SIMULTANEOUS MODELLING AND CLUSTERING OF VISUAL FIELD DATA

A thesis submitted for the degree of Doctor of Philosophy

by

Mohd Zairul Mazwan Bin Jilani

Department of Computer Science

Brunel University London

December 2017

Abstract

In the health-informatics and bio-medical domains, clinicians produce an enormous amount of data which can be complex and high in dimensionality. This scenario includes visual field data, which are used for managing the second leading cause of blindness in the world: glaucoma. Visual field data are the most common type of data collected to diagnose glaucoma in patients, and usually the data consist of 54 or 76 variables (which are referred to as visual field locations). Due to the large number of variables, the six nerve fiber bundles (6NFB), which is a collection of visual field locations in groups, are the standard clusters used in visual field data to represent the physiological traits of the retina. However, with regard to classification accuracy of the data, this research proposes a technique to find other significant spatial clusters of visual field with higher classification accuracy than the 6NFB.

This thesis presents a novel clustering technique, namely, Simultaneous Modelling and Clustering (SMC). SMC performs clustering on data based on classification accuracy using heuristic search techniques. The method searches a collection of significant clusters of visual field locations that indicate visual field loss progression. The aim of this research is two-fold. Firstly, SMC algorithms are developed and tested on data to investigate the effectiveness and efficiency of the method using optimisation and classification methods. Secondly, a significant clustering arrangement of visual field, which highly interrelated visual field locations to represent progression of visual field loss with high classification accuracy, is searched to complement the 6NFB in diagnosis of glaucoma. A new clustering arrangement of visual field locations can be used by medical practitioners together with the 6NFB to complement each other in diagnosis of glaucoma in patients.

This research conducts extensive experiment work on both visual field and simulated data to evaluate the proposed method. The results obtained suggest the proposed method appears to be an effective and efficient method in clustering visual field data and

improving classification accuracy. The key contributions of this work are the novel model-based clustering of visual field data, effective and efficient algorithms for SMC, practical knowledge of visual field data in the diagnosis of glaucoma and the presentation a generic framework for modelling and clustering which is highly applicable to many other dataset/model combinations.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the most important person in my PhD journey, Dr. Stephen Swift, who has given continuous guidance, kind support, motivation, full commitment, and patience. Along the journey of my research, he has shared much knowledge which was very helpful and made my research flowed with ease. Not to forget, my second supervisor Dr. Allan Tucker for some tips and advices for the research work and the paper we published. Dr. Stephen Swift and Dr. Allan Tucker are the most friendly and easy-going supervisors, and knowledgeable experts in the area of this research, and this made my PhD studies became enjoyable work.

I am using this opportunity to also express my gratitude to the lecturers from Brunel Language Centre, Alison Cheetham and David Jones, for providing the face to face tutorials in academic writing. The lessons I gained from these two people were incredibly valuable and cannot easily be obtained outside my PhD journey.

My deepest appreciation goes to my parents, Jilani and Maznah, who forever support me for the spirit and the encouragement regardless of difficult situations I faced throughout my studies. Special thanks I dedicate to friends of mine at Brunel University and in Malaysia for consistent inspiration and laughter that made my PhD life in the UK more interesting.

A big thanks to the government of Malaysia for sponsoring this work, otherwise this work would not have manifested.

Spiritually, praise to the Almighty God for His grace. Without His blessed and excellent health been granted to me, I would not be able to produce this work.

AUTHOR'S DECLARATION

I confirm that this work is my own except where references are cited. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. The work was done under the guidance of Dr Stephen Swift and Dr. Allan Tucker, at the department of Computer Science, Brunel University London, United Kingdom.

Mohd Zairul Mazwan Bin Jilani

“Not everything that can be counted counts, and not everything that counts can be counted”

- Albert Einstein, Physicist -

Table of Contents

List of Figures

List of Tables

List of Equations

List of Algorithms

List of Definitions

Mathematical Notation

UI(a,b)	A random uniformly distributed uniform integer between a and b inclusive
$\| a \|$	An absolute value of a scalar
$\Re$	A real number
[…]	An operator concatenates matrices and/or vectors by column
$\in$	An element (member) of a set
t	A time point of a time series
{…}	A set
[[…], [..], ..]	List of clusters
Σ	Summation
Π	Product

Δ	A change in a variable
e	A mathematical constant, approximately equal to 2.71828 (5 d.p)

Abbreviations

6NFB	Six Nerve Fiber Bundles
AGIS	Advanced Glaucoma Intervention Study
CSA	Classical Simulated Annealing
FCA	Fast Simulated Annealing
FCV	Fold Cross Validation
FD	Fold
GSA	Generalised Simulated Annealing
HC	Hill Climbing
IOP	Intraocular pressure
J48	A Decision Tree Classifier
MNB	Multinomial Naïve Bayes
MNBU	Naïve Bayes Multinomial Updateable
NB	Naïve Bayes
NHS	National Health Services
PACG	Primary Closed-Angle Glaucoma
POAG	Primary Open-Angle Glaucoma
RGF	Restricted Growth Functions
RGFGSA	Restricted Growth Functions Generalised Simulated Annealing
RMHC	Random Mutation Hill Climbing
RRHC	Random Restart Hill Climbing
SA	Simulated Annealing
SMC	Simultaneous Modelling and Clustering

TN	True Negative
TP	True Positive
TSP	Travelling Salesman Problem
VF	Visual Field
WK	Weighted Kappa Statistics

Tools

Eclipse Kepler Service Release 2
WEKA Version 3.6.11
Matlab R2012b
RGui Version 3.4.2
R Studio Version 1.1.383
Refwork (Write-N-Cite III)
Microsoft Excel 2010
Microsoft Word 2010

Publications

M. Z. M. B. Jilani, A. Tucker and S. Swift (2016) "Simultaneous Modelling and Clustering of Visual Field Data", *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS).* , 213-218.

M. Z. M. B. Jilani, A. Tucker and S. Swift (2017) "An Application of Generalised Simulated Annealing towards the Simultaneous Modelling and Clustering of Glaucoma", submitted to the Journal of Heuristics, (based on Chapter 5).

Chapter 1

Introduction

1.1 Overview

This thesis presents research on a clustering technique (Simultaneous Modelling and Clustering - SMC) that is applied towards visual field data. The technique, which is also a model-based clustering technique, is devised to find a significant clustering arrangement of visual field locations that indicates visual field loss. The practicality of the clustering technique is investigated and tested on the data by implementing a number of optimisation and classification methods. Chapter 1 consists of eight sections. Section 1.1 provides an overview of the work presented in this thesis. Research background is discussed in section 1.2. The problems being addressed and the motivation behind this work are discussed in section 1.3. Research aim is set out in section 1.4. Building from research aim, research objectives and questions are formulated in section 1.5. Section 1.6 describes the methodology and details out SMC experiments conducted within this research. Research novelty is demonstrated in section 1.7. Finally section 1.8 summarises the chapters presented in this thesis.

1.2 Research Background

According to a report from IBM Marketing Cloud, "10 Key Marketing Trends For 2017," 90% of the world's data has been collected in the last two years (Loechner, 2016). With the advent of computerised technologies in the health-informatics and bio-informatics domain, there has been an explosion in the amount of data collected and they are rarely analysed (Herland, Khoshgoftaar and Wald, 2014). Visual field data, which are collected from patients to manage glaucoma, are a high volume dataset available on digital storage at clinics, and they need to be analysed to a satisfactory level in order to explain visual loss in patients. Due to the fact that limited knowledge can be obtained from a vast amount

of data by clinicians, analysing visual field data using machine learning techniques, therefore, could bridge the gap between clinicians and computer technology. Furthermore, prediction of glaucoma progression using machine learning could support medical practitioners in providing data-driven decision-making for a diagnosis of this irreversible disease in patients.

Visual field data are high dimensionality in nature. In other words, the data comprise of numerous variables (more than 52) which represent sensitivity of visual field locations in patients. Early exploratory experiments (classifying visual field data) on these 52 variables found that the predictive accuracies of the data were between 75.4% to 84.9%. Finding a collection of the variables on this dataset that could highly improve predictive accuracy using computer algorithm is an area to research. As such, clustering (Tran, Xue and Zhang, 2016), which is a machine learning technique, can be employed to reduce high dimensional data and in some cases, the goal of cluster analysis is a better understanding of the data. With clustering, some of the 52 visual field locations of visual field data can be grouped together to represent significant locations that indicate visual field loss. Finding latent groups of visual field locations within the data can provide useful information to medical practitioners and this is an open research opportunity.

Garway Heath and colleagues (Garway-Heath *et al.*, 2000) established the six nerve fiber bundles (6NFB) that map the visual field locations with the regions of the optic nerve head. Several studies (VanBuren *et al.*, 2016; Morales *et al.*, 2016) have used the 6NFB in analysing visual field data to diagnose glaucoma progression. However, this thesis presents a novel technique, namely, Simultaneous Modelling and Clustering (SMC), which searches other clustering arrangements of visual field that could better predict glaucoma progression.

1.3 Research Motivation and Problem

It is known that glaucoma is the second leading cause of blindness in the world and it is an irreversible disease. To date, there is no gold standard (X. Zhang *et al.*, 2017; Viswanathan, Fitzke and Hitchings, 1997) for analysing visual field data to provide identification and progression of visual field loss. Furthermore, the high-dimensionality of visual field data are not easy to comprehend without suitable analysis tools in order to interpret glaucoma severity in patients. Therefore, the problems motivate this research to cater for new knowledge about glaucoma progression and support medical practitioners in providing treatment to patients. Moreover, with additional knowledge which is resulted from data analyses, appropriate treatments can be delivered to patients to avoid and slow the progression of glaucoma in patients for a quality of life (Otori *et al.*, 2017; Sleath *et al.*, 2017; De Keyser, De Belder and De Groot, 2017). This can be achieved by means of advanced analyses using machine learning on visual field data.

With the data in hand, this research is motivated to propose a novel clustering and classification technique, as well as providing high prediction on glaucoma progression. Broadly speaking, there is no single method would be universally optimal across all classification problems as presented in the "no free lunch theorem" (Wolpert and Macready, 1997). As such, a new approach of clustering and classification is developed to provide prediction of visual loss.

