_Component_4	Principal components calculated from standardized inputs for dimensionality reduction
Cluster Assignments	Cluster, Sub-cluster	Initial clusters and sub-clusters based on PCA for detailed grouping
Normalized & Weighted Properties	YTS_Normalized, Fatigue_Normalized, Toughness_Normalized, Corrosion_Normalized, Weldability_Normalized; YTS_Weighted, Fatigue_Weighted, Toughness_Weighted, Corrosion_Weighted, Weldability_Weighted	Normalized and weighted performance metrics per alloy
Multi-Criteria Optimization	TOPSIS_Score	Multi-criteria ranking of alloys considering all weighted properties
Mixing Ratio	Mixing Ratio	Normalized mixing ratio for recycling evaluation
LIME Contributions	YS_Zn_contrib, YS_Cu_contrib, YS_Mg_contrib, YS_Fe_contrib, YS_Mn_contrib, YS_Si_contrib, YS_Cr_contrib, YS_Ti_contrib, YS_Zr_contrib, MT_Zn_contrib, MT_Cu_contrib, MT_Mg_contrib, MT_Fe_contrib, MT_Mn_contrib, MT_Si_contrib, MT_Cr_contrib, MT_Ti_contrib, MT_Zr_contrib	Estimated per-element contributions to Yield Strength (YS) and Toughness (MT) from LIME regression analysis
Additional Info	Temper	Alloy temper condition

Note: This dataset enables full reproduction of PCA, clustering, sub-clustering, and multi-criteria evaluation, and recycling workflows. Also, full reproducibility procedure for both low Cu and high Cu sub-sets is present in **Section 11.3.4.3** by replacing the 6xxx series dataset with 7xxx series dataset.

Additionally, the corresponding codes discussed in below **Section 11.4**, provided alongside the datasets support the full visualisation procedures and analysis algorithms applied to these alloy series.

11.3.6 Supplementary Dataset

All raw and processed datasets supporting this research are provided as supplementary material in CSV format. The file names are listed below:

- **Dataset 1: Chapter 3: Preliminary Analysis**
 - *Preliminary_Analysis_Raw_Dataset.csv*
 - *Preliminary_Analysis_Processed_Dataset.csv*
- **Dataset 2: Chapter 4: Non-Age-Hardenable Series Dataset**
 - *Non_Age_Hardenable_Alloys_Raw.csv*
 - *Non_Age_Hardenable_Alloys_Processed.csv*
- **Dataset 3: Chapter 4: Age-Hardenable Series Dataset**
 - *Age_Hardenable_Alloys_Raw_Data.csv*
 - *Age_Hardenable_Alloys_Processed.csv*
- **Dataset 4: Chapter 5: 6xxx Series Alloys Dataset**
 - *6xxx_Alloys_Raw_Dataset.csv*
 - *6xxx_Alloys_Processed_Dataset.csv*
- **Dataset 5: Chapter 6: 7xxx Series Alloys Dataset**
 - *7xxx_Alloys_Raw_Dataset.csv*
 - *7xxx_High_Cu_Processed_All_Temper_Data.csv*
 - *7xxx_Low_Cu_Processed_All_Temper_Data.csv*

All datasets are provided to ensure transparency, reproducibility, and ease of analysis.

Note: Raw and processed datasets (CSV files) are available from the author upon request.

11.4 Codes for the Project

This appendix contains the complete Python code implementations of the optimisation pipeline used in this research. The codes include PCA & K-means clustering, sub-clustering with TOP-SIS-based optimisation algorithm, recycling algorithm, LIME analysis, Thermo-Calc validation, and recycling software application. As a case study, the examples and demonstrations focus on the 6xxx series aluminium alloys, with replacement parameters for other series also defined.

