Journal of Sustainable Metallurgy
https://doi.org/10.1007/s40831-025-01112-4

RESEARCH ARTICLE

Facilitating Recycling of 6xxx Series Aluminum Alloys by Machine Learning-Based Optimization

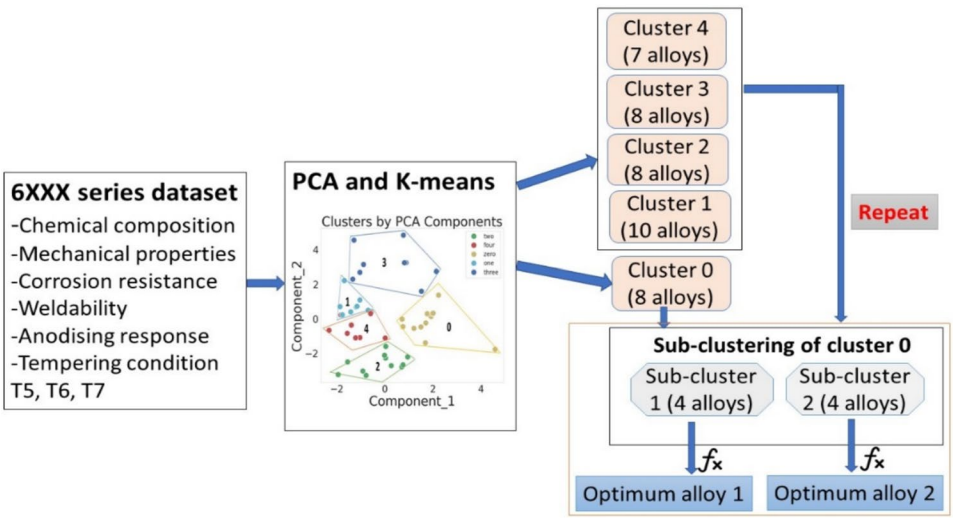
Tanu Tiwari[1] · **Chamini Mendis**[1] · **Dmitry Eskin**[1]

Received: 13 November 2024 / Accepted: 23 April 2025

Abstract

Aluminum alloys throughout the last century have experienced extensive development, owing to their unique strength-to-weight ratio. This led to generating multiple alloy grades. However, large number of grades present challenges when it comes to the recycling of aluminum scrap, which is the current and future trend in aluminum alloy production and application. Therefore, there is an urgent need to decrease the number of alloying grades while preserving their performance. In this study, we designed an optimization loop based on Machine Learning (ML) and material science knowledge for the 292 sets of data collected on 42 grades of 6xxx series aluminum alloys, focusing on their mechanical, service, and technological properties under T5, T6, and T7 tempering conditions. K-means clustering and principal component analysis algorithms were applied to form various clusters of alloys and are further re-clustered into fine sub-clusters. An optimal alloy (OA) for each sub-cluster was identified based on optimization criteria. After successive iteration, we were able to reduce 42 grades of the 6xxx series into a set of 10 OA's each performing optimally. This method not only support the capability of machine learning in selecting OA's but also introduce a future direction for recycling practices in the aluminum industry.

Graphical Abstract



Keywords 6xxx Series aluminum alloys · Machine learning · Alloy optimization · Recyclability

The contributing editor for this article was Zhi Sun.

Extended author information available on the last page of the article

Published online: 12 May 2025

Introduction

Aluminum is the third most common element and the most abundant metal (8%) in the earth's crust. The versatility of aluminum makes it the most widely used structural metal after steel due to its high strength-to-weight ratio, making it easy to design and construct lightweight and sturdy-structures [1].

The mechanical properties of pure aluminum are significantly enhanced by the addition of up to 7% of major alloying elements, such as manganese, copper, silicon, zinc, and magnesium. Furthermore, minor alloying elements (less than 0.5%) are added to further improve its properties. Due to the exceptional combination of properties, aluminum demand is projected to double by the year 2050, leading to the continuous development of new aluminum alloys [2]. At the same time the fraction of recycled (scrap) alloys should increase to at least 50% [3–5].

Traditionally, aluminum alloy design relies on trial and error, driven by the domain knowledge of materials researchers and the current requirements of manufacturers, as well as proprietary considerations. This is done by varying alloying element concentrations and processing conditions to improve mechanical properties. However, this approach is extremely time- and cost-intensive and does not cover the vast design space of potential alloys.

Materials scientists, in discovering new alloys, often rely on the thermodynamic information presented by phase diagrams. However, the relationship between changes in single input variables and the target property often cannot be interpreted by a human. Recently,'ab initio'methods have been used to discover alloys, involving structural calculations from scratch. However, this approach cannot be generalized for all alloy design issues due to the limitations of the method, a number of assumptions, nonlinearity, and the high dimensionality of alloy property variations with composition [6]. In addition, ab initio approaches do not fully harness the information of alloys that are already known. The enormous complexity due to the interplay of structural, chemical, and microstructural degrees of freedom makes the rational design of materials with targeted properties rather difficult. Even with a wide potential solution space, rapid testing and fabrication will not guarantee alloys with the desired properties. Therefore, there is an urgent need to narrow down the solution space [7].

In recent years machine learning (ML) entered in the field of material science to address complex materials science problems. Due to its low computational cost and better prediction performance, ML techniques play important role in discovery of new materials, material analysis and material design. Several studies have been done to develop new alloys with superior characteristics [8]. Li et al. explored 7xxx series aluminum alloys using ML-based composition and process optimization. They identified a lean composition optimum alloy having superior ultimate tensile strength under T6 tempering condition compared to other 7xxx series alloys [9]. Xue et al. developed an ML-model and an adaptive learning strategy for high-property shape memory alloys [10]. Raccuglia et al. used an ML strategy to model the synthesis laws from failed experiments in order to design new materials [11]. Devi et al. predicted mechanical properties of aluminum alloys with three ML methods including linear regression (LR), K-nearest neighbor (KNN), and artificial neural network (ANN) [12].

Similar, ML approaches have been applied composition design of piezoelectric materials [13, 14], superconducting materials [15], stainless steel [16], and high-entropy alloys [17], as well as to structure and property predictions such as diffusion [18–20], lattice misfit [21], and fatigue [22, 23].

These studies are largely focused on developing new materials with improved properties, rather than optimizing the existing alloy grades. Therefore, they often overlook the challenge of having a large range of alloy grades, which could exacerbate recycling issues in the long run.

Some initiatives have been taken to reduce the solution space of aluminum alloys to a few best-performing alloys; however, the number of alloys keeps increasing with further alloy development [24–26]. The main challenge in recycling wrought aluminum alloys is maintaining the correct chemical composition while minimizing the addition of primary aluminum and alloying elements. In mixed scrap, excessive concentrations of critical elements (Fe, Cu, Mn, Mg, Zn, Si) often necessitate dilution with primary aluminum, making careful scrap sorting essential for direct reuse. Understanding compositional tolerance limits is crucial, as impurities can accumulate and impact alloy properties, emphasizing the need for optimized compositions [27–29].