Currently, the 6NFB are commonly used in analysing visual field data. Obtaining other sets of bundles (or clusters which are the term used in this study) in visual field data with high predictive accuracy is one of the research motivations of this thesis.

Visual field data used within this research consist of 52 variables (excluding the blind spot) and a short number of observations per patient. With these properties of the data, it is a challenge to model the data for predicting visual field loss. Modelling data using a high number of variables may result an inefficient classification and low accuracy (J.

Zhang *et al.*, 2017). Therefore, searching subsets of highly correlated variables in the data is one of solutions in modelling the data. This is similar to dimensionality reduction and feature selection. However, dimensionality reduction and feature selection is not a favourable approach for this research as they do not use all variables and might remove some significant variables of the data to infer glaucoma progression (Kriegel, Kröger and Zimek, 2009).

1.4 Research Aim

The main aim of this research is to investigate the effectiveness and efficiency of a novel clustering technique on visual field data. The focus of the research is to develop algorithms of SMC that cluster and classify the data with high accuracy towards the prediction of glaucoma deterioration. Additionally, the developed algorithms can be a generic approach of modelling and clustering for other datasets. To achieve this aim, several machine learning techniques are used in these algorithms.

1.5 Research Objectives and Questions

This research is conducted to provide solutions to the research problem and contribute knowledge for medical practitioners and machine learning community. The following objectives are established for this study:-

a. To propose a model-based clustering technique (SMC) on visual field data using compatible algorithms that improves prediction accuracy of visual field loss.

b. To predict and compare visual field loss of the 6NFB and other clusters of visual field locations that significantly represent visual field loss.

c. To describe and summarise the other clusters of visual field that improve predictive accuracy of visual field loss.

Based on the research problems and objectives, research questions are constructed. The work in this research is investigated to the answer to the research questions. The research questions (RQ) are:-

***Research Question* 1 (*RQ*1)**

What is the baseline accuracy for predicting visual field loss across all patients when the visual field points are aggregated to the 6NFB?

***Research Question* 2 (*RQ*2)**

Can visual field data be clustered using model-based clustering to improve the baseline accuracy as in question 1 (RQ1)?

***Research Question* 2-a (*RQ*2-a)**

Can model-based clustering techniques improve on the base line accuracy (c.f. RQ1)?

***Research Question* 2-b (*RQ*2-b)**

Do visual field points arrangement from the model-based agrees with the clinical evidence of glaucoma deterioration?

***Research Question* 3 (*RQ*3)**

Does the choice of heuristic search techniques in the model-based clustering technique effectively improve prediction accuracy of visual field loss?

***Research Question* 4 (*RQ*4)**

Can the SMC approach improves visual field loss prediction using a subset of the patients' records?

1.6 Methodology

This research applied the exploratory experimental research (Franklin, 2005) method by conducting empirical experiments on data using the method proposed in this study: SMC.

Simultaneous Modelling and Clustering

SMC comprises of two techniques from the field of machine learning: clustering and modelling. The data are clustered and modelled using classification techniques. The clustering process is undertaken by performing a systematic search of significant groups of visual field locations (through optimisation/search) within a number of iterations, where initially clusters are formed in random. The range of initial clusters for visual field locations is between two and 52 inclusive. The data (visual field) are clustered based upon the groups of visual field locations by computing the average value of the data for the groups. The clustered data are then modelled using several classification techniques. However, SMC presented in this work can be used in conjunction with any appropriated modelling technique. Classification of the data is measured by percentage accuracy (%). The proposed technique (SMC) is explored by implementing several optimisation methods, strategies of modelling, and fine-tuned procedures in order to cluster the data with high accuracy. The final result of SMC is a cluster arrangement of the visual field locations.

The SMC Experiments

SMC is designed to search an optimal solution (cluster) in the visual field dataset. Several algorithms are developed within this research using optimisation methods and classifiers. Algorithms of SMC are tested by executing experiments on visual field data. Additional experiments are carried out on synthetic data to validate the viability of SMC. Moreover beside 10 fold cross validation, additional two modelling strategies (two-fold and no-fold) are introduced in experiments to further explore SMC. Thereafter, a fine-tuned algorithm of SMC, which addresses the issues discovered in the preceding experiments, is developed and tested using an advanced optimisation method. The processes entailed in this research are depicted in Figure 1-1.

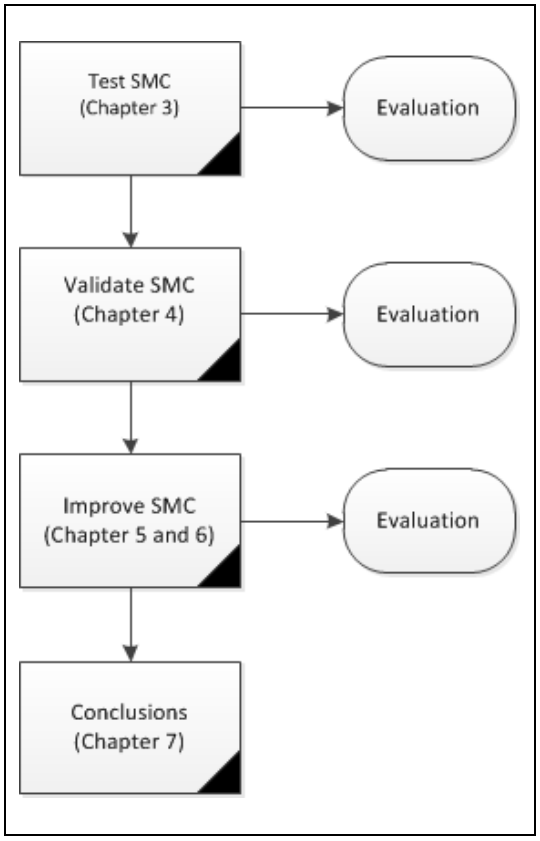


Figure 1-1: Process Diagram for Methodology

1.7 Research Novelty

This research contributes six key areas in analysing visual field data.

1.7.1 A Novel Technique of Clustering

A novel approach of clustering and classification is presented in this research. SMC is a model-based clustering technique and a discrete optimisation problem that searches clustering arrangement of visual field locations for prediction of glaucoma deterioration. Extensive research has been done in prediction using classification. Common techniques of clustering data use data point distance or proximity to cluster data. Meanwhile, this research uses model-based clustering technique where data points are clustered based on classification accuracy performance. The SMC clustering and classification technique implements an optimisation method to search a cluster arrangement in visual field data with high classification.

1.7.2 Generalised Simulated Annealing (GSA)

This research demonstrates a novel development of the GSA method in the discrete problem (SMC). GSA is an advanced annealing method which is proven effective in solving continuous problems (Menin and Bauch, 2017; Xiang, Gubian and Martin, 2017). However, very little work seems have been performed on applying GSA in discrete problems. The GSA method is tailored for SMC to search an optimum cluster arrangement in visual field data. The optimisation method appears to be compatible with the problem as the best prediction accuracy has shown an improvement (compared to the 6NFB) by 2.07% with 10-fold cross validation.

1.7.3 Restricted Growth Functions Generalised Simulated Annealing (RGFGSA)

This work produces a novel algorithm that incorporates the element of Restricted Growth Functions (RGFs) in the existing algorithm of SMC, namely, RGFGSA. RGFs have been used in other research for the purpose of removing redundancy (degeneracy) of solution in a search (Tucker, Crampton and Swift, 2005). Inspired from the literature, RGFGSA is invented and has appeared to be an efficient way (measured using Weighted Kappa) to remove degeneracy in the search of combinatorial problem.

1.7.4 Noisy Fitness Reduction

Noisy fitness is ubiquitous in data analysis especially involving real world problems. Noisy fitness was found in this study and SMC performance has appeared to be deteriorating when using probabilistic optimisation methods such as Simulated Annealing. This was due to tolerance in accepting worse solutions in an optimisation search system. This research contributes to noisy fitness reduction at algorithm level by fine-tuning the optimisation method used in SMC. Application of an RGF in SMC appears to have improvements in the search method to reduce noisy fitness.

1.7.5 Cross Validation in SMC

The study researches a compatible algorithm that can be used in SMC that applied in visual field data and synthetic data. Few optimisation methods and technique of modelling have been applied in order to find the best algorithms that compatible with SMC. The experiments results have shown that modelling strategy has shown slightly improvement to the SMC performance (high accuracy and good clusters). The modelling strategy 2-fold cross validation with 10 repeats has appeared to be an efficient and effective strategy in SMC to cluster and classify data.

1.7.6 Visual Field Datasets

Another contribution of this piece of work that relates to the health-informatics perfective is that the SMC technique could highly improves prediction accuracy of visual field data using segmented data (visual field). Splitting the data into a few datasets is also to avoid bias in data sampling. Modelling the data using SMC on subset datasets has shown predictive accuracy can be predicted up to 90% accurate and this could improve confident to physicians in diagnosis patients.

1.8 Thesis Structure

This thesis is organised as follows:-

Chapter 2 provides a literature review of the main concepts of glaucoma, visual field data, machine learning, methods and techniques that are used within this research. The subject areas: glaucoma and visual field data are first described. Then the topic of machine learning such as classification and clustering, which is suitable for analysing visual field data, is discussed. Next, the relevant techniques from machine learning are demonstrated. Finally, data exploration is presented.

Chapter 3 details a set of baseline experiments that classify visual field data to obtain a baseline accuracy of visual loss. Then, SMC is introduced and tested on visual field data. A number of optimisation methods such as Random Mutation Hill Climbing, Random Restart Hill Climbing and Simulated Annealing are used in the experiments. The results of the SMC experiments are then compared with the baseline accuracy of visual loss.

Chapter 4 demonstrates the use of simulated dataset (synthetic) to validate SMC. Additionally, two modelling strategies are introduced in these experiments to explore the

performance of predictive accuracy of SMC. Moreover, additional experiments using *K*-means clustering technique is presented. The objective of the *K*-means experiments is to obtain the performance of a non-model-based clustering technique as a benchmark for SMC.