11.4.1 Optimisation Algorithm

11.4.1.1 Combined Approach of PCA and K-means clustering

```
# Importing libraries

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Separate categorical and non-categorical variables

non_categorical_vars = df_segmentation[['Si', 'Fe', 'Cu', 'Mn', 'Mg', 'Mg:Si', 'Modulus
of Toughness', 'Yeild strength', 'Fatigue']]
categorical_vars = df_segmentation[['Weldability', 'Corrosion', 'Anodising']]

# Map categorical values to numerical scores

mapping = {'Poor': 0, 'Fair': 1, 'Good': 2, 'Excellent': 3}
categorical_vars['Weldability'] = categorical_vars['Weldability'].map(mapping)
categorical_vars['Corrosion'] = categorical_vars['Corrosion'].map(mapping)
categorical_vars['Anodising'] = categorical_vars['Anodising'].map(mapping)

# Standardise the non-categorical variables

scaler = StandardScaler()
df_non_categorical_scaled = pd.DataFrame(
    scaler.fit_transform(non_categorical_vars),
    columns=['Si', 'Fe', 'Cu', 'Mn', 'Mg', 'Mg:Si', 'Toughness', 'Yeild strength', 'Fa-
tigue strength']
)

# Combine scaled numerical and encoded categorical variables

df_segmentation_scaled = pd.concat([df_non_categorical_scaled, categorical_vars],
axis=1, join='inner')

# PCA Analysis

pca = PCA()
pca.fit(df_segmentation_scaled)

# Plot cumulative explained variance

fig = plt.figure(figsize=(10, 8), facecolor='white')
ax = fig.add_subplot(111, facecolor='white')
ax.plot(range(1, len(pca.explained_variance_ratio_) + 1),
```



```

        pca.explained_variance_ratio_.cumsum(), marker='o', linestyle='--', mark-
        ersize=15)
ax.set_title('Explained Variance by Components', fontsize=34)
ax.set_xlabel('Number of Components', fontsize=34)
ax.set_ylabel('Cumulative Explained Variance', fontsize=34)
ax.set_xticks(range(1, len(pca.explained_variance_ratio_) + 1))
ax.tick_params(axis='both', labelsize=28)
ax.grid(True, color='black', linestyle='--', linewidth=1)
for spine in ax.spines.values():
    spine.set_visible(True)
    spine.set_color('black')
    spine.set_linewidth(2)
plt.show()

# Apply PCA with chosen number of components

pca = PCA(n_components=4)
scores_pca = pca.fit_transform(df_segmentation_scaled)

# Elbow Method using K-means

WCSS = []
for i in range(1, 21):
    kmeans_pca = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans_pca.fit(scores_pca)
    WCSS.append(kmeans_pca.inertia_)

fig = plt.figure(figsize=(10, 8), facecolor='white')
ax = fig.add_subplot(111, facecolor='white')
ax.plot(range(1, 21), WCSS, marker='o', linestyle='--', markersize=15)
ax.set_title('K-means with PCA Clustering', fontsize=34)
ax.set_xlabel('Number of Clusters', fontsize=34)
ax.set_ylabel('WCSS', fontsize=34)
ax.set_xticks(np.arange(0, 22.5, 2.5))
ax.tick_params(axis='both', labelsize=28)
ax.grid(True, color='black', linestyle='--', linewidth=1)
for spine in ax.spines.values():
    spine.set_color('black')
    spine.set_linewidth(2)
plt.show()

# Silhouette Score Analysis

silhouette_scores = []
for k in range(2, 21):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(df_segmentation_scaled)
    score = silhouette_score(df_segmentation_scaled, kmeans.labels_)
    silhouette_scores.append(score)
    print(f"Number of Clusters: {k}, Silhouette Score: {score:.4f}")

plt.figure(figsize=(10, 8))
plt.plot(range(2, 21), silhouette_scores, marker='o', linestyle='--', mark-
        ersize=15)
plt.title('Silhouette Score for k-means Clustering', fontsize=34)
plt.xlabel('Number of Clusters', fontsize=34)
plt.ylabel('Silhouette Score', fontsize=34)
plt.xticks(np.arange(0, 22.5, 2.5), fontsize=28)
plt.yticks(np.arange(0.10, 0.26, 0.02), fontsize=28)
plt.gca().set_facecolor('white')
plt.grid(True, color='black', linestyle='--', linewidth=1)
for spine in plt.gca().spines.values():
    spine.set_color('black')
    spine.set_linewidth(2)
plt.show()

# Final K-means Clustering and Cluster Assignment

```

```

kmeans_pca = KMeans(n_clusters=5, init='k-means++', random_state=42)
y_kmeans = kmeans_pca.fit_predict(scores_pca)

# Create a DataFrame with PCA scores and cluster labels

df_segmpca_kmeans = pd.DataFrame(scores_pca, columns=['Component_1', 'Component_2', 'Component_3'])
df_segmpca_kmeans['Segment_K-means_PCA'] = y_kmeans
df_segmpca_kmeans['Segment'] = df_segmpca_kmeans['Segment_K-means_PCA'].map({0: 'zero', 1: 'one', 2: 'two', 3: 'three', 4: 'four'})

# Visualise Clusters

sns.set_style("whitegrid", {"axes.edgecolor": "black", "axes.linewidth": 2})
plt.figure(figsize=(10, 8))
sns.scatterplot(
    x=df_segmpca_kmeans['Component_2'],
    y=df_segmpca_kmeans['Component_3'],
    hue=df_segmpca_kmeans['Segment'],
    palette=['g', 'r', 'c', 'b', 'y', 'k', 'orange'],
    s=470,
    edgecolor='black'
)
plt.title('Clusters by PCA Components', fontsize=34)
plt.xlabel('Component_1', fontsize=34)
plt.ylabel('Component_2', fontsize=34)
plt.legend(fontsize=17)
plt.xticks(fontsize=28)
plt.yticks(fontsize=28)
for spine in plt.gca().spines.values():
    spine.set_edgecolor('black')
    spine.set_linewidth(2)
plt.grid(False)
plt.show()

```

Note: Repeat the above code for other series by adjusting *n_components* and *k* as per Table 11.13.