Additionally, the wide range of current standard alloys makes the recycling challenging and needs to be narrowed down. Fewer alloy grades make it easier to sort the scrap and allow the producers and users to mix different alloys upon recycling without the loss of quality. Therefore, there is an urgent need to optimize aluminum alloy compositions while minimizing the number of alloy grades in use.

In our previous work [30], we attempted to narrow down the solution space of the 6xxx series dataset, consisted of tensile properties and alloying compositions in the T6 tempering condition, using a combined approach of K-means clustering and principal component analysis (PCA). We divided the dataset into five clear clusters consisting similar properties within them. Furthermore, we used the Local Interpretable Model-Agnostic Explanations (LIME)

algorithm to explain the clusters and suggested a metallurgical reasoning behind clustering.

In this study, we expanded our analysis of the 6xxx series aluminum alloys by developing a comprehensive design framework for optimal alloy selection. We have expanded our dataset to include fatigue strength, technological characteristics, corrosion resistance (CR), and anodizing response, in addition to the already existing tensile properties. We have also expanded the dataset to include the T5 and T7 tempering conditions, whereas previously, all the datasets were in the T6 tempering condition. This work not only streamlines the alloy selection process but also provides the metallurgical reasoning behind the clustering and optimal alloy selection process. The main focus of this work is to develop a universal methodology for reducing the number of alloys grades through optimal alloy selection, marking a significant step forward in aluminum alloy optimization and recycling practices. This methodology enables faster identification of alloys that not only meet but exceed performance expectations. While rapid fabrication and testing are valuable, they are less effective when the solution space is vast. Narrowing this space increases the likelihood of selecting alloys with the desired properties.

The 6xxx alloy series was selected as an example, as these alloys are widely used (mostly as extrusions) in construction and automotive industries. The developed methodology, however, can be applied to other alloying series as well as to an expanded range of process conditions.

Ultimately, this work contributes to a more sustainable and efficient future for the aluminum industry by optimizing alloy selection and minimizing resources spent on trial and error. The process of selecting and narrowing down optimized alloys directly enhances the recycling efficiency. By reducing the number of alloys in circulation while still meeting necessary application requirements, we facilitate better sorting, minimize contamination, and improve the performance of secondary aluminum. This approach aligns with sustainable material practices and supports a circular economy.

Materials and Methods

Framework

A systematic design framework for optimal alloy selection within the 6xxx series aluminum alloys is shown in Fig. 1. It consists of the following key stages: data collection → combined PCA and K-means clustering → sub-clustering → optimum alloy selection algorithm. Initially, dataset collection was done by collecting chemical compositions, mechanical properties, technological and service properties at T5, T6, and T7 tempering conditions. Following data collection, clustering was performed. Once clear clusters containing similar range of properties within them were formed and supported by metallurgical reasoning, the dataset was sub-clustered to get more compact sub-clusters, from which an optimum alloy was selected in the next step. An optimal alloy selection algorithm was employed to select the alloys with optimal properties within each sub-cluster. This framework enabled the design and selection of the most promising 6xxx series aluminum alloys with the optimal set of properties which not only met but exceed the required performance criteria.

Data Collection and Processing

In this study, we selected commercially available ASTM standard 6xxx series wrought aluminum alloys to construct database using Encyclopaedia of Aluminum and its Alloys [31], ASM Specialty Handbook [2], MakeItFrom website [32], and various literature sources. This series of wrought

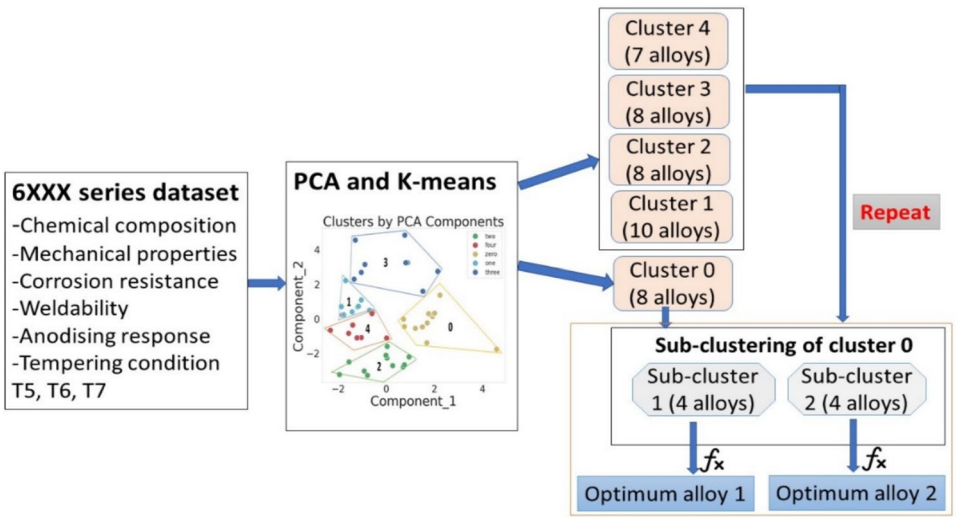


Fig. 1 A design framework for optimum alloy selection

aluminum alloys is widely used in automotive and construction industries, hence need to be highly recyclable. It covers data collected on chemical compositions, mechanical properties, technological and service properties, and processing of 6xxx series aluminum alloys at T5, T6, and T7 tempering conditions to capture the broader range of properties at various stages of artificial aging. The dataset was tabulated into a CSV format. All the implementations of ML algorithms were done in Anaconda's Jupyter Notebook environment [33]. Data visualization and plotting of all graphs were performed with the help of the Matplotlib and Seaborn libraries [34, 35]. The above setup allowed running machine learning algorithms quite smoothly. Unlike previous studies where the database was comprised mostly of numerical features, our database contained numerical as well as categorical features, i.e., weldability, corrosion resistance, and anodizing response. These three properties are important for 6xxx series alloys that are used in construction and automotive industries. At the preprocessing stage, categorical variables were converted into numerical representations to enable their integration into the modeling process. These variables were then concatenated with the already existing numerical features (i.e., chemical composition and mechanical properties), by the use of a joining function.

PCA and K-Means Clustering

K-means clustering and principal component analysis algorithms were applied in grouping the 6xxx series dataset into various clusters containing alloys with similar ranges of composition and properties. The unsupervised machine learning algorithm, K-means clustering, separates the dataset into a number of clusters by assigning data points to the nearest cluster centroid. On each iteration, data points that share similar features get assigned to the same centroid [36].