Chapter 5 extends the works carried out in Chapter 3 and 4 using an advanced optimisation method: Generalised Simulated Annealing (GSA). GSA has shown promising results in continuous problems, therefore, an algorithm is designed using the optimisation method for the discrete problem (SMC). GSA, which is also a highly parameter-based optimisation method, is advantaged by a few adjustable parameters and additional feature compared to the other methods used in this research in order to extensively search solutions in data. GSA is applied in SMC and experiments are run on both datasets (visual field and synthetic data). The three modelling strategies, which were proven effective in SMC, are remained to be used in these experiments.

Chapter 6 continues the work carried out in Chapter 5. Chapter 6 introduces a Restricted Growth Function based Generalised Simulated Annealing method to solve the degeneracy issues and noisy fitness found in Chapter 5. The new algorithm, RGFGSA, uses a different way of grouping technique (in clustering data) for representing a solution during a search. Additionally within Chapter 6, supplementary experiments are presented to further explore the algorithm and visual field dataset. These extended experiments are aimed at investigating the performance of SMC and RGFGSA method as well as improving classification accuracy (predictive accuracy).

The supplementary experiments manipulate the size of number of moves (small change) within the RGFGSA algorithm. Whilst, the three subsets of the visual field data investigates the performance of SMC and RGFGSA on certain criteria of the visual field data such as early, middle and latest dataset.

Chapter 7 summarises the whole thesis. This chapter recaps outcomes and findings within this research and presents improvements can be made for future work.

Chapter 2

Literature Review

2.1 Overview

Chapter 2 provides a review of research areas including the problem domain and techniques which are used within this research. The chapter is presented in five sections. Section 2.2 discusses visual fields and glaucoma. Vision is one of the main senses in humans that can be deteriorated without systemic symptoms. In this section, the mechanism of glaucoma is discussed. In managing glaucoma, visual field tests are commonly carried out on patients. Section 2.3 discusses the analysis of visual field data. Machine learning is one of techniques that can be applied to analyse visual field data towards discovering new and hidden knowledge. Section 2.4 looks at relevant techniques of machine learning, which are suitable for this study. The emergence of overwhelming volumes of data makes the importance of machine learning in analysing big data such as visual field data. Finally, Chapter 2 is summarised in section 2.5.

2.2 Visual Fields and Glaucoma

The eye is one of the primary human sensory organs and has connections to the visual cortex of the brain. The function of this organ is called vision. The entire scope of vision that can be seen by human eyes is referred to as the visual field. The term visual field refers to a portion of a subject's surroundings that is visible at any one time. It is the area of a scene that can be seen when the eyes focused on a certain object (Jennifer Skillen, 2007). Visual field impairment may happen due to a number of factors such as age, blood pressure, strokes, physical accidents and eye disease (Ananya Mandal, 2012; Klaver *et al.*, 1998). However, there have been reported that visual field loss may due to surgery

such as macular hole surgery (Pendergast and McCuen, 1996). Amongst other diseases that have major impact the visual field of eye are cataracts and glaucoma (Resnikoff *et al.*, 2004).

Visual field loss due to glaucoma cannot be noticed by patients until the damage has become severe. For this reason, glaucoma is often called the "silent blinder" because it is usually a painless process that mostly affects the periphery of vision first (Glaucoma Research Foundation, 2014). According to (Tielsch *et al.*, 1991) 50% of people with glaucoma were not aware of the disease they have. There is also evidence that patients with less than a college education are more likely to be unfamiliar with the disease (Gasch, Wang and Pasquale, 2000).

Glaucoma is a group of eye disorders that have few symptoms or almost none in their early stages, but eventually leads to damage of the optic nerve (the bundles of nerve fibres that carries information from the eye to the brain), which then leads to vision loss or complete blindness (Weinreb and Khaw, 2004). Visual field damage due to glaucoma is irreparable, however the disease can be prevented with early detection and in some cases the progression can be slowed (Weinreb and Khaw, 2004). Therefore, early diagnosis, detection and follow-up are the best ways to manage the disease. There are also series of surgery treatments can be carried out in glaucoma patient such as Trabeculectomy (Weinreb and Khaw, 2004; Khaw, Wells and Lim, 2003). A Trabeculectomy is surgical procedure carried out in patients to reduce the pressure in the eye that causes damage to the optic nerve.

Glaucoma is one of the major causes of blindness in the world after cataracts. Globally, it was estimated that 285 million people across all ages have visual impairment in 2010, of which 39 million were blind (Pascolini and Mariotti, 2012). Meanwhile, it is estimated that 79.6 million people aged over 40 years old will be suffering glaucoma by 2020 with

most of them being Asian and or female (Quigley and Broman, 2006). According to the World Health Organisation, globally up to 75% of blindness (of many causes such as cataracts and trachoma) could be prevented (World Health Organization, 2007). The same report presents that out of 37 million people who sufferings from blindness, 12% were blind due to glaucoma (World Health Organization, 2007).

Visual impairment can seriously impact a person's quality of life. A recent study has found that severity of visual field loss impacts patient quality of life among adult people aged 40 and above in Los Angeles, California, United State (McKean-Cowdin *et al.*, 2007). The quality of life in the study is related to difficulties in driving, distance and peripheral vision activities, and a sense of dependency. Furthermore, this study (ibid.) supports a study in community-dwelling elderly people that visual field loss due to glaucoma was associated with disability, diminished enjoyment of reading and watching television, and a higher risk of falling incidents (Ramrattan *et al.*, 2001).

Glaucoma can occur for a number of reasons. Many cases are caused by a build-up of pressure in the eye (NHS Choices, 2016). Pressure in the eye is termed as intraocular pressure (IOP). This pressure in the eye can damage the optic nerve thus effecting vision. The most common forms of glaucoma are primary open-angle glaucoma (POAG) (Dada, Dave and Mithal, 2009) and primary angle-closure glaucoma (PACG) (Quigley, 2011) which both occur due to the IOP. It was found that PACG was responsible for the vast majority (91%) of bilateral glaucoma blindness in China (P. J. Foster and Johnson, 2001). There were a few studies presented that presence of glaucoma correlated significantly with exfoliation syndrome (Ritch, Schlötzer-Schrehardt and Konstas, 2003). Exfoliation syndrome (XFS) is an age-related disease characterized by the production and progressive accumulation of a fibrillar extracellular material in many ocular tissues (Ritch and Schlötzer-Schrehardt, 2001). The syndrome can lead to both open-angle glaucoma and angle-closure glaucoma. Even though there are kinds of glaucoma diseases have been

classified based on the disease mechanism, there is still limited information available on the exact magnitude of the problem (Thylefors and Negrel, 1994).

2.2.1 The Disease Mechanism

Visual loss due to glaucoma commonly relates to elevation pressure in the eyes resulting from extra fluid (aqueous humour) building up in the front part of the eye. Aqueous humour is the clear watery fluid that fills the space between the cornea and lens of the eye. The fluid is produced by the ciliary body behind the iris and drains from the anterior chamber of the eye through the trabecular meshwork. Failure of the aqueous humour to drain properly from the eye increases intraocular pressure (IOP) and can lead to glaucoma. Eye pressure is measured in mm of Hg (mercury) where normal pressure is between 10 - 21 mm Hg (James C. Tsai, 2016). High IOP is more than 21mmHg (>21mmHg) (Thylefors and Negrel, 1994). IOP damages the optic nerve where the visual information is sent to the brain. This type of glaucoma disease mechanism is called POAG which a progressive optic neuropathy (Weinreb and Khaw, 2004). POAG, which is also known as chronic open-angle glaucoma, is formed by the cornea and iris remains open, but the trabecular meshwork is partially blocked. This causes pressure in the eye to gradually increase. While PACG which is called primary closed-angle glaucoma, occurs when the iris bulges forward to narrow or block the drainage angle formed by the cornea and iris. As a result, fluid unable to circulate through the eye and pressure increases. Angle-closure glaucoma may occur suddenly (acute angle-closure glaucoma) or gradually (chronic angle-closure glaucoma). There are a number of studies that indicate a favorable result on glaucoma patients by reducing IOP, e.g. (Heijl *et al.*, 2002; Anderson, Drance and Schulzer, 1998).

Figure 2-1 (source: (NHS Choices, 2016)) shows the flow of aqueous humour where the blockage of the trabecular meshwork causes IOP and Figure 2-2 (source: (John Berdahl, 2016)) shows the elevation of IOP that damages the optic nerve.