Table 11.15. Parameter values for other aluminium alloy series

Alloy Series	n_components	Optimal <i>k</i>
7xxx_Cu	4	4
7xxx_No_Cu	4	3
2xxx	4	6
5xxx	4	3
3xxx	4	4

11.4.1.2 Sub-clustering and Optimisation Algorithm

```

# Importing libraries

from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score

# Normalise data

def normalize_data(data):
    scaler = MinMaxScaler()
    return scaler.fit_transform(data)

```

```

# TOPSIS

def topsis(subcluster_data):
    normalized_data = normalize_data(subcluster_data)
    ideal_solution = np.max(normalized_data, axis=0)
    negative_ideal_solution = np.min(normalized_data, axis=0)

    distance_to_ideal = np.sqrt(np.sum((normalized_data - ideal_solution) ** 2,
axis=1))
    distance_to_negative_ideal = np.sqrt(np.sum((normalized_data - nega-
tive_ideal_solution) ** 2, axis=1))

    closeness = distance_to_negative_ideal / (distance_to_ideal + distance_to_neg-
ative_ideal)
    return closeness

# Find optimum alloy

def find_optimum_alloy(subcluster_data):
    closeness_scores = topsis(subcluster_data)
    optimum_index = np.argmax(closeness_scores)
    optimum_alloy = subcluster_data[optimum_index]
    return optimum_alloy, optimum_index, closeness_scores

# Plot clusters

def plot_clusters(data, labels, cluster_number):
    plt.figure(figsize=(8, 6))
    data = np.array(data)
    plt.scatter(data[:, 0], data[:, 1], c=labels, cmap='viridis', marker='o')
    plt.title(f'Cluster {cluster_number} - Subclusters Plot')
    plt.xlabel('Si')
    plt.ylabel('Fe')
    plt.colorbar(label='Subcluster Label')
    plt.show()

# Perform K-Means for sub-clustering

def perform_clustering(cluster_data, num_subclusters):
    if len(cluster_data) < 3:
        print("Only one data point, skipping clustering.")
        return [cluster_data], [0] * len(cluster_data)

    kmeans = KMeans(n_clusters=num_subclusters, random_state=0)
    subcluster_labels = kmeans.fit_predict(cluster_data)

    silhouette_avg = silhouette_score(cluster_data, subcluster_labels)
    print(f"Silhouette Score for {num_subclusters} subclusters: {silhouette_avg}")

    subclusters = [[] for _ in range(num_subclusters)]
    for i, label in enumerate(subcluster_labels):
        subclusters[label].append(cluster_data[i])

    return subclusters, subcluster_labels

# Automatically decide sub-cluster number (Mathematical model)

def decide_num_subclusters(cluster_size):
    if cluster_size < 3:
        return 1
    elif cluster_size <= 10:
        return 2
    elif cluster_size <= 20:
        return 3
    elif cluster_size <= 30:
        return 4
    else:

```

```

        return 5

# Load and Preprocess Data

with open('6xxx_.csv', newline='') as csvfile:
    data = list(csv.DictReader(csvfile))

def convert_to_mpa(toughness_pa):
    return float(toughness_pa) # Already in MPa

rating_dict = {'Poor': 0, 'Fair': 1, 'Good': 2, 'Excellent': 3}

# Convert required fields

for row in data:
    row['Toughness'] = convert_to_mpa(float(row['Toughness']))
    row['Corrosion'] = rating_dict[row['Corrosion']]
    row['Weldability'] = rating_dict[row['Weldability']]
    row['Anodising'] = rating_dict[row['Anodising']]
    for key in row.keys():
        if key not in ['Alloy grade', 'Cluster', 'Corrosion', 'Weldability', 'Ano-
dising']:
            row[key] = float(row[key])

# Group Data by Clusters

clusters = {}
for i in range(4): # Assumes 4 main clusters
    clusters[i] = [row for row in data if int(row['Cluster']) == i]

# Apply Sub-clustering and Optimum Alloy Selection

for cluster_number, cluster_data in clusters.items():
    cluster_data_for_clustering = [
        [row['YTS'], row['Fatigue strength'], row['Toughness'], row['Corrosion'],
row['Weldability']]
        for row in cluster_data
    ]

    num_subclusters = decide_num_subclusters(len(cluster_data))
    subclusters, subcluster_labels = perform_clustering(cluster_data_for_clustering, num_subclusters)
    plot_clusters(cluster_data_for_clustering, subcluster_labels, cluster_number)

    for subcluster_number, subcluster_data in enumerate(subclusters):
        print(f"\nCluster {cluster_number}, Subcluster {subcluster_number}:")
        if subcluster_data:
            closeness_scores = topsis(subcluster_data)
            for alloy, score in zip(subcluster_data, closeness_scores):
                print(f"{alloy} - Closeness Score: {score:.4f}")
            optimum_alloy, optimum_index, _ = find_optimum_alloy(subcluster_data)
            print(f"Optimum Alloy: {optimum_alloy}")
            print(f"Optimum Alloy Index: {optimum_index}")
        else:
            print("No data in this subcluster.")

```

Note: Same Sub-clustering and optimum alloy selection algorithm applies to other series i.e., 7xxx (with Cu and without Cu), 2xxx, 3xxx and 5xxx series alloys.

11.4.1.3 Recycling Algorithm

```

# Alloy data (e.g., sub-cluster 01 of cluster 0 of 6xxx series)

```

```

alloys = [
    {'Si': 0.6, 'Fe': 0.35, 'Cu': 0.5, 'Mn': 0.15, 'Mg': 1.0,
     'Yield strength': 296.0, 'Modulus of Toughness': 46.5, 'Fatigue': 119.3,
     'Corrosion': 'Excellent', 'Weldability': 'Excellent', 'Anodising': 'Fair'},

    {'Si': 0.55, 'Fe': 0.2, 'Cu': 0.275, 'Mn': 0.275, 'Mg': 0.85,
     'Yield strength': 260.0, 'Modulus of Toughness': 25.2, 'Fatigue': 100.0,
     'Corrosion': 'Good', 'Weldability': 'Excellent', 'Anodising': 'Good'},

    {'Si': 0.6, 'Fe': 0.25, 'Cu': 0.2, 'Mn': 0.1, 'Mg': 0.9,
     'Yield strength': 270.0, 'Modulus of Toughness': 25.2, 'Fatigue': 110.0,
     'Corrosion': 'Good', 'Weldability': 'Good', 'Anodising': 'Poor'},

    {'Si': 0.6, 'Fe': 0.35, 'Cu': 0.275, 'Mn': 0.15, 'Mg': 1.0,
     'Yield strength': 270.0, 'Modulus of Toughness': 29.5, 'Fatigue': 110.0,
     'Corrosion': 'Good', 'Weldability': 'Excellent', 'Anodising': 'Good'},

    {'Si': 0.6, 'Fe': 0.37, 'Cu': 0.275, 'Mn': 0.15, 'Mg': 1.0,
     'Yield strength': 270.0, 'Modulus of Toughness': 31.9, 'Fatigue': 110.0,
     'Corrosion': 'Good', 'Weldability': 'Excellent', 'Anodising': 'Excellent'},
]

# Corresponding closeness scores for each alloy

closeness_scores = [
    0.7722700008035586,
    0.37215467497755456,
    0.21237410773401758,
    0.46104307602295397,
    0.5118384307334206
]

# Identify the optimum alloy with the highest closeness score

optimum_alloy_index = np.argmax(closeness_scores)
print(f"Optimum Alloy Index: {optimum_alloy_index}")
print(f"Optimum Alloy (Based on Closeness Score): {alloys[optimum_alloy_index]}")

# Convert qualitative ratings to numerical values

rating_dict = {'Poor': 0, 'Fair': 1, 'Good': 2, 'Excellent': 3}
for alloy in alloys:
    alloy['Corrosion'] = rating_dict[alloy['Corrosion']]
    alloy['Weldability'] = rating_dict[alloy['Weldability']]
    alloy['Anodising'] = rating_dict[alloy['Anodising']]

# Fix atleast 50% mixing for the optimum alloy (Present case its 95%)

mixing_ratios = np.zeros(len(alloys))
mixing_ratios[optimum_alloy_index] = 0.95

# Distribute the remaining 50% (Here 5%) based on relative closeness scores

remaining_ratio = 0.05
for i in range(len(alloys)):
    if i != optimum_alloy_index:
        mixing_ratios[i] = (closeness_scores[i] / np.sum(closeness_scores)) * remaining_ratio

# Normalise to ensure the ratios sum exactly to 1

mixing_ratios /= np.sum(mixing_ratios)
# Display the mixing ratios

print("\nOptimal Mixing Ratios (sum = 1):")
for i, ratio in enumerate(mixing_ratios):
    print(f"Alloy {i} Mixing Ratio: {ratio:.4f}")
# Step 5: Compute the mixed alloy's final composition