Principal component analysis is a technique of reducing dataset dimensionality to a lower dimensionality space such that most of the dataset information is preserved by lesser number of principal components, which is determined by the PCA variance plot [37].

The combined approach of PCA and K-means clustering involved the following steps [38]:

Step 1: Standardize the dataset—Normalize variable scales if they are measured in different values scales.

Step 2: PCA—Use PCA after the dataset standardization to reduce the dataset dimensionality and identify the number of principal components.

Step 3: Select principal components—By the amount of variation, through a variance plot, determine how many principle components to retain.

Step 4: Dataset Transformation—Apply selected principal components to the original dataset.

Step 5: Decide the number of clusters—Apply an elbow method to the transformed dataset.

Step 6: Perform K-means clustering—Assign each data point to the nearest cluster centroid until there was no further reassignment of the data points.

Step 7: Evaluate clustering quality—Assess the clustering quality using Within-Cluster Sum of Squares (WCSS) or silhouette score to determine the optimal number of clusters.

Step 8: Clustering analysis—This includes analyzing the resultant clusters for derivations or insights to understand the underlying patterns within the dataset.

Step 9: Sub-clustering analysis of clusters obtained from combined approach of PCA and K-means clustering.

For details on PCA and K-means clustering, refer to the Supplementary Material Text S1.

Optimization of Clusters

Following sub-clustering, a "technique for Order Preference by Similarity to Ideal Solution (TOPSIS)" subcategory of **Multi-Criteria Decision Analysis algorithm (MCDA)** was employed, which allowed for structured decision-making by evaluating multiple property criteria simultaneously to predict the optimum alloys with the best combination of properties within each sub-cluster.

The TOPSIS approach consisted of the following steps (For details see Supplementary Material):

Step 1: Convert categorical variables to numerical values.

Step 2: Construct properties matrix**.**

Step 3: Normalization of property matrix consisting of all the properties of the alloys.

Step 4: Determine the ideal-best solution, T*, and ideal-worst solutions, T^{-}.

Step 5: Calculate the distance of each alloys from the Ideal-best and Ideal-worst Solutions.

Step 6: Closeness Coefficient, i.e., TOPSIS score.

Step 7: Ranking of all alloys based on TOPSIS score (C^{i}), to identify the optimum alloy.

Step 8: Optimum alloy prediction by the highest TOPSIS score.

This method ensured an optimum alloy is selected for each sub-cluster having a superior combination of properties covering the broader range of properties of already existing alloys within the same sub-cluster, which contributed toward reducing the number of alloys [39, 40].

Results and Discussion

Data Collection

An accuracy of machine learning prediction depends on the availability of a high-quality dataset. We gathered 292 sets of data on 42 wrought 6xxx series aluminum alloys (see Supplementary Material). Data collection involves a number of quantitative and qualitative properties in a number of tempering as described above. These tempering conditions, i.e., T5, T6, T7, represent stages of artificial aging [41]. Every stage of artificial aging gives the alloys a unique set of properties. Note that here in this work we only considered artificial aging as this is the main temper to achieve high strength, used in structural applications. Since, 6xxx series alloys are primarily heat-treatable, our study focused on heat-treated conditions rather than work-hardened (H-states) or annealed (O-condition) alloys. This aligns with industrial practices. The methodology can be expanded to other process states and alloying systems but this goes beyond the scope of this work that is focused on the methodology development.

The numerical columns in the dataset represented chemical compositions spanning the following intervals: $0.35 \leq Si \leq 1.35$, $0.075 \leq Fe \leq 0.5$, $0.1 \leq Cu \leq 0.95$, $0.03 \leq Mn \leq 0.75$, and $0.35 \leq Mg \leq 1.4$. The tensile properties depended on the tempering conditions employed. For instance, for T6 tempering the properties ranges were as follows: $160 \leq$ Yield strength ≤ 430, $205 \leq$ Ultimate tensile strength ≤ 483, $3 \leq$ Elongation at fracture ≤ 15, $13.14 \leq$ Modulus of toughness[1] [42] ≤ 49.5, and $62 \leq$ High cycle fatigue strength ≤ 180. It can be noted that we standardized the chemical composition data to ensure consistency and 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. Also, this preprocessing step 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 anodizing response, were transformed into numerical values as Poor = 0, Fair = 1, Good = 2, and Excellent = 3. The range of total solute content (Mg + Si + Cu) was between 1 and 3 wt%, suggesting that a larger solute content was associated with a greater precipitation of strengthening precipitates. 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 [43].

Combined PCA and K-Means Clustering

A combined approach of PCA and K-means clustering was employed to examine the similarities present in the 6xxx series aluminum alloy dataset and organize them into clusters consisting similar compositions and properties.

We applied standard scaling to the entire dataset after concatenating converted categorical features with the numerical features (i.e., compositions, yield strength, modulus of toughness, fatigue strength) using a joining function. Standard scaling eliminated the disparities between absolute values (see Table 1) by dataset transformation into a more consistent scale (see Table S1) and assigning a standard deviation of 1 and a mean of 0 to all alloy attributes. If we had not standardized the data, properties with larger numerical ranges, such as yield strength, would have disproportionately influenced the clustering process. This could result in the clustering being dominated by such properties, making it difficult to interpret how compositions and other properties are related. To clearly visualize the dataset and minimize its dimensionality, we employed a PCA variance plot as shown in Fig. 2 (a) between the cumulative explained variance (i.e., total variance explained by each component) and the number of principal components. Conventionally 80% of the dataset information should be retained but at last the choice is ours which number of parameters we need to clearly visualize

[1] Modulus of toughness $= 0.5 \times (UTS + YS) \times EI$

Table 1 Composition and property variability across clusters

Group	0 (Medium)	1 (High)	2 (Lowest)	3 (Highest)	4 (Low)
Si, wt%	0.55–0.6	0.9–1.35	0.35–0.58	0.8–1.35	0.7–1.25
Fe, wt%	0.35–0.4	0.2–0.5	0.08–0.18	0.1–0.25	0.2–0.25
Cu, wt%	0.2–0.28	0.1–0.6	0.1–0.25	0.85–0.95	0.1–0.2
Mn, wt%	0.1–0.27	0.7–0.95	0.03–0.12	0.5–0.85	0.1–0.6
Mg, wt%	0.85–1	0.7–0.95	0.35–0.68	0.9–1.4	0.5–0.8
Yield strength, MPa	240–296	270–352	160–214	350–430	195–270
Modulus of toughness (MOT), MPa%	25–35	14–36	15–23	35–49	20–31
Fatigue strength	90–110	95–100	62–88	110–180	88–110
Corrosion	Good	Good–Excellent	Good–Excellent	Fair-Good	Good
Weldability	Excellent	Good	Excellent	Poor-Good	Excellent
Anodizing	Fair-Excellent	Good	Good	Fair-Good	Good
Mg:Si	1.667	< 1	1.33–1.68	1–1.667	< 1
Cu content	medium	high	low	High	low
Excess Si	Stoichiometry	Large excess	Slight excess	Excess	Large excess
Mg + Si + Cu, wt%	1.9 (medium)	2.3 (high)	1.3 (lowest)	3.4 (high)	1.8 (low)