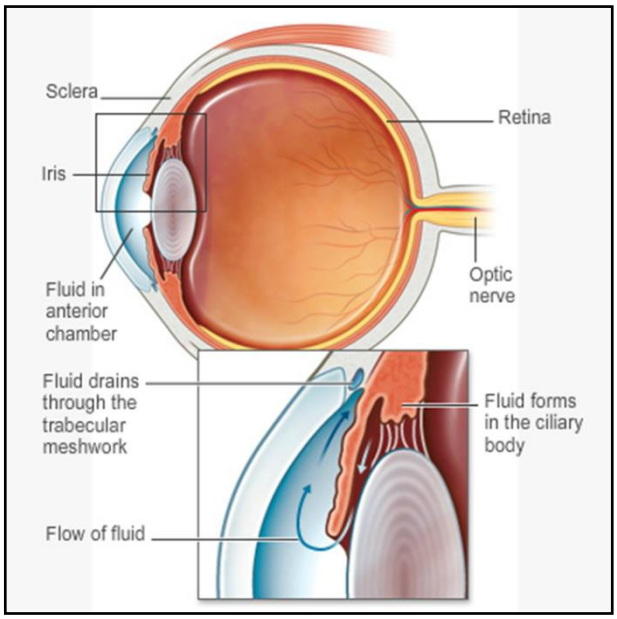


Figure 2-1: Flow of Extra Fluid in the Eye that Causes IOP

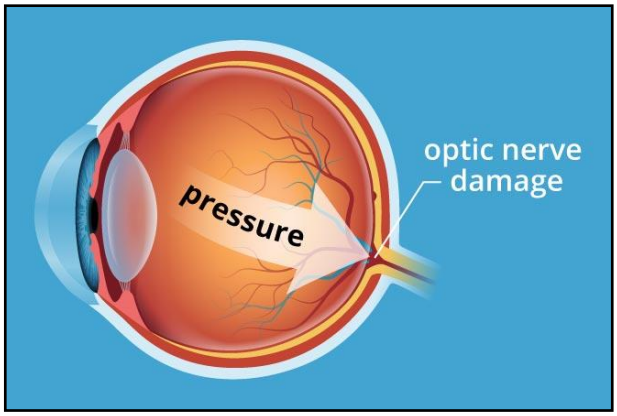


Figure 2-2: Elevated Intraocular Pressure that Damages the Optic Nerve

The optic nerve, which is also referred to as the optic disc, is the whole or a part of the anterior-most part of the optic nerve head or the entire optic nerve head. It is the part of the eye where damage happen due to IOP (Hayreh, 2011) and the damage in the optic nerve is known as optic neuropathy. Progressive optic neuropathy refers to slow degeneration of the vision (Hutchinson, 2012). However the aetiology of POAG is multifactorial which there is no single mechanism could adequately explain the great impact that leads to optic nerve damage (Fechtner and Weinreb, 1994). Occasionally, a patient with glaucoma may have a sudden onset of decreased or blurred vision, eye pain and reddening of the eye, seeing halos around lights, headaches, and sometimes experiencing nausea and vomiting (D. A. Lee and Higginbotham, 2005). Patients with glaucoma see the world differently compared to those with healthy eyes. Figure 2-3 (source : (Christian Nordqvist, 2016)) shows the comparison between the normal vision and glaucomatous vision.

Figure 2-3: Vision Comparison between Healthy Eyes and Glaucoma Eyes

2.2.2 Visual Field System

The visual field is the area from which humans are able to perceive visual signals, when the eyes are in a stationary position and looking straight ahead. In the human brain system, visual field seen by the eyes is mapped to human cortex (Tovée, 2008; Wandell, Dumoulin and Brewer, 2007). Figure 2-4 (source:(Kristie Draskovic, John J. McSoley,

2016)) shows the organisation of visual field system including the connections from the retina to the cerebral hemispheres.

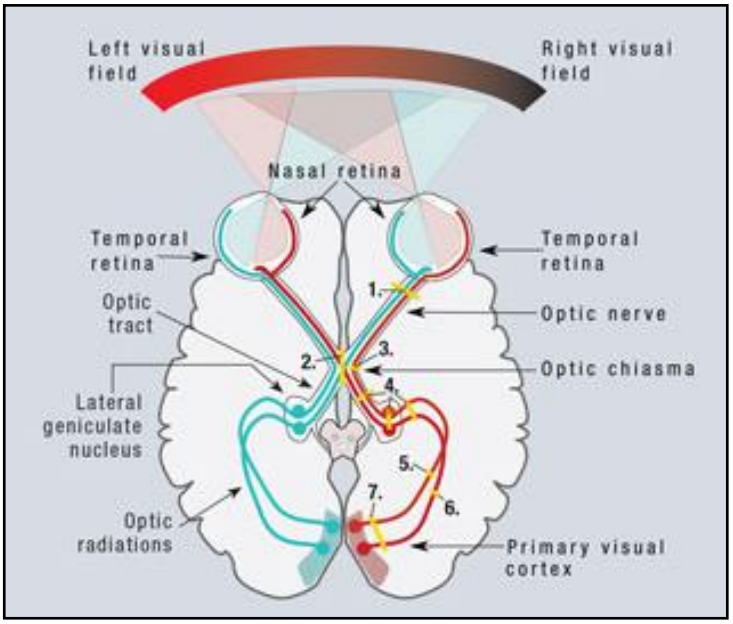


Figure 2-4: The Connections from the Retina to the Cerebral Hemispheres

The normal field vision that can be seen by the human eye extends more than 90 degrees temporally, 60 degrees both nasally and superiorly, and about 70 degrees inferiorly (Heijl and Patella, 2002). These areas of fields are shown in Figure 2-5.

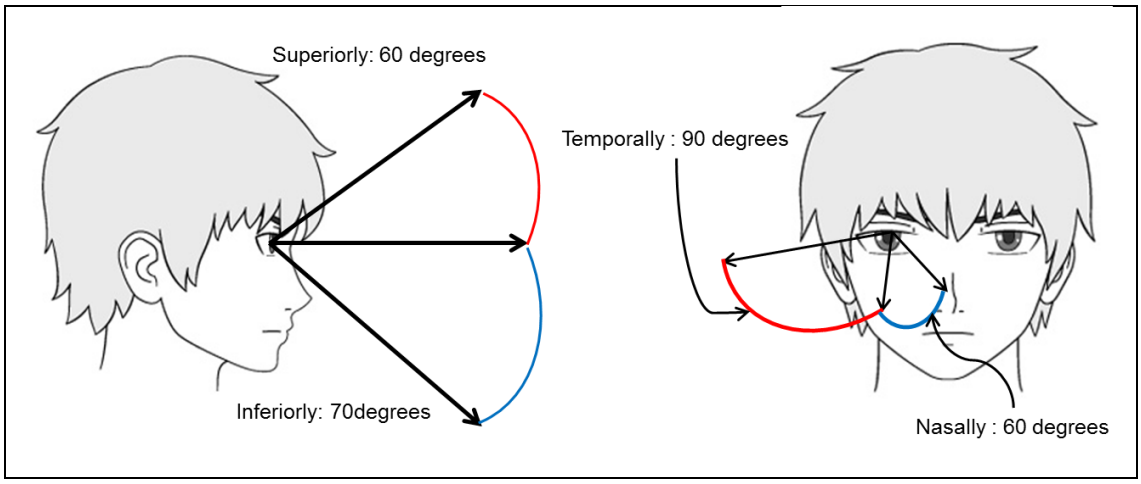


Figure 2-5: Normal Eyes Visual Field Areas

2.2.3 Examination of Glaucoma

Glaucoma deterioration is irreversible, therefore early detection of the disease and early treatment are the best ways to managing the disease from further damages. There are several common tests of glaucoma. According to Glaucoma Research Foundation (Glaucoma Research Foundation, 2013) the following are the common tests conducted to diagnose glaucoma:-

Name of Test	**Test Description**
Tonometry	Measure pressure of the inner eye.
Ophthalmoscopy	Examine the shape and colour of optic nerve to diagnose glaucoma.

Name of Test	Test Description
Perimetry	Visual field test to discern eye vision condition which has been affected by glaucoma.
Gonioscopy	Examine the condition of the iris and cornea angle whether open and wide or narrow or close.
Pachymetry	Measure thickness of the cornea.

Table 2-1: List of Glaucoma Tests

Visual field data are obtained by means of a perimetry test. Perimetry tests capture patient visual sensitivity in the central and peripheral vision (Flammer *et al.*, 1985). Visual field data are commonly used in managing glaucoma (Jay and Murdoch, 1993). With the presence of high technology embedded in the perimetry test system such as Humphrey Field Analyser, visual field data are massively produced by clinicians to diagnose patients. The overwhelming volume and under analysed nature of visual field data available at clinics can be exploited to discover new knowledge by using advanced predictive techniques such as machine learning. This research used a visual field dataset which have been granted the access with permission from the Moorefield Eye Hospital, London for research and development in managing glaucoma.

2.3 Visual Field Analysis

In managing glaucoma disease, a visual field test and its corresponding data are commonly used to diagnose visual impairment in patient. Analysis of visual field data helps physicians to understand the condition and progression of the disease in patients from the complex data. There are few analysis methods can be used towards analysing

visual field data such as statistical methods (for example: pointwise linear regression) and data mining (Turpin *et al.*, 2001). One of visual field data analyses is probability analysis (Spry and Johnson, 2002) which is mostly available in commercial perimetry equipment to perform the analysis automatically. In probability analysis, each individual's pointwise threshold sensitivity estimates or global indices is compared with the available threshold estimates which found among a population of ages matched normal individuals. The analysis allows simple quantification of the likelihood that any test location or index falls outside the distribution found in normal individuals. Another method as presented by Fitzke *et al.* is using linear regression of the luminance sensitivity at each stimulus location, which the method was proven efficient to detect true progression in glaucomatous field loss (Fitzke *et al.*, 1996). Whilst Heijl *et al.* (Heijl, Lindgren and Olsson, 1987) performed statistical analysis on visual field data which the design of the analyses have been based on empirical results obtained from a large number of normal subjects. Since noise in visual field data is inevitable, Liu *et al.* (Liu, Cheng and Wu, 1994) and Henson *et al.* (Henson, Spenceley and Bull, 1997) carried out significant analyses on visual field data to enhance the representation of visual field data to the glaucoma disease and enabling the quantification of the disease condition. For prediction studies on glaucoma, Nouri Mahdavi *et al.* (Nouri-Mahdavi *et al.*, 2007) and Caprili Joseph and his colleagues (Caprioli *et al.*, 2011) carried out prediction of glaucoma deterioration using linear regression and a number of statistical methods. Recently, advanced techniques from artificial life and evolutionary computation have been applied. Pavlidis and colleagues (Pavlidis *et al.*, 2013) modelled the progression of glaucoma using cellular automata modelling of visual field data to understand the mechanism underlying the onset of glaucoma. More research examples of retrospective analyses of visual field data in the research of glaucoma management are:-

Research	Description and Finding	Reference
Predicting glaucomatous visual field deterioration through short multivariate time series modelling	Development of a novel computational method based on genetic algorithm that bypasses the size restrictions of traditional statistical multivariate time series methods. The method was used to predict and model glaucomatous visual field deterioration. The method provided an effective prediction of glaucoma compared to other conventional statistical methods.	(Swift and Liu, 2002)
Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons	At least 25% to 35% retinal ganglion cells loss is associated with statistical abnormalities in automated visual field testing.	(Kerrigan-Baumrind *et al.*, 2000)
Predictive factors for glaucomatous visual field progression in the advanced glaucoma intervention study	Both increasing age and greater IOP fluctuation increase the odds of VF progression by 30% (For each 5-year increment in age and 1-mmHg increase in IOP fluctuation). The higher risk conferred by IOP fluctuation was consistently observed in eyes with and without a history of cataract extraction.	(Nouri-Mahdavi, Hoffman, Coleman *et al.*, 2004)