```

```

mixed_alloy = {key: 0 for key in alloys[0].keys()}
for i, alloy in enumerate(alloys):
    for key, value in alloy.items():
        mixed_alloy[key] += value * mixing_ratios[i]

# Output the final blended alloy composition

print("\nResulting Mixed Alloy Composition:")
for key, value in mixed_alloy.items():
    print(f"{key}: {value:.4f}")

```

Note: This approach can be extended to other sub-clusters of 6xxx series aluminium alloys and to entirely different series (e.g., 2xxx, 3xxx, 5xxx, 7xxx) by simply replacing the alloy data and closeness scores generated using **Section 11.3.1.2**.

11.4.2 Metallurgical Reasoning behind Clustering

11.4.2.1 LIME Analysis on Clusters

```

# Import libraries

import lime.lime_tabular
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

# Load dataset (Here cluster 0)

RR = pd.read_csv('6xxx_Cluster_0.csv')
X = RR[['Si', 'Fe', 'Cu', 'Mn', 'Mg']]
y = RR['YS']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
model = RandomForestRegressor(max_depth=6, n_estimators=10, random_state=0)
model.fit(X_train, y_train)

# Predictions

y_pred = model.predict(X_test)
rmse = mean_squared_error(y_test, y_pred, squared=False)
print(f"RMSE: {rmse:.4f}")

# LIME explanation

explainer = lime.lime_tabular.LimeTabularExplainer(
    X_train.values,
    feature_names=X_train.columns.tolist(),
    verbose=True,
    mode='regression'
)

# Explain a single instance (index 1)

j = 1
exp = explainer.explain_instance(X_test.values[j], model.predict, num_features=9)
exp.show_in_notebook(show_table=True)

```

Note: Replace with dataset for various clusters and other series of aluminium alloys

11.4.2.2 Thermo-Calc Validation of Clusters

```
# Load dataset

df = pd.read_csv('6xxx_.csv')

# Group by Cluster and calculate mean values for selected phases and yield strength

cluster_summary = df.groupby('Cluster').agg({
    'Mg2Si': 'mean',
    'Q-phase': 'mean',
    'Yield strength': 'mean'
}).reset_index()

# Sort clusters by increasing yield strength

cluster_summary = cluster_summary.sort_values(by='Yield strength').reset_index(drop=True)

# Plot

fig, ax1 = plt.subplots(figsize=(12, 6))
ax1.grid(False)
# Define positions for grouped bars
x = np.arange(len(cluster_summary))
bar_width = 0.35

# Plot the phase volume fractions as grouped bars

bars_mg2si = ax1.bar(x - bar_width / 2, cluster_summary['Mg2Si'], bar_width, label='Mg2Si', color='lightblue')
bars_qphase = ax1.bar(x + bar_width / 2, cluster_summary['Q-phase'], bar_width, label='Q-phase', color='lightgreen')
ax1.set_xlabel('Cluster (Sorted by Yield Strength)', fontsize=21)
ax1.set_ylabel('Phase Volume Fraction', fontsize=21)
ax1.set_title('Cluster-wise Phase Volume Fractions vs Yield Strength', fontsize=21)
ax1.set_xticks(x)
ax1.set_xticklabels(cluster_summary['Cluster'], rotation=45, fontsize=20)
ax1.tick_params(axis='y', labelsize=20)

# Add secondary y-axis for yield strength

ax2 = ax1.twinx()
ax2.grid(False)
ax2.plot(x, cluster_summary['Yield strength'], label='Yield Strength', color='red', marker='o', linestyle='-', linewidth=2)
ax2.set_ylabel('Yield Strength (MPa)', color='red', fontsize=20)
ax2.tick_params(axis='y', labelsize=20)

ax1.legend(loc='upper left', fontsize=15)
ax2.legend(loc='upper right', fontsize=15)
plt.tight_layout()
plt.show()
```

Note: This code can be reused for other aluminium alloy series (e.g., 2xxx, 5xxx, 3xxx) by replacing the phase names in the .agg() function with the phase names shown in Table 11.14. The rest of the plotting logic remains unchanged.