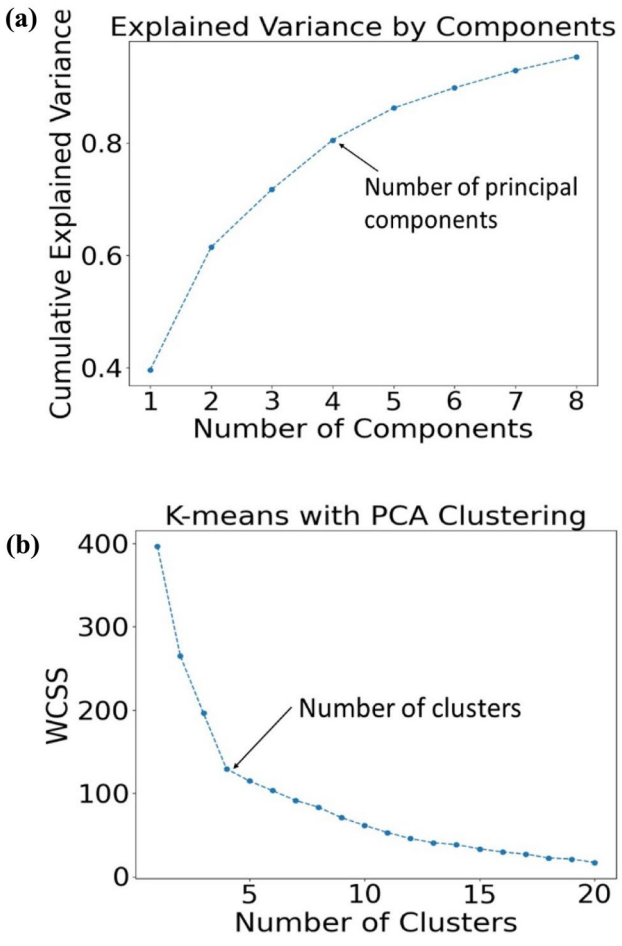


Fig. 2 A PCA variance plot to detect the number of principal components (**a**) and an elbow plot to find the optimal number of clusters (**b**)

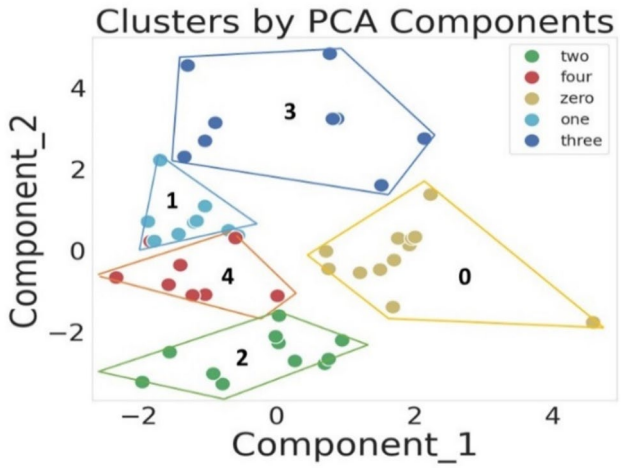


Fig. 3 Clustering results for T6

the clusters [37, 44], therefore we selected four principal components, as indicated in the curve. Consequently, we reduced the dataset dimensionality from 12 features (12 D space) to 4 components for easier cluster visualization in a 2D or 3D space.

We employed an elbow plot on the low-dimensional feature dataset, as illustrated in Fig. 2 (b) as WCSS vs the number of clusters [36]. The point on the x-axis at which there was a kink in the elbow curve represented the number of clusters present in the dataset. As shown in the curve the kink occurred between points 4 and 5. After employing both values of the number of clusters, we found 5 well-defined clusters corresponding to point 5.

Figure 3 illustrates five well-defined clusters at T6 temper.

Similar clustering patterns were observed at T7 and T5 tempering conditions, as shown in Fig. S-1 (a) & (b) (refer to Supplementary Material), with similar alloy grades present within them. Same colors specify similar clusters.

Clustering results remained same at various processing conditions, i.e., rolling, forging, and extrusion of 6xxx series aluminum alloys (the results are not shown here for brevity), which further supported our clustering findings that remained unchanged even if processing and tempering conditions varied. We have discussed a detailed explanation of the T6 tempering condition in our previous paper [30], because it is the most commonly used condition in 6xxx series aluminum alloys. The current clustering results were quite similar but more clearly defined as they include a wider range of properties. For further analysis, we extracted this information (i.e., clusters numbered from 0 to 4) into a CSV file and manually examined whether the algorithm correctly grouped the alloys before further processing. The results confirmed that the algorithm successfully categorized alloys with similar properties and composition ranges into the same clusters. Table 1 illustrates the range of chemical compositions and properties values found in each cluster.

For example, Cluster 2 consists of alloys that possess the lowest range of tensile properties, due to low amount of solute content, i.e., Mg + Si + Cu while it showed excellent corrosion resistance and weldability. On the other hand, Cluster 3 comprised alloys containing higher range of tensile properties but showed medium corrosion resistance and weldability. It is well-established that, in general, as the strength of these alloys increases, corrosion resistance and weldability tend to decrease, and vice versa as can be seen in Table 1. This inverse relationship is a key characteristic of the 6xxx alloys, making it crucial to evaluate these properties together to get a comprehensive view of how the alloys perform.

Sub-clustering

After obtaining five well-defined clusters comprising alloys of similar properties and compositions, we further refined these clusters by implementing same combined approach of PCA and K-means clustering algorithm as we performed in Section"Combined PCA and K-Means Clustering". This refinement was performed to create finer and more compact sub-clusters containing a closer range of properties, so that in next step optimum alloys could be selected.