Research	Description and Finding	Reference
Visual field progression in glaucoma: total versus pattern deviation analyses	Pattern deviation analyses classified approximately 15% fewer eyes as having progressed than did the total deviation analyses. In eyes classified as progressing by both the total and pattern deviation methods, total deviation analyses tended to detect progression earlier than the pattern deviation analyses.	(Artes *et al.*, 2005)
Exploring early glaucoma and the visual field test: classification and clustering using Bayesian Networks	A set of Bayesian Network classifiers was tested on an independent glaucoma dataset, obtaining good results both on pre- (50% sensitivity at 90% specificity) and post- (85% sensitivity at 90% specificity) diagnosis data.	(Ceccon *et al.*, 2014)
Estimating progression of visual field loss in glaucoma	Less than one in three eyes of patients with glaucoma had any progressive field loss. Average changes in threshold sensitivities of less than 1 dB/year could not be detected with seven fields done over 6 years. Larger changes or increased frequency of visual field testing would need to occur before smaller changes could be detected statistically.	(Katz *et al.*, 1997)

Table 2-2: Examples of Research in Analysing Visual Field Data for Glaucoma

Analysing visual field data especially predicting the deterioration of glaucoma is integral in managing this irreversible disease. Sufficient and reliable information about glaucoma deterioration in patients could help medical practitioners to provide appropriate treatments for prevention or slow the disease from deteriorating. For instance, if the progression of the disease is predicted to worsen, increasing the medication or necessary surgery could be taken to avoid the irreversible defect. However, to date, there is no ideal method of analysing visual field data for identifying of glaucoma progression. Therefore, analysing the high dimensional visual field data available from clinical using advanced analytical tools would benefit medical practitioners and help towards providing the best treatments in patients.

2.3.1 Perimetry

Perimetry, which is non-invasive clinical procedure, is the most common type of visual field test performed to capture patients' eyes sensitivity in diagnosing and managing glaucoma (Mikelberg, 1986). The Octopus Perimeter and the Humphrey VF Analyzer (HFA) are the most widespread instruments used by clinicians for assessing patient's visual field. Visual field tests assess the potential presence of visual impairment which could be caused by glaucoma. The purpose of visual field test is threefold (Spry and Johnson, 2002). Detection of early sensitivity deficit of sight is the focal of the test. Secondly, it is used for determination of characteristic spatial pattern of sensitivity loss for differential diagnosis. Thirdly, it works to monitor patient for evidence of progression, stability or improvement of visual field deficits. Discerning visual field sensitivity in patients using the perimetry test is a complex task due to the test variability and sensitivity to the subtle changes in visual field. Distinguishing progressive glaucomatous visual field loss from test variability therefore represents a complex task (Spry and Johnson, 2002). Estimation of threshold sensitivity made using algorithms available in commercial instrumentation is not perfectly repeatable (Bebie, Fankhauser and Spahr, 1976). The

sensitivity of each test location can vary physiologically and become more or less sensitive over time. Due to these issues, there were studies on reducing variability in visual field data using image processing techniques (Fitzke *et al.*, 1995; Crabb *et al.*, 1995).

Data produced by the perimetry tests at clinical are overwhelming. Interpretation of visual field test results by clinician often tricky and a decision to be made based on the result is difficult even though there are cases clear-cut changes in the visual field can be seen easily (Mikelberg, 1986). Therefore, analysing a big data of visual field available at clinicians using machine learning technique is essential to support the clinicians' interpretation of visual field data about the disease (Ceccon *et al.*, 2014; Sacchi *et al.*, 2014). There are three main kinds of perimetry tests namely, Kinetic Perimetry, Static Perimetry and Automated Perimetry (Heijl and Patella, 2002; Haley, 1986).

- **Kinetic Perimetry**

  Kinetic perimetry test captures patient vision sensitivity with moving object stimulus from a non-visible of the visual field to a seeing area along a set meridian.

- **Static Perimetry**

  In static perimetry test, a specific location of the visual field is presented with stimulus and remains constant. The retinal sensitivity is determined by varying the brightness of the specific test location.

- **Automated Perimetry**

  Automated perimetry test offers obvious advantages to the manual test such as the test is more rapid, provides more quantitative information, the stimulus to patients can be presented in random, unpredicted fashion. With the random and unpredictable fashion of stimulus, the patients do not know where the next stimulus is and this improves fixation and test reliability.

The two most commonly used types of perimetry are Goldmann kinetic perimetry and threshold automated static perimetry. Automated static perimetry is the most important clinical tool for measuring visual function outside fovea. Threshold testing is performed to quantify precise visual sensitivity, while suprathreshold testing is used mainly to establish whether visual function is within the normal range (Heijl and Patella, 2002).

In perimetry, eye sensitivity is measured in apostilbs and decibels. Apostilbs is the unit of light intensity while decibel (dB) is the unit of retinal sensitivity (Haley, 1986). Apostilbs and decibels are inversely proportional to each other. The high value in apostilbs is equivalence to the low value in decibels. Figure 2-6 (source : (Haley, 1986)) exhibits the scale of stimulus intensity comparison between decibels and apostilbs.

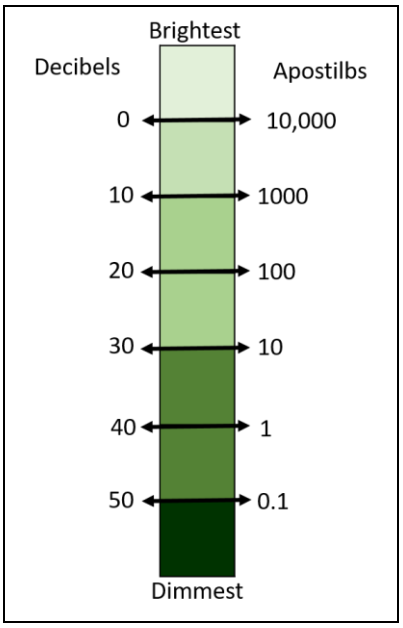


Figure 2-6: Stimulus Intensity Scale of Decibels and Apostibls

2.3.2 Visual Field Test Procedures

There are a variety of methods to measure the visual fields. Visual field testing is performed to one eye at a time, with the opposite eye completely covered to avoid errors. During the test, patient places chin in a type of bowl to support static position and one eye is covered. Lights of various intensity and size are randomly projected around. Patient pushes a button when light is seen. This process produces a computerised map of the visual field drawing the area where each eye can or cannot see. Concluding glaucomatous progression based on visual field results is a tricky task. The procedures of the test which cause patient fatigue, learning influence, and changes in the physiologic state of the eye, affect the visual field results (Boden *et al.*, 2004; Iester *et al.*, 2000; Birch, Wishart and O'Donnell, 1995).

2.3.3 Visual Field Data

Visual field data are used in this research work to test the proposed technique of clustering and classification, which the technique is devised to discover a clustering arrangement and patterns of visual loss in the data. The data are numerical type characterise the sensitivity of the eye to light. A type of visual field data is also depending on the type of test carried out in patient. Typically it is 54 or 76 visual field locations (including the blind spot). Figure 2-7 exhibits the visual field data with the 54 visual field locations including the blind spots which designated are by grey boxes.

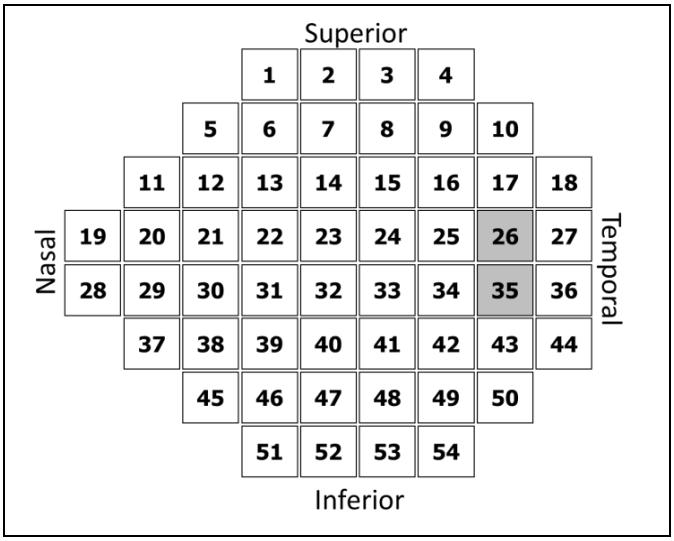


Figure 2-7 : The 54 Locations of Visual Field of the Right Eye (Including the Blind Spots)

The 54 locations of visual field data represent the rounding retina which contains the light sensitive tissue lining the back of the eye. The retina converts the light rays into impulses that transmit through the optic nerve to the brain where images are interpreted. Figure 2-8 shows the 54 locations of the visual field data mapped to the retina for the right eye.

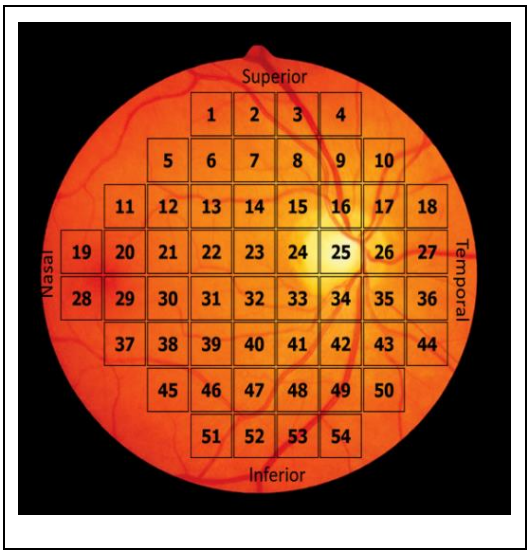


Figure 2-8: The 54 Locations of the Visual Field Data Mapped to the Retina

In the retina there are millions of nerves that connect to the brain via optic nerve (Hayreh, 2011; Bourne, 2006). The optic nerve is located in the back of the eye and also named as the second cranial nerve. The millions of nerves are formed in bundles which transfer visual information from the retina to the vision centers of the brain via electrical impulses. The visual field corresponds to the topographic arrangement of photoreceptors in the eye. In clinical analysis of visual field data, the 54 locations are sectored into six sections (also known as six nerve fiber bundles – 6NFB) to map with the optic nerve fiber bundles based upon physiological trait (Garway-Heath *et al.*, 2000). Each sector is a collection of visual field locations that are grouped together. The sensitivity value of each sector is averaged according to the mapping visual field with the six sectors. Figure 2-9 exhibits the 54 locations of visual field mapped to the 6NFB including the blind spots (location 26 and 35). This 6NFB clustering arrangement is used as the benchmark for the experimental results throughout this study.