Table 11.16. Phase formation in aluminium alloy series

Alloy Series	Phases to Include in groupby().agg()
2xxx	'Al2Cu_C': 'mean', 'S phase': 'mean', 'Q_ALCUMGSI': 'mean', 'AL31MN6NI2': 'mean', 'AL7CU4NI': 'mean', 'Yeild strength': 'mean'
5xxx	'AL18MG3TM2': 'mean', 'AL6MN': 'mean', 'MG2SI_C1': 'mean', 'T_PHASE': 'mean', 'ALMG_BETA': 'mean', 'Yeild strength': 'mean'
3xxx	'AL15SI2M4': 'mean', 'AL6MN': 'mean', 'Mg2Si': 'mean', 'Yeild strength': 'mean'

11.4.3 Recycling Software Application Code for Sorted 6xxx series alloys

```
# Import libraries

import streamlit as st
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

# Load Dataset

data = pd.read_csv("6xxx_.csv")

# Train Linear Regression Model

features = data[['Cu', 'Mg', 'Mn', 'Si', 'Fe']]
targets = data[['Yield Strength', 'Toughness', 'Fatigue Strength']]
model = LinearRegression()
model.fit(features, targets)

# Prediction Function

def predict_properties(composition):
    composition_df = pd.DataFrame([composition])
    predicted_properties = model.predict(composition_df)
    return predicted_properties[0]

# Normalise Data

def normalize_matrix(properties_matrix):
    min_vals = np.min(properties_matrix, axis=0)
    max_vals = np.max(properties_matrix, axis=0)
    return (properties_matrix - min_vals) / (max_vals - min_vals)

# TOPSIS Calculation

def calculate_topsis_scores(normalized_matrix):
    ideal_best = np.max(normalized_matrix, axis=0)
    ideal_worst = np.min(normalized_matrix, axis=0)
    distances_best = np.linalg.norm(normalized_matrix - ideal_best, axis=1)
    distances_worst = np.linalg.norm(normalized_matrix - ideal_worst, axis=1)
    return distances_worst / (distances_best + distances_worst)

# Recycling Algorithm

def mix_alloys(user_alloys):
    properties = ['Avg Typical Yield Strength', 'Avg Typical Toughness', 'Avg Typical Fatigue Strength']
    alloy_properties = np.array([[alloy[prop] for prop in properties] for alloy in user_alloys])

    normalized_matrix = normalize_matrix(alloy_properties)
    topsis_scores = calculate_topsis_scores(normalized_matrix)
    max_topsis_index = np.argmax(topsis_scores)
    # Step 1: Initialise mixing ratio

    mixing_ratios = np.zeros(len(user_alloys))
```



```

    mixing_ratios[max_topsis_index] = 0.5

    # Step 2: Distribute remaining 50% among other alloys

    remaining_ratio = 0.5
    other_scores = [topsis_scores[i] for i in range(len(user_alloys)) if i !=
max_topsis_index]
    total_other_scores = np.sum(other_scores)

    for i in range(len(user_alloys)):
        if i != max_topsis_index:
            mixing_ratios[i] = (topsis_scores[i] / total_other_scores) * remain-
ing_ratio

    mixing_ratios /= np.sum(mixing_ratios)

    # Step 3: Compute weighted composition

    mixed_alloy_composition = {key: 0 for key in user_alloys[0]['Composi-
tion'].keys()}
    for i, alloy in enumerate(user_alloys):
        for key, value in alloy['Composition'].items():
            mixed_alloy_composition[key] += value * mixing_ratios[i]

    recycled_properties = {
        'Avg Typical Yield Strength': sum(alloy['Avg Typical Yield Strength'] *
mixing_ratios[i] for i, alloy in enumerate(user_alloys)),
        'Avg Typical Toughness': sum(alloy['Avg Typical Toughness'] * mixing_ra-
tios[i] for i, alloy in enumerate(user_alloys)),
        'Avg Typical Fatigue Strength': sum(alloy['Avg Typical Fatigue Strength'] *
mixing_ratios[i] for i, alloy in enumerate(user_alloys)),
    }

    return mixed_alloy_composition, recycled_properties, mixing_ratios

# Find Similar Alloy

def find_similar_alloy(composition, temper_condition):
    composition_series = pd.Series(composition)

    # Filter by both tempering and processing conditions
    filtered_data = data[
        (data['Tempering'] == temper_condition) &
        (data['Processing'] == processing_condition)
    ]

    if filtered_data.empty:
        return None

    distances = filtered_data[['Cu', 'Mg', 'Mn', 'Si', 'Fe']].apply(
        lambda row: np.linalg.norm(row - composition_series), axis=1)

    if distances.empty:
        return None

    closest_index = distances.idxmin()
    closest_distance = distances.loc[closest_index]
    if closest_distance > 0.5:
        return None

    return filtered_data.loc[closest_index]

# Streamlit UI

st.markdown(
    """
    <style>
    .stApp {