First, we applied PCA to reduce the dataset dimensionality, enabling better visualization of the sub-clusters. Then, we used K-means clustering to identify number of sub-clusters within each cluster. Both the clustering results and our domain understanding confirmed the presence of two clear sub-clusters per cluster also called hyper-parameter tuning where we tune the parameters, i.e., number of clusters and number of components. This combined

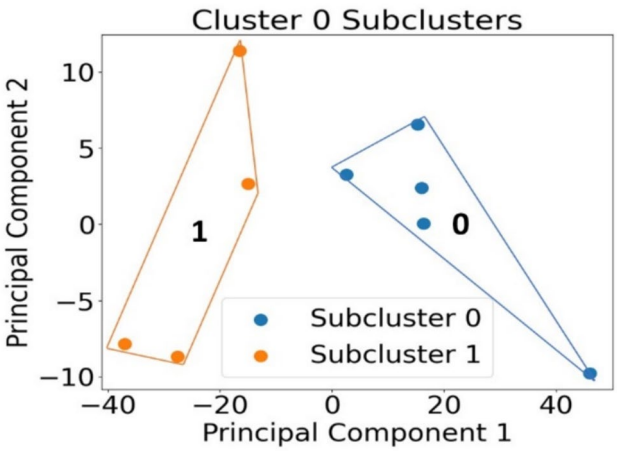


Fig. 4 Sub-clustering analysis of Cluster 0

approach of PCA and K-means clustering resulted in 10 clear sub-clusters, each containing a compact range of properties and compositions as seen in Fig. 4 for Cluster 0 and in Fig. S-2 (refer to Supplementary Material) for other clusters. This step ensured that we did not lose a significant amount of information while selecting the optimal alloys from the sub-clusters [45, 46].

Let us consider Cluster 0, which was sub-divided into two sub-clusters, i.e., Sub-cluster 0 and Sub-cluster 1. Sub-cluster 0 exhibit a higher range of tensile properties, moderate corrosion resistance, poor anodizing response, while Sub-cluster 1 consists of a moderate range of tensile properties but excellent corrosion resistance and weldability. This trend was seen across all sub-clusters, suggesting the alloys were grouped with a finer range of properties.

The next step involved optimization of these sub-clusters to identify representative alloys with the best combination of properties.

Optimum Alloy Selection

After obtaining sub-clusters containing a narrower range of properties, next step was to optimize them selecting the best combination of properties alloys by using TOPSIS algorithm. TOPSIS is a powerful method of multi-criteria decision-making, intuitively ranking the alloys with respect to their distance from the ideal solution and efficiently balancing the trade-offs among conflicting criteria.

Consider for example Sub-cluster 0 of Cluster 0, which involves four alloys with different ranges of properties listed in Table S1.

Step 1: Converting categorical ratings to numerical values: Ratings of corrosion resistance, weldability, and anodizing response were converted to numerical

values, i.e., Excellent = 3, Good = 2, Fair = 1, and poor = 0.

Step 2: Construction of a property matrix: A property matrix was created for all alloy properties of Sub-cluster 0 of Cluster 0 as follows (see Table S1 for clarity):

270	270	240	270
25.5	29.5	25	31.9
110	110	102.7	110
2	2	2	2
2	3	3	3
0	2	3	3

Step 3: Normalization of the property matrix: All properties were converted to normalized values (i.e., shown in Table S1 in parenthesis) by using Eq. S3 (see Supplementary Material).

Step 4: Determine ideal-best and ideal-worst solutions following an algorithm described in Supplementary Materials (Calculation S1).

Step 5: Distance of each alloy from ideal-best and ideal-worst solutions (refer to Calculation S2 for details):

The ideal-best solution for alloys were as follows:
$D_A^* = 0.915, D_B^* = 0.2896, D_C^* = 0.0717$, and $D_D^* = 0$.
The distance for each alloy from the ideal-worst solution:

$D_A^- = 0.098,\ D_B^-, D_C^- =, 0.095,$ and $D_D^- = 0.916$

Step 6: Rank the alloys: Ranking the four alloys in Sub-cluster 0 of Cluster 0 based on the TOPSIS score (refer to calculation S3 for details): Alloy D: $C_D = 1$ (best choice), Alloy C: $C_C \approx 0.9260$, Alloy B: $C_B \approx\approx 0.6903$, and Alloy A: $C_A \approx 0.0441$.

Based on the TOPSIS method calculated above, **Alloy D** (Table S1) is the best choice alloy, indicating it offers a balanced performance across all properties compared to other alloys as shown in Fig. S-3 (refer to Supplementary Material) where we also employed other MCDM methods (i.e., AHP, VIKOR [39]) for comparison, each time Alloy D gets the best score.

The other optimum alloys chosen for the remaining sub-clusters followed the same logic. The optimum alloys chosen following the optimization of each sub-cluster and the parent cluster are shown in Table 2. Through optimization, the chosen alloys were guaranteed to match the specified criteria with a wide range of properties and applications. By narrowing down the solution space, we ensure that only the most effective and efficient alloys are selected, leading to cost savings, improved product performance and simplified sorting of scrap in the aluminum industry. Additionally, in most clusters the optimum alloys selected coincide with commercial grade alloys offering better combination of properties for similar applications as shown in Table 2.

While we have reduced the alloys within clusters to a few top-performing options, the clusters will always remain available to meet any unique demands in properties, processing, or economics.

Metallurgical Understanding of Clustering

The clustering by ML methods will only be reliable if it had some metallurgical/physical background. The general metallurgical reasoning behind clusters obtained by the combined approach of PCA and K-means clustering, as well as the rationale behind optimal alloy selection is discussed here. According to the well-known sequence of precipitation upon aging in Al–Mg-Si alloys, the precipitation in ternary 6xxx series alloys proceeds as follows: supersaturated solid solution (SSS) → GP zones → β″ → β′ → β [43, 57]. β″ is the main hardening phase in 6xxx series aluminum 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 as has been summarized in our previous paper [30] and briefly outlined below for some ranges of alloying elements (with hardening phases in bold) [57]:

At a lower Cu concentration: SSS—**GP zones—β″ (a modification with Cu)**—Si—numerous variations of β′ (with Cu) including β′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 β′C, Q′ and θ′—Q (AlMgSiCu), Mg_2Si, Al_2Cu and Si.

These precipitation paths may also occur at intermediate copper concentrations during artificial aging. Also, Table S2 (Supplementary Material) shows the role of each alloying element in the precipitation hardening and the effects on mechanical and functional properties.

Based on this general understanding of the metallurgy of 6xxx series alloys, we can suggest the following metallurgical reasons behind grouping the alloys into different clusters. An example for Cluster 0 is given below, other clusters are listed in Supplementary Material (Text S2).

Cluster 0 (Medium strength alloys): Composition range (wt%): Si: (0.55–0.6), Fe: (0.35–0.4), Cu: (0.2–0.28), Mn: (0.1–0.27), Mg: (0.85–1) determines formation of the following phases: medium amount of β′′, Q′, β′.