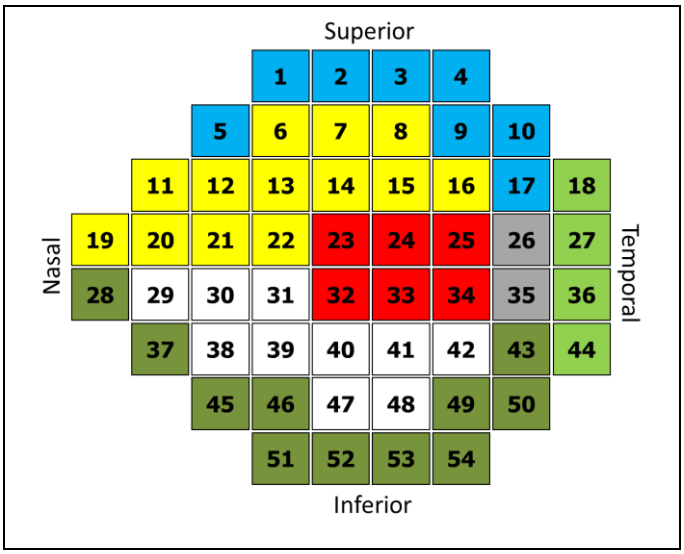


Figure 2-9: The 54 Visual Field Locations Mapped to the 6NFB for the Right Eye

2.3.4 Visual Field Loss

Visual field loss may occur due to glaucoma. Glaucoma affects the optic nerve, which can then cause to blindness if it remains undetected and no appropriate treatments are given (Drance, 1969). Glaucoma usually has few or no initial symptoms until it diagnosed with a suitable test (D. A. Lee and Higginbotham, 2005). Early visual field loss normally occurs predominantly in the inner 20˚ of the visual field and is referred to as paracentral scotomas. The other two visual field defects include peripheral and central nasal steps (as depicted in Figure 2-10) (Mikelberg, 1986).

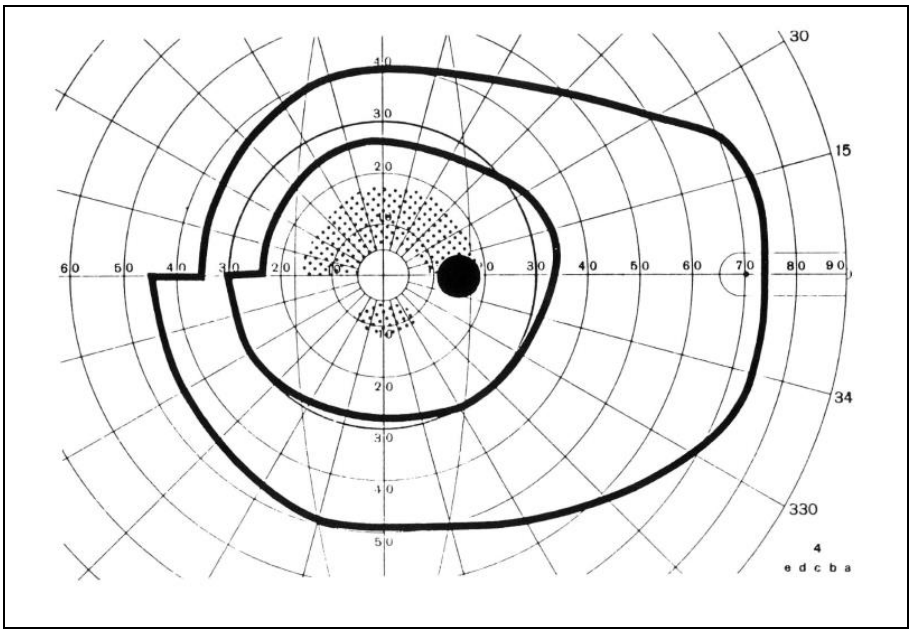


Figure 2-10: Visual Field Record Illustrating Superior Arcuate Scotoma, Inferior Paracentral Scotoma, Central, and Peripheral Nasal Steps

Characteristic of visual field deficit due to glaucoma presents dark areas in the grid of visual field in Humphrey visual field 24-2 test (Figure 2-11 - source:(Pradeep Ramulu, 2017)).

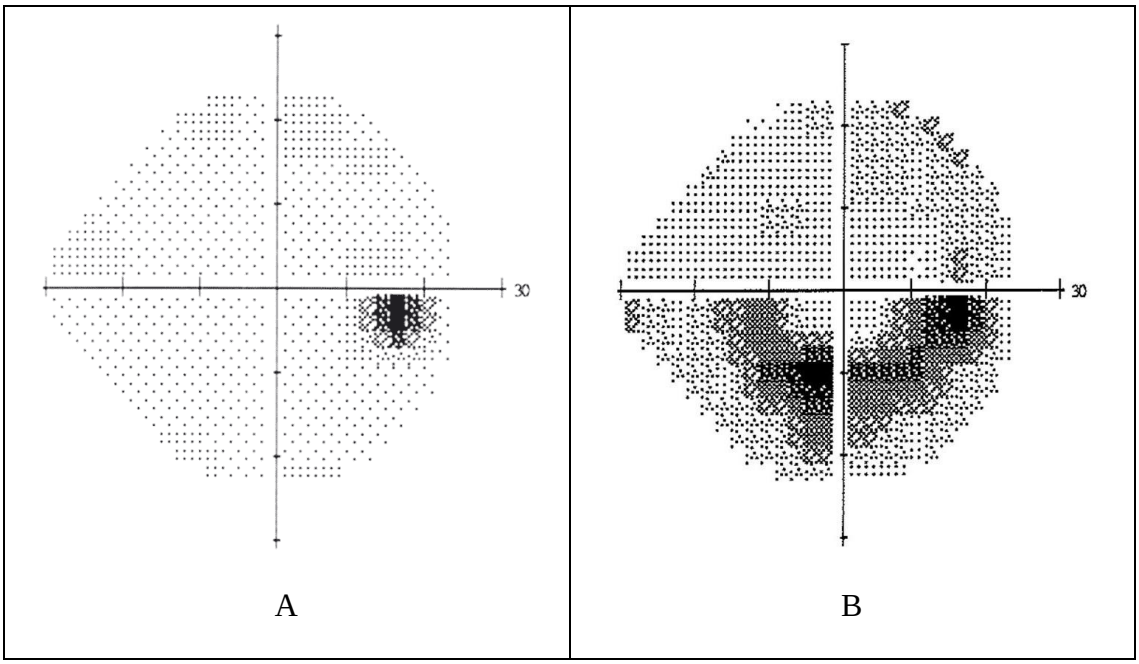


Figure 2-11: Comparison of Health Eye and Glaucoma Eye

(A: Health Eye, B: Glaucoma Eye)

2.3.5 Advanced Glaucoma Intervention Study

There are many ways of quantifying numerical visual field data to diagnose glaucoma. The Advanced Glaucoma Intervention Study or with its convention name AGIS is created to quantify visual field data in diagnosing the severity of glaucoma. For example Nouri-Mahdavi and his colleagues (Nouri-Mahdavi, Hoffman, Gaasterland *et al.*, 2004) performed a prediction study on glaucomatous progression using AGIS score using pointwise linear regression analysis. In AGIS, a quantitative method is developed to assess the test reliability and measure the severity of glaucomatous visual field defects with the 24-2 threshold program of the Humprey Visual Field Analyzer (Advanced Glaucoma Intervention Study Investigators, 1994). The AGIS defect score is derived from the number and depth of clusters of adjacent depressed (measure in decibels) test sites in the upper and lower hemifields and in the nasal area of the total deviation. Figure

2-12 exhibits the printout of visual field that represents the area of defects taken into account in AGIS scoring. Scoring of visual field defect to derive an AGIS severity on a patient is based on the following six precepts:-

a. Defects may occur in the upper or lower hemifield or in the nasal. Defects in one or more subdivision may also occur.
b. Visual field locations above and below of the center of the physiologic blind spot are not taken into account in AGIS scoring. The visual field locations with label 1 and 2 as shown in Figure 2-6.
c. The number of depression and the depth of depression contribute to AGIS scoring.
d. Defect in hemifield is considered when three or more adjacent test sites are affected.
e. The visual field is to be considered defective when the depression of a patient's threshold at a site is sufficiently large.
f. The AGIS score only takes account defects cause by glaucoma.

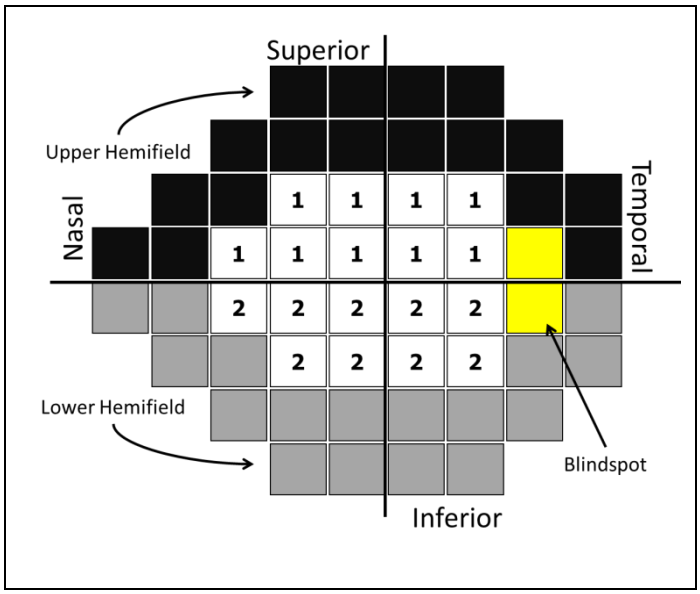


Figure 2-12: Visual Field Layout of AGIS Scoring Area

The AGIS score ranges from 0 to 20 to represent visual loss condition. In the AGIS metric, the score is categories into several categories of the eye condition. The categories are defined as ‘none’, ‘mild’, ‘moderate’, ‘severe’ and ‘end stage’ as shown in Table 2-3.