```

```

        background-color: #f0f5f9;
        text-align: center;
    }
</style>
"""
unsafe_allow_html=True
)
st.title(":recycle: Alloy Recycling Software")
st.write("Welcome to the Alloy Recycling Application. Recycle your alloys and create new compositions!")
st.subheader("Recycle 6XXX Series Alloys")
num_alloys_6XXX = st.number_input("Enter the number of 6XXX series alloys for recycling:", min_value=1, value=1, step=1)

alloy_data_6XXX = []
for i in range(num_alloys_6XXX):
    st.header(f"Alloy 6XXX {i+1}")
    use_typical_values = st.checkbox(f"Use average typical values for Alloy 6XXX {i+1}", key=f"typical_6XXX_{i}")

    if use_typical_values:
        yield_strength = st.number_input(f"Average Typical Yield Strength (MPa):", min_value=0.0, key=f"ys_typical_{i}")
        toughness = st.number_input(f"Average Typical Toughness (Pa):", min_value=0.0, key=f"t_typical_{i}")
        fatigue_strength = st.number_input(f"Average Typical Fatigue Strength (MPa):", min_value=0.0, key=f"fs_typical_{i}")
    else:
        cols = st.columns(2)
        with cols[0]:
            ys_min = st.number_input("Min Yield Strength (MPa):", min_value=0.0, key=f"ys_min_{i}")
            t_min = st.number_input("Min Toughness (Pa):", min_value=0.0, key=f"t_min_{i}")
            fs_min = st.number_input("Min Fatigue Strength (MPa):", min_value=0.0, key=f"fs_min_{i}")
        with cols[1]:
            ys_max = st.number_input("Max Yield Strength (MPa):", min_value=ys_min, key=f"ys_max_{i}")
            t_max = st.number_input("Max Toughness (Pa):", min_value=t_min, key=f"t_max_{i}")
            fs_max = st.number_input("Max Fatigue Strength (MPa):", min_value=fs_min, key=f"fs_max_{i}")
        yield_strength = (ys_min + ys_max) / 2
        toughness = (t_min + t_max) / 2
        fatigue_strength = (fs_min + fs_max) / 2

# Alloy composition
composition = {}
for element in ["Si", "Cu", "Mg", "Mn", "Fe"]:
    use_avg = st.checkbox(f"Use average composition for {element}:", key=f"avg_{element}_{i}")
    if use_avg:
        composition[element] = st.number_input(f"Avg Composition of {element} (%)", min_value=0.0, max_value={"Si": 1.5, "Fe": 0.5, "Cu": 1.0, "Mn": 1.0, "Mg": 1.5}[element], key=f"{element}_avg_{i}")
    else:
        min_val = st.number_input(f"Min Composition of {element} (%)", min_value=0.0, key=f"{element}_min_{i}")
        max_val = st.number_input(f"Max Composition of {element} (%)", min_value=min_val, key=f"{element}_max_{i}")
        composition[element] = (min_val + max_val) / 2

alloy_data_6XXX.append({
    "Avg Typical Yield Strength": yield_strength,
    "Avg Typical Toughness": toughness,

```

```

        "Avg Typical Fatigue Strength": fatigue_strength,
        "Composition": composition
    })

temper_condition_6XXX = st.selectbox(
    "Select Tempering Condition for 6XXX Series Alloy:",
    ["T5", "T6", "T7"]
)

processing_condition_6XXX = st.selectbox(
    "Select Processing Condition for 6XXX Series Alloy:",
    ["Extrusion", "Forging", "Plate", "Rolled", "Sheet"]
)

if st.button("Recycle 6XXX Alloys"):
    recommended_alloy_6XXX, recycled_properties_6XXX, mixing_ratios_6XXX = mix_al-
    loys(alloy_data_6XXX)

    st.subheader("Recycled Alloy Composition:")
    for key, value in recommended_alloy_6XXX.items():
        st.write(f"{key}: {value:.2f} %")

    st.subheader("Mixing Ratios for Alloys in the Recycling Process:")
    for i, ratio in enumerate(mixing_ratios_6XXX):
        st.write(f"Alloy {i+1} - Mixing Ratio: {ratio:.4f}")

    st.subheader("Recycled Alloy Properties:")
    for property_name, value in recycled_properties_6XXX.items():
        unit = "MPa" if "Strength" in property_name else "Pa"
        st.write(f"{property_name}: {value:.2f} {unit}")

    similar_alloy = find_similar_alloy(recommended_alloy_6XXX, temper_condi-
    tion_6XXX)
    st.subheader("Similar Alloy:")
    if similar_alloy is not None:
        st.write(f"Similar Alloy Grade: {similar_alloy['Alloy grade']}")
    else:
        st.write("No similar alloy found.")