Table 2 Properties and applications of optimum alloys in each cluster

Clusters	Optimum alloys (their standard analogues) and their properties		Applications
Cluster 0	6061, Sub-cluster 0	Low strength, modulus of toughness, and good CR, weldability, anodizing	Truck frames, Rail coaches, Military, commercial bridges, Ships, Towers, Aerospace, Rivets, Motorboats, Boiler making [47]
	6040, Sub-cluster 1	Good strength, modulus of toughness, and excellent CR, weldability, fair anodizing	Automotive, small engine parts, sports products [48]
Cluster 1	6012, Sub-cluster 0	Moderate strength, toughness, weldability, anodizing and excellent CR	Manufacturing parts which require easy machining and are suitable for anodization. [49]
	6070, Sub-cluster 1	High strength, modulus of toughness, weldability, moderate CR, anodizing	Pipelines and heavy duty welded structures [50]
Cluster 2	6063, Sub-cluster 0	Low strength, toughness, excellent CR, weldability and anodizing response	Transportation, architectural and, industrial applications [51]
	6463, Sub-cluster 1	Moderate strength, modulus of toughness, and excellent CR, weldability, anodizing response	Manufacturing extruded architectural and trim sections [52]
Cluster 3	6033, Sub-cluster 0	High strength, modulus of toughness, and good CR, weldability, anodizing	Recreational, medical and after-market automotive parts used in fly reels, oxygen regulators and small gas-powered engines [53]
	6011 A, Sub-cluster 1	High strength, modulus of toughness, moderate CR, weldability, anodizing	Automobile vehicles [54]
Cluster 4	6081, Sub-cluster 0	High strength, moderate modulus of toughness, good CR, weldability, and anodizing	Structural components, such as frames for bicycles, motorcycles, and automobiles [55]
	6009, Sub-cluster 1	Low strength, modulus of toughness, good strength, good CR, weldability and anodizing response	Components of car body, such as door panels, roofs, front and rear bumpers, side skirts, and wheel arches [56]

At an average Cu content of 0.275% and with a small excess of Si as Mg:Si ~ 1.67 (1.73 is the stoichiometry in Mg_2Si) the main hardening phase is β′′ may be assisted at later stages with the Q′ phases [57, 58]. 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 the intergranular corrosion (IGC) resistance [59, 60]. Weldability is excellent due to the low presence of liquation cracking-causing elements, i.e., Cu and the moderate formation of Mg_2Si phases [61]. Anodizing response is excellent due to moderate amount of constituent Si and Mg_2Si particles, resulting in softer surfaces and increased surface gloss [62].

Similar metallurgical reasoning applies to other clusters. Table 2 provides concise descriptions of the wider range of properties and applications covered by optimal alloys selected by optimization algorithm. This approach serves as a **pre-screening** tool to suggest alloys with the balanced combination of properties and a theoretical metallurgical reasoning behind its selection process. It is important that each of the optimum alloys has an analogue among the standard alloy grades, which helps to validate the selection as the standard alloys and their properties are well documented. A comparison is performed in Table 3 for the optimum alloys selected for Cluster 0, contrasting the reported property values from the MatWeb website [63] with predicted values from our database using the optimization algorithm. Note that the MatWeb database was not used in building our database and, therefore, can be considered as an independent source of data. This comparison reveals minimal disparity between the two data sources. Here, Alloy A is ASTM 6061, and Alloy B is ASTM 6040.

Table 3 Comparison of predicted vs measured properties of optimum alloys for cluster 0

Property	Alloy A (predicted)	Alloy A (reported)	Alloy B (predicted)	Alloy B (reported)
Yield strength	276	276	296	296
Elongation	12	17	15	15
Fatigue strength	96.5	96.5	119	119
Corrosion resistance	Good	Good	Excellent	Excellent
Weldability	Excellent	Excellent	Excellent	Excellent
Anodizing	Excellent	Excellent	Good	Good

In this work we developed an ML-based methodology and identified the optimal alloys for each sub-cluster of the 6xxx alloying series, effectively reducing the number of alloys while keeping the level of properties; and successfully found the metallurgical reasoning behind clustering process.

In future we will focus on recycling issues within sub-clusters by determining the mixing ratios of the alloys allowing the preservation of the optimum properties.

Conclusion

In this study, we developed a methodology of reducing the number of standard alloys while retaining the level of properties within each sub-cluster of the alloying series. This methodology was demonstrated for the 6xxx series of aluminum alloys (Al–Mg-Si–(Cu) under T5, T6, and T7 tempering conditions using PCA and K-means clustering and optimization algorithm. From an initial set of 42 alloys, we identified 10 optimum alloys, all of which have analogues in the existing commercial alloys. These selected alloys not only maintain desirable mechanical and functional properties but also improve recyclability due to the lesser requirements for scrap sorting.

The clustering approach effectively categorized alloys, ensuring that widely used commercial alloys were included in each cluster. Notably, 6061, 6081, and 6063 alloys corresponded to low, medium, and high property clusters, respectively.

A key limitation of the current methodology is that generalizing 42 alloys into 10 optimized ones may not fully account for niche applications with unique property requirements. However, the clustering process groups alloys with similar overall properties, ensuring that a broad range of applications is covered. In specialized cases, the full range of alloys remains available for tailored needs. Additionally, experimental validation is still required to confirm our findings, though well-established reference data supported our optimization approach. This provides a framework for improved alloy classification, reducing compositional variability, and enhancing sustainability in recycling when applied to well-sorted materials.

Future research will focus on refining recycling strategies within sub-clusters by determining the optimal mixing ratios of alloys while maintaining desirable properties.

This methodology provides a sustainable approach to alloy selection and enhances efficiency in the aluminum industry, facilitating both performance improvement and material recyclability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40831-025-01112-4.

Acknowledgments The authors extend their gratitude to Prof. H. Assadi and Prof. I. Chang for their insightful conversations. The corresponding author is grateful to Brunel University of London for providing funding for the scholarship and EPSRC for funding the Circular Metals Centre (UKRI/EPSRC grant EP/V011804/1).

Author Contributions Conceptualization, DE and TT; Methodology, TT; Formal Analysis, TT, DE; Investigation, TT; Data Curation, TT; Writing – Original Draft Preparation, TT, DE; Writing – Review and Editing, DE, CM; Supervision, DE, CM; Project Administration, DE; Funding Acquisition, CM, DE.

Funding This research was funded by Brunel University of London, UKRI/EPSRC grant EP/V011804/1 and was carried out within the Circular Metals Centre framework.