Agis Score	AGIS Category
0	None
1-5	Mild
6-11	Moderate
12-17	Severe
18-20	End-Stage

Table 2-3: AGIS Metric with Category Label

In this study however applies reclassified AGIS categories (Table 2-4). Sacchi (Sacchi *et al.*, 2014) used the reclassified AGIS category in modelling and predicted visual field data up to 85% accurate with additional variables, which are normally discarded in clinical analysis. While other study done by Sullivan Mee et al. (Sullivan-Mee *et al.*, 2005) used four AGIS category of glaucoma severity by combining the severe and end-stage category into one group. This makes the prediction of AGIS more efficient compared to the five AGIS category.

AGIS Score	AGIS Category	Reclassified Category
0	None	Mild defect
1-5	Mild	
6-11	Moderate	Moderate defect
12-17	Severe	Severe defect
18-20	End-Stage	

Table 2-4: Reclassified AGIS Category

2.4 Machine Learning

Recent years have seen a tremendous growth in the volume of data generated in life sciences and health-informatics due to the emergence of automation and computation tools. This scenario results from the field of big data and machine learning are much discussed on its practicality to the real world (Athey, 2017; Barocas *et al.*, 2017; Landhuis, 2017). The enormous volume of health informatics and biomedical data available at clinics is a research advantage to data scientists to make use the data for pattern recognition analysis and discover new knowledge. Many studies have been carried out on big data analytics especially in health and medical fields recently (Xie, Draizen and Bourne, 2017; Stylianou and Talias, 2017; Park, Chang and Nam, 2017; Finkelstein, 2017). However, there are challenges to data scientists when dealing with high volume, velocity, variety and veracity of data even though the opportunities for making use of this data are vast (R. Fang *et al.*, 2016). Often health informatics and biomedical data are too complex to comprehend the patterns such as gene expression data (Signor, Arbeitman and Nuzhdin, 2016). Therefore, the application of machine learning techniques using a high-performance computing platform would help domain experts to make use of data beyond just recording numbers on digital storage. For example this research applies machine learning on the multidimensional data of visual field which comprises of 54 variables. Machine learning is a branch of artificial intelligence (AI) that gives computers the ability to learn using computational statistics without being explicitly programmed. Machine learning focuses on the development of computer programs (or algorithms) that self-adaptable when exposed to new data (Alpaydin, 2014). Computer programs or algorithms are developed to solve problems such as supervised learning and unsupervised learning (also known as tasks) which is used in this research. Supervised learning, which involves classification analysis, is a task of inferring a function from data based on inputs (independent variables) and output (dependent variable) of the data. The output of supervised learning is a function that classifies data to its label with a certain degree of accuracy of the classification. On the other hand, unsupervised learning performs data

clustering where the label and/or category of each data item is unknown. Similar to classification, clustering (Jain, Murty and Flynn, 1999) infers a function that maps the data into sensible groups based on structures and patters available in the data using a certain approach. A common data clustering technique is *K*-means which finds groups of data based on the distance of data points (Jain *et al.*, 1999). There was a large volume of research have used clustering and classification techniques in biomedical and health-informatics to diagnosis diseases (K. R. Foster, Koprowski and Skufca, 2014; Almeida *et al.*, 2014; Chan *et al.*, 2002; Dreiseitl *et al.*, 2001). Burgansky-Eliash *et al.* (Burgansky-Eliash *et al.*, 2005) have used machine learning techniques such as linear discriminant analysis, support vector machine, recursive partitioning and regression tree on optical coherence tomography (OCT) data for glaucoma detection. As for this research, both clustering and classification are used simultaneously to search the best clusters and improve classification accuracy of the datasets, as well as predicting glaucoma progression. The technique of simultaneous clustering and classification of visual field data is presented in chapter three of this thesis.

2.4.1 Clustering

Cluster analysis is a task of grouping data, where objects in the same group share the common attributes (homogeneity) and dissimilar attributes to other objects in other groups (Rokach and Maimon, 2005). Formally, the clustering structure is represented as a set of subsets $C = \{C_1, C_2, ..., C_k\}$ of $S = \{1,2,...,n\}$, such that $S = \bigcup_{i=1}^{k} C_i$ and $C_i \cap C_j = \varnothing$ for $i \neq j, 1 \leq i, j \leq n$.

Cluster analysis is broadly applied on health-informatics and bio-medical data for dimensionality reduction and subset selection in order to explore the complex data and understanding the structure of data. For example in healthcare, cluster analysis is used to

identify types of depression and to detect pattern in spatial or temporal distribution of disease. Tasoulis et al. (Tasoulis, Plagianakos and Vrahatis, 2004) investigated the application of unsupervised clustering on gene expression microarray data using *k*-windows clustering algorithm due to the challenge of the tremendous data to analyse. Similar to this study, a study on identification of glaucoma in ophthalmology has improved the accuracy of classification for supporting diagnosis of glaucoma using clustering technique in feature selection (ST APOR, 2006).

There are many clustering techniques available in the literatures that may confound practitioners to select suitable technique to solve a problem at hand. However, the selection of clustering technique can be made based on (1) the manner in which clusters are formed, (2) the structure of the data, and (3) sensitivity of the clustering technique to changes that do not affect the structure of the data (Jain *et al.*, 1999). The most common techniques are easy to apply include Hierarchical Clustering and *K*-means clustering (Ronan, Qi and Naegle, 2016; Wu *et al.*, 2008).

Another method of clustering is model-based clustering (Rokach and Maimon, 2005) which has inspired the focal subject of this research. Model-based clustering is an alternative way of clustering techniques besides the distance clustering technique such as *K*-mean and hierarchical clustering. These two types of clustering techniques are largely heuristic and not based on formal models (Fraley and Raftery, 2002). Therefore, instead of using the distance clustering technique, the model-based clustering technique in this study uses data modelling performance as the key indicator of good clusters (this topic is discussed further in chapter 3). SMC is a model-based clustering technique proposed in this research that finds significant objects (variables) to form a cluster in data by using classification performance as an indicator of the quality of the clusters.

Model-based clustering was employed by Yeung *et al.* (Yeung *et al.*, 2001) towards gene expression data and produced clusters of quality comparable to a leading heuristic clustering algorithm. Additionally, the method was found superior performance on synthetic data, which the model selected the right number of clusters (ibid.). Another model-based clustering method studied by Zhang *et al*. (W. Zhang *et al.*, 2017) on the Cancer Genome Atlas data demonstrated an improvement in results compared to existing methods in their simulation studies. Meanwhile, Bose and his colleague (Bose and Chen, 2009) demonstrated their study, which slightly similar to this study, in predicting customer churn using clustering and classification. However, the attributes (or variables) which have been used in classification were obtained from distance-based clustering techniques.

K-Means Clustering

K-means clustering algorithm, which is also used in this research, is a partitioning clustering method by moving data objects from one cluster to another based upon a measurement (commonly measured by data object distance) to determine the similarity of data objects, starting from an initial partitioning. This technique requires a pre-set number of clusters by user. The fundamental of this technique is to compute the distance (the commonly used is Euclidean distance) between data objects and the clusters' centroid and minimising the average distance of the data objects in the cluster. The data objects are randomly assigned to the cluster that has minimum distance from the centroid of the cluster. The best clustering arrangement is formed when each data object is in close proximity to a cluster centroid (**Appendix 2D**).

2.4.2 Classification

Classification is one of the common analyses carried out in research. With classification, the inferred function is produced based on supervised learning that enables a new observation to be classified to which class it belongs. For instance in classifying visual field data, a record of visual field at the time *t* is labelled with the AGIS at the time *t*+1. This means that classifying the previous record of visual field test with the next test's AGIS category is a form of prediction of glaucoma progression. With the classification function learned from this visual field data, a test (visual field test) result in a patient obtained today could predict the AGIS for the next test. Furthermore, the pattern of glaucoma progression in patients can be studied. This example is comprehensively discussed in Chapter 3. There is a large volume of research in disease classification. Asaoka *et al* (Asaoka *et al.*, 2014) demonstrated good results using random forest classifiers for distinguishing visual field of preperimetric open angle glaucoma patient and visual field of healthy eyes. Tucker *et al.* (Tucker *et al.*, 2005) studied the spatial and temporal relationship on glaucomatous visual field deterioration data. Within this study, the structure of visual field characteristics on the onset of glaucoma could be understood. In imaging data analysis, Yang *et al.* employed neural network classifier on imaging data of retina to differentiate the cataract severity (M. Yang *et al.*, 2013). There are numerous classification methods can be employed towards big data for pattern recognition, discovering new knowledge and prediction. However, a choice of classifiers is crucial in order to provide a meaningful solution to a problem because each method can be best for certain problems but not all. One of the approaches of selecting classification method is by performing empirical experiment, which is carried out in this study. The performances of the methods are compared and the best classifier performance is selected (Woods and Laederach, 2017; Shavlik, Mooney and Towell, 1991). Fairly similar to this study, Wan and Freitas (Wan and Freitas, 2017) have employed four classifiers, which are Naïve Bayes, Tree Augmented Naïve Bayes, Bayesian Network Augmented Naïve Bayes and *K*-Nearest Neighbour classifier, to evaluate the performance of features selection in

bioinformatics datasets. However one must be conscious of the misclassification issue if one method dominates all the other methods (Kiang, 2003). The followings are the classification methods used in this work:-