```

11.4.4 Extended Recycling Software Application Code for Sorting and Recycling of 2xxx, 3xxx, 5xxx, 6xxx, 7xxx, 8xxx

```

# NEW FEATURES FOR ALL-SERIES MIXING

# Import Libraries

import os
os.environ["CUDA_VISIBLE_DEVICES"] = "-1" # Force CPU

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
import tensorflow as tf
import plotly.graph_objects as go

# Hide GPU completely

tf.config.set_visible_devices([], 'GPU') (# Optional depending on GPU availability)

# Features and Targets (Extended)

```

```

features = ["Si", "Fe", "Cu", "Mn", "Mg", "Cr", "Zn", "Ti", "Ag", "Li", "Ni", "Zr", "Ga", "V"]
targets = ["Yield Strength", "Fatigue Strength", "Toughness", "Weldability", "Corrosion", "Electrical", "Thermal"]

# Encode categorical targets

mapping = {'Poor':0, 'Fair':1, 'Good':2, 'Excellent':3}
for col in ["Weldability", "Corrosion"]:
    if col in data.columns:
        data[col] = data[col].map(mapping)

# ANN Model for property prediction

# ANN Model for property prediction (single ANN version)
ann_model = Sequential([
    Dense(256, input_shape=(len(features),), activation='relu'),
    BatchNormalization(),
    Dropout(0.15),
    Dense(128, activation='relu'),
    BatchNormalization(),
    Dropout(0.15),
    Dense(64, activation='relu'),
    Dense(32, activation='relu'),
    Dense(len(targets), activation='linear')
])

# Compile model

ann_model.compile(optimizer=Adam(learning_rate=0.0005), loss='mse')

# Early stopping callback

early_stop = EarlyStopping(monitor='val_loss', patience=40, restore_best_weights=True)
lr_schedule = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=15, min_lr=1e-6)

# Train ANN

history = ann_model.fit(
    X_train, y_train,
    validation_data=(X_test, y_test),
    epochs=500,
    batch_size=16,
    verbose=0,
    callbacks=[early_stop, lr_schedule]
)

# Predict

y_train_pred_scaled = ann_model.predict(X_train, verbose=0)
y_test_pred_scaled = ann_model.predict(X_test, verbose=0)

# Inverse scaling

y_train_pred = y_scaler.inverse_transform(y_train_pred_scaled)
y_train_true = y_scaler.inverse_transform(y_train)
y_test_pred = y_scaler.inverse_transform(y_test_pred_scaled)
y_test_true = y_scaler.inverse_transform(y_test)

# Large Batch Handling

# PCA + K-means clustering to handle >6 alloys

comp_matrix = np.array([list(a["Composition"].values()) for a in user_alloys],
dtype=np.float32)
scaler = StandardScaler()

```

```

comp_scaled = scaler.fit_transform(comp_matrix)
pca = PCA(n_components=min(3, comp_scaled.shape[1]))
comp_pca = pca.fit_transform(comp_scaled)

# Optimal clusters selection using silhouette score

best_score = -1
optimal_k = 2
max_k = min(10, num_alloys - 1)
for k in range(2, max_k + 1):
    km = KMeans(n_clusters=k, random_state=42)
    labels = km.fit_predict(comp_pca)
    if len(set(labels)) > 1:
        score = silhouette_score(comp_pca, labels)
        if score > best_score:
            best_score = score
            optimal_k = k

# Process each cluster separately

kmeans = KMeans(n_clusters=optimal_k, random_state=42)
clusters = kmeans.fit_predict(comp_pca)

# Visualisation: Compare recycled alloy vs cluster alloys

fig = go.Figure()
fig.add_trace(go.Bar(
    x=list(mixed_comp.keys()),
    y=list(mixed_comp.values()),
    name="Recycled Alloy",
    marker_color="green"
))
for i, alloy in enumerate(cluster_alloys):
    fig.add_trace(go.Bar(
        x=list(alloy["Composition"].keys()),
        y=list(alloy["Composition"].values()),
        name=f"Cluster {c+1}_Alloy {i+1}",
        opacity=0.6
    ))
fig.update_layout(
    title=f"Cluster {c+1}: Alloy Compositions vs Recycled Alloy",
    xaxis_title="Elements",
    yaxis_title="Composition (%)",
    barmode="group",
    template="plotly_white"
)
st.plotly_chart(fig, use_container_width=True)

```