Data Availability The corresponding author can provide the raw or processed data required to reproduce the findings upon request. Dataset for 6xxx series aluminum alloys can be downloaded from Brunel University of London repository: https://doi.org/10.17633/rd.brunel.28471826.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Luo Z, Soria A (2007) Prospective study of the world aluminum industry. JRC Sci Technol Rep 22951.
2. Davis JR (ed) (1993) Aluminum and aluminum alloys. ASM International, Materials Park, pp 3–88
3. Aluminum E (2019) European aluminum. Belgium, Brussels
4. Italpres. (n.d.). EU aluminum demand to 2050: Primary and recycled metal ratio, parity objective. https://www.italpres.com/aluminum-die-casting-news/eu-aluminum-demand-2050-primary-recycled-metal-ratio. Accessed 12 Feb 2025
5. Eheliyagoda D, Li J, Geng Y, Zeng X (2022) The role of China's aluminum recycling on sustainable resource and emission pathways. Resour Policy 76:102552. https://doi.org/10.1016/j.resourpol.2022.102552
6. Giofré D, Junge T, Curtin WA, Ceriotti M (2017) Ab initio modelling of the early stages of precipitation in Al-6000 alloys. Acta Mater 140:240–249

7. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201. https://doi.org/10.1038/nmat3568
8. Liu Y, Zhao T, Ju W, Shi S (2017) Materials discovery and design using machine learning. J Materiomics 3(3):159–177. https://doi.org/10.1016/j.jmat.2017.08.002
9. Li J, Zhang Y, Cao X, Zeng Q, Zhuang Y, Qian X, Chen H (2020) Accelerated discovery of high-strength aluminum alloys by machine learning. Commun Mater 1:73. https://doi.org/10.1038/s43246-020-00074-2
10. Xue D, Xue D, Yuan R, Zhou Y, Balachandran PV, Ding X, Sun J, Lookman T (2017) An informatics approach to transformation temperatures of NiTi-based shape memory alloys. Acta Mater 125:532–541. https://doi.org/10.1016/j.actamat.2016.12.009
11. Raccuglia P, Elbert KC, Adler PD, Falk C, Wenny MB, Mollo A, Zeller M, Friedler SA, Schrier J, Norquist AJ (2016) Machine-learning-assisted materials discovery using failed experiments. Nature 533:73–76. https://doi.org/10.1038/nature17439
12. Devi MA, Prakash CPS, Chinnannavar RP, Joshi VP, Palada RS, Dixit R (2020) An informatic approach to predict the mechanical properties of aluminum alloys using machine learning techniques. In: Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE, pp 536–541. https://doi.org/10.1109/ICOSEC49089.2020.9215277.
13. Gao EJ, Liu Y, Wang Y, Hu X, Yan W, Ke X, Zhong L, He Y, Ren X (2017) Designing high dielectric permittivity material in barium titanate. J Phys Chem 121:13106–13113. https://doi.org/10.1021/acs.jpcc.7b04636
14. Yuan R, Liu Z, Balachandran PV, Xue D, Zhou Y, Ding X, Sun J, Xue D, Lookman T (2018) Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning. Adv Mater 30:1702884. https://doi.org/10.1002/adma.201702884
15. Stanev V, Oses C, Kusne AG, Rodriguez E, Paglione J, Curtarolo S, Takeuchi I (2018) Machine learning modeling of superconducting critical temperature. Npj Comput Mater 4:29. https://doi.org/10.1038/s41524-018-0085-8
16. Xu W, Diaz R, del Castillo PEJ, Van Der Zwaag S (2009) A combined optimization of alloy composition and aging temperature in designing new UHS precipitation hardenable stainless steels. Comput Mater Sci 45:467–473. https://doi.org/10.1016/j.commatsci.2008.11.006
17. Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, Dai L, Lookman T, Su Y (2019) Machine learning assisted design of high entropy alloys with desired property. Acta Mater 170:109–117. https://doi.org/10.1016/j.actamat.2019.03.010
18. Zeng Y, Li Q, Bai K (2018) Prediction of interstitial diffusion activation energies of nitrogen, oxygen, boron, and carbon in BCC, FCC, and HCP metals using machine learning. Comput Mater Sci 144:232–247. https://doi.org/10.1016/j.commatsci.2017.12.030
19. Wu H, Lorenson A, Anderson B, Witteman L, Wu H, Meredig B, Morgan D (2017) Robust FCC solute diffusion predictions from ab-initio machine learning methods. Comput Mater Sci 134:160–165. https://doi.org/10.1016/j.commatsci.2017.03.052
20. Deng Z, Yin H, Jiang X, Zhang C, Zhang K, Zhang T, Xu B, Zheng Q, Qu X (2018) Machine learning aided study of sintered density in Cu-Al alloy. Comput Mater Sci 155:48–54. https://doi.org/10.1016/j.commatsci.2018.07.049
21. Jiang X, Yin HQ, Zhang C, Zhang RJ, Zhang KQ, Deng ZH, Liu GQ, Qu XH (2018) A materials informatics approach to Ni-based single crystal superalloys lattice misfit prediction. Comput Mater Sci 143:295–300. https://doi.org/10.1016/j.commatsci.2017.09.061
22. Wang B, Zhao W, Du Y, Zhang G, Yang Y (2016) Prediction of fatigue stress concentration factor using extreme learning machine. Comput Mater Sci 125:136–145. https://doi.org/10.1016/j.commatsci.2016.08.035
23. Rovinelli A, Sangid MD, Proudhon H, Ludwig W (2018) Using machine learning and a data-driven approach to identify the small fatigue crack driving force in polycrystalline materials. Npj Comput Mater 4:35. https://doi.org/10.1038/s41524-018-0094-7
24. Mangos J, Birbilis N (2021) Aluminum alloy design and discovery using machine learning. arXiv preprint arXiv:2105.14806
25. Berboucha M (2018) Scientists use artificial intelligence to discover new materials. Forbes. https://www.forbes.com/sites/meriameberboucha/2018/04/22/scientists-use-artificial-intelligence-to-discover-new-materials/#79cf676738c4. Accessed 13 Feb 2025
26. Salema Project (2024) Sustainable aluminium applications for the aerospace and automotive industries. Available online: https://salemaproject.eu/
27. Boin UM, Bertram M (2005) Melting standardized aluminum scrap: a mass balance model for Europe. JOM 57:26–33
28. Paraskevas D, Kellens K, Dewulf W, Duflou JR (2015) Environmental modelling of aluminum recycling: a life cycle assessment tool for sustainable metal management. J Clean Prod 105:357–370
29. Duflou JR, Tekkaya AE, Haase M, Welo T, Vanmeensel K, Kellens K, Dewulf W, Paraskevas D (2015) Environmental assessment of solid state recycling routes for aluminum alloys: can solid state processes significantly reduce the environmental impact of aluminum recycling? CIRP Ann 64(1):37–40
30. Tiwari T, Jalalian S, Mendis C, Eskin D (2023) Classification of T6 tempered 6xxx series aluminum alloys based on machine learning principles. JOM 75:4526–4537. https://doi.org/10.1007/s11837-023-06025-9
31. Totten GE, Tiryakioglu M, Kessler O (2018) Encyclopedia of aluminum and its alloys, two volume set. CRC Press, Boca Raton. https://doi.org/10.1201/9781351045636
32. Make it From (2022) Aluminum alloys—materials—engineering. https://www.makeitfrom.com/material-group/Aluminum-Alloy. Accessed 18 April 2022
33. Rolon-Mérette D, Ross M, Rolon-Mérette T, Church K (2016) Introduction to Anaconda and Python: installation and setup. Quant Methods Psychol 16:S3–S11. https://doi.org/10.20982/tqmp.16.5.S003
34. Sial AH, Rashdi SYS, Khan AH (2021) Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. Int J 10:277–281. https://doi.org/10.30534/ijatcse/2021/391012021
35. Waskom ML (2021) Seaborn: Statistical data visualization. J Open Source Softw 6:3021. https://doi.org/10.21105/joss.03021
36. Kodinariya TM, Makwana PR (2013) Review on determining number of cluster in K-means clustering. Int J 1:90–95
37. Daffertshofer A, Lamoth CJ, Meijer OG, Beek PJ (2004) PCA in studying coordination and variability: a tutorial. Clin Biomech 19:415–428. https://doi.org/10.1016/j.clinbiomech.2004.01.005
38. Ding C, He X (2004) K-means clustering via principal component analysis. In: Proceedings of the Twenty-First International Conference on Machine Learning. https://doi.org/10.1145/1015330.1015408.
39. Yadav SK, Joseph D, Jigeesh N (2018) A review on industrial applications of TOPSIS approach. Int J Serv Oper Manag 30:23–28. https://doi.org/10.1504/IJSOM.2018.091438
40. Diaby V, Campbell K, Goeree R (2013) Multi-criteria decision analysis (MCDA) in health care: a bibliometric analysis. Oper Res Health Care 2:20–24. https://doi.org/10.1016/j.orhc.2013.03.001
41. Series A, Element PA (2023) Aluminum alloys and heat treatment. https://www.cabww.com/uploads/case_studies/HeatTreatAlum-wp.pdf. Accessed 1 Jan 2023
42. Kuhn H (2000) ASM Handbook, vol. 8 (Mechanical Testing and Evaluation). ASM International, Materials Park, OH. https://doi.org/10.31399/asm.hb.v08.9781627081764.

43. Chakrabarti DJ, Laughlin DE (2004) Phase relations and precipitation in Al–Mg–Si alloys with Cu additions. Prog Mater Sci 49:389–410. https://doi.org/10.1016/S0079-6425(03)00031-8
44. Harrison JM, Howard D, Malven M, Halls SC, Culler AH, Harrigan GG, Wolfinger RD (2013) Principal variance component analysis of crop composition data: a case study on herbicide-tolerant cotton. J Agric Food Chem 61(26):6412–6422. https://doi.org/10.1021/jf400606t
45. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
46. Witten IH, Frank E, Hall MA, Pal CJ (2011) Data mining: practical machine learning tools and techniques. Morgan Kaufmann, San Francisco
47. Campbell FC (ed) (2012) Lightweight materials: understanding the basics. ASM International, Materials Park
48. Anderson K, Weritz J, Kaufman JG (eds) (2019) Properties and selection of aluminum alloys. ASM International, Materials Park
49. Shahsavari MH, Ahmadi F (2015) Evolution of texture and grain size during equal channel angular extrusion of pure copper and 6012 aluminum. J Mod Process Manuf Prod 4:47–58
50. Sivakumar VR, Kavitha V, Saravanan NS, Nanjundamoorthi TT, Chanakyan C (2023) Tribological behavior on stir casted metal matrix composites of Al 6070 and TiC reinforcement with Taguchi S/N ratios. Mater Today Proc 77:455–461. https://doi.org/10.1016/j.matpr.2022.11.222
51. Dhoska K, Markja I, Bebi E, Sulejmani A, Koça O, Sita E, Pramono A (2023) Manufacturing process of the aluminum alloy AA6063 for engineering applications. J Integr Eng Appl Sci 1:1–13. https://doi.org/10.15157/jieas.2023.1.1.xxx-xxx
52. Sekhar YC, Gopichand A, Sukumar RS, Kumar NP (2017) Optimization of surface roughness of 6463 aluminium alloy and brass materials in CNC milling operation using Taguchi's design. Optim 6:2. https://doi.org/10.15680/IJIRSET.2017.0602018
53. Finn ME (2018) Machining of aluminum alloys. ASM International, Materials Park
54. Klemenc J, Glodež S, Steinacher M, Zupanič F (2023) LCF behaviour of high strength aluminum alloys AA 6110A and AA 6086. Int J Fatigue 177:107971. https://doi.org/10.1016/j.ijfatigue.2023.107971
55. Suriyanarayana A (2021) Experimental investigation of effect of rubber ash and molybdenum disulphide on Al6081 based hybrid MMC. IJAEM 2:2395–5252. https://doi.org/10.35629/5252-030310551063
56. Staley JT, Lege DJ (1993) Advances in aluminum alloy products for structural applications in transportation. Le J Phys IV. https://doi.org/10.1051/jp4:1993728
57. Eskin DG (2003) Decomposition of supersaturated solid solutions in Al–Cu–Mg–Si alloys. J Mater Sci 38(2):279–290. https://doi.org/10.1023/A:1021109514892
58. Zang R, Ning Y, Ding L, Jia Z, Xiang K, Liu Q, Li Y (2022) Study on properties and precipitation behavior of 6000 series alloys with high Mg/Si ratios and Cu contents. Mater Charact 194:112402. https://doi.org/10.1016/j.matchar.2022.112402
59. Zou Y, Liu Q, Jia Z, Xing Y, Ding L, Wang X (2017) The intergranular corrosion behavior of 6000-series alloys with different Mg/Si and Cu content. Appl Surf Sci 405:489–496. https://doi.org/10.1016/j.apsusc.2017.02.045
60. Larsen MH, Walmsley JC, Lunder O, Mathiesen RH, Nisancioglu K (2008) Intergranular corrosion of copper-containing AA6xxx AlMgSi aluminum alloys. J Electrochem Soc. Doi 10(1149/1):2976774
61. Katoh M, Kerr HW (1987) Investigation of heat-affected zone cracking of GTA welds of Al-Mg-Si alloys using the Varestraint test. Weld J 66(12):251–259
62. Aggerbeck M, Canulescu S, Dirscherl K, Johansen VE, Engberg S, Schou J, Ambat R (2014) Appearance of anodised aluminum: Effect of alloy composition and prior surface finish. Surf Coat Technol 254:28–41. https://doi.org/10.1016/j.surfcoat.2014.05.047
63. MatWeb (2024) Aluminum reference. https://www.matweb.com/reference/aluminum. Accessed 18 Oct 2024

Authors and Affiliations

Tanu Tiwari[1] · Chamini Mendis[1] · Dmitry Eskin[1]

✉ Tanu Tiwari
Tanu.Tiwari@brunel.ac.uk

Chamini Mendis
Chamini.Mendis@brunel.ac.uk

Dmitry Eskin
Dmitry.Eskin@brunel.ac.uk

[1] BCAST, Brunel University of London, Uxbridge, Middlesex UB8 3PH, UK