- Decision Tree

The decision tree algorithm has been successfully used in expert systems for capturing hidden knowledge in data. The decision tree classifier is a non-parametric classifier, which can handle both categorical and numerical data. The decision tree classifier offers many benefits such as simple and easy to understand. The classifier can handle a variety of input data that is nominal, numeric and textual, able to process erroneous datasets or missing values and it is a high performance classifier with small number of efforts (Bhargava *et al.*, 2013). Kim and colleagues (Kim, Cho and Oh, 2017) have used decision tree (C.5.0) algorithm to model glaucoma data (based on retinal nerve fiber layer thickness and visual field). The decision tree algorithm J48, which is a simple binary tree C4.5 decision tree for both discrete and continuous attributes (Y. Yang and Chen, 2016; Peng, Chen and Zhou, 2009), was used within this research work and (Fageeri *et al.*, 2017; Patil and Sherekar, 2013) demonstrated that the method has shown better accuracy and more efficiency than Naïve Bayes classifier. In other domain of problem, (Alam and Pachauri, 2017) have devised the J48 tree classifier on credit card fraud detection with 85.5% accurate. A decision tree is developed in two steps that is learn a model using training dataset and test the model using testing dataset to assess the model. In a training dataset, decision tree is constructed by getting significant attributes (variable) to be the root and nodes of the tree. The significant attributes are computed using entropy and information gain (for C4.5 type of tree) as in Equation 2-1 and 2-3.

$$E(C) = -\sum_{i=1}^{c} P_i \log_2 p_i \qquad \textbf{Equation 2-1}$$

Where *C* in Equation 2-1 denotes the computation of entropy of a target variable. Meanwhile, Equation 2-2 is used to compute the entropy of a target variable (*C*) with a condition of an attribute (*X*). While in Equation 2-3 is to obtain gain information which indicates the significant attribute to be a node of the tree.

$$E(C, X) = \sum_{c \in X} P(c)E(c) \qquad \textbf{Equation 2-2}$$

$$Gain(C, X) = E(C) - E(C, X) \qquad \textbf{Equation 2-3}$$

Selection of an attribute to be a node in the tree is based upon the highest value of gain information (Equation 2-3). A branch with entropy of 0 will be a leaf node while a branch with entropy more than 0 needs further splitting. **Appendix 2A** shows the example of entropy computation for constructing a decision tree.

- Naïve Bayes

Probabilistic approaches to classification are a common machine learning task, examples include Bayesian based classifiers (Q. Wang *et al.*, 2007; Androutsopoulos *et al.*, 2000; John and Langley, 1995), the performance of which is good in terms of accuracy. In the recent work, the classifier's predictive performance (accuracy 90.3%) has shown the best in predicting the chemical Ames mutagenicity with 5-fold cross validation (H. Zhang *et al.*, 2017). The Naïve Bayes (NB) classifier is particularly appropriate when the dimensionality of the inputs (variables) is high (Krichene and Krichene, 2017). The choice of NB classifier advantages this work as classification of the real dataset (visual field) can be up to 52 variables (synthetic data 45 variables). NB is a classification algorithm for binary (two-class) and multi-class classification problems. The NB algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set

based on the Bayes's Theorem with the conditional independence assumptions. Bayes' Theorem is formulated as (Leung, 2007):

$$P(c \mid t_i) = \frac{P(c)P(t_i \mid c)}{P(t_i)} \qquad \textbf{Equation 2-4}$$

Where c is a class of the target variable (C) ($c \in C$)

$P(c|t_i)$ is the posterior probability of class (target) given predictor (attribute). $P(c)$ is the prior probability of class. $P(t_i|c)$ is the likelihood which is the probability of predictor given class. $P(t_i)$ is the prior probability of predictor. An example of the Naïve Bayes is discussed in **Appendix 2B**.

- Multinomial Naïve Bayes

Multinomial Naïve Bayes (MNB) implements the naive Bayes algorithm for multinomial distributed data. Unlike NB, MNB takes into account a specific instance of a NB classifier which uses a multinomial distribution for each of the features (attributes). The MNB model is typically used for discrete counts and widely used in text categorisation problems (Frank and Bouckaert, 2006). The main reasons this classifier is so common in research is that fast, easy to implement and relatively effective. The method was found good performance in terms of prediction accuracy in text mining research (Al-Jefri *et al.*, 2017; Diab and El Hindi, 2017; Kale *et al.*, 2017). This classifier was used in this research owing to its efficiency in terms of run time (Matwin and Sazonova, 2012; Rennie *et al.*, 2003), and the visual field data seems compatible with this classifier. Pilot experiments were conducted on visual field data prior to the formal experiments to test the efficiency of MNB and its prediction accuracy. The results were slightly good (74.70% accurate – using the six nerve fiber bundle with 25 samples) and the method was the fastest among the existing methods. The following discusses the computation of the MNB probability for classifying a document (test document) (Kibriya *et al.*, 2004):-

The basic Bayes' Theorem as in Equation 2-4 applies.

Let *C* denotes the class. Then t_i is the assigned document to the class. The $P(c)$ is estimated by dividing the number of documents belong to class *c* by the total numbers of documents. While $P(t_i \mid c)$ is the probability of obtaining a document like t_i in class *c* and calculated as:-

$$P(t_i \mid c) = \alpha \prod_n P(w_n \mid c)^{f_{ni}} \qquad \textbf{Equation 2-5}$$

Where α is a constant that drops out because of the normalisation step. The f_{ni} denotes the count of word *n* in the test document t_i and $P(w_n \mid c)$ is the probability of word *n* given class *C*.

$$\widehat{P}(w_n \mid c) = \frac{1 + F_{nc}}{N + \sum_{x=1}^{N} F_{xc}} \qquad \textbf{Equation 2-6}$$

Let F_{nc} denotes the number of the *n* word in class *C*. The *N* denotes the size of vocabulary or unique words available in the training documents. Where F_{xc} is the count word *x* in all the training documents belonging to class *c*. The normalisation factor $P(t_i)$ in Equation 2-7 can be computed using:-

$$P(t_i) = \sum_{k=1}^{|C|} P(k) P(t_i \mid k) \qquad \textbf{Equation 2-7}$$

Where $P(t_i)$ is probability of a document t_i, $P(k)$ is the probability of the class and $P(t_i|k)$ is the probability of document t_i given class k.

The MNB computation example is explained in **Appendix 2C**.

- Multinomial Naïve Bayes Updateable

Multinomial Naïve Bayes Updateable (MNBU) is another classifier under the Bayesian theorem family that available in the WEKA machine learning toolbox. MNBU works similar to MNB, however this classifier extends the updateable classifier function (George-Nektarios, 2013) in the WEKA tools. The classifier applies incremental classification model that can learn using one instance at a time. Many studies have applied updateable classifiers with promising results (Ajaz and Hussain, 2015; Balachandran and ANITHA, 2014; Balachandran and Anitha, 2012). Thus, it is worthwhile to employ this classifier for the sake of testing the proposed approach of clustering and modelling at very cost-efficient of algorithm's runtime.

2.4.3 Cross Validation

Cross validation (Browne, 2000) is a technique of evaluation of predictive model by partitioning data into training and testing sets. A training set of data is used to build a model while testing set of data is used to test the model. Predictive models in this research is evaluated using *K*-fold cross validation. *K*-fold cross validation randomly partitions *K* subsets of the original dataset equally. *K*-1 subsets of the dataset (training dataset) are used for building a model. Meanwhile, the remaining single subset (testing dataset) is used to test a model. This process of training and testing predictive model is repeated *K* times, where each individual fold is used as the testing dataset. Then the total estimation of predictive model of *K* times is averaged to produce a single estimation. The advantage of *K*-fold cross validation is that it provides unbiased estimate of models (Bengio and Grandvalet, 2004). Additionally, this approach of modelling allows all data points to have an equal chance of being used in the training and testing datasets in constructing a model.

2.4.4 Heuristic Search

Heuristic search is one of the branches of computer science that solves problems (optimisation problem) by finding an approximate solution. The method generates a possible solution from a search space (Definition 2.1) for a certain problem, and a possible solution is assessed based on the defined goal (commonly termed a fitness function). In the context of computer science and mathematics, a solution to a problem is translated into a fitness function (Definition 2.1 and Definition 2.2). An application of heuristic search can be a quick way for problem solving and guarantees to find a good solution in reasonable time. However, the solution produced is might not be the best solution as there is an element of stochastic in the search. This research applies heuristic search techniques in classification and clustering of visual field data finding the possible clusters of visual field location that represent the significant vision deterioration in patient due to glaucoma. Moreover, instead of using synthetic data, some of the paradigm problems in computer science that can be used to test a heuristic search algorithm are the Travelling Salesman Problem (TSP) (Hoffman, Padberg and Rinaldi, 2013) and Bin Packing Problem (Lodi, Martello and Vigo, 2002).

Definition 2-1 Search Space

The search space is the feasible region of a problem space that has all sets of possible solutions.

Definition 2-2 Fitness Function

A fitness function is a single objective function that summarises the aim of certain solution to a problem by means of a merit of measurement.

Definition 2-3 Fitness Value

A fitness value is the value generated from a fitness function.

The performance of a heuristic algorithm is determined by the efficiency and effectiveness of an algorithm. The effectiveness of an algorithm refers to the quality of solutions found by the algorithm in terms of the practicality of the solution to a problem. In optimisation problem solving, a search may have potential to obtain local optima solutions (Definition 2.4) which this affects the efficiency of a heuristic search. Moreover, efficiency in a heuristic search also refers to runtime of the search algorithm to obtain a global optimum (Definition 2.5). This is termed as convergence in this research (further discussed in Chapter 3).

Definition 2-4 Local Optima

Local optima are defined as the relative best solutions within a neighbour solution set.

x is a local optimum of $f(X) \rightarrow \mathbb{R}$, where $x \in \forall X$ when $f(x)$ is least better than $f(x')$ as x' is the best solution and $x' \in \forall X$.

Definition 2-5 Global Optimum

Global optimum is the best solution among all local optima solutions.

x' is a global optimum of $f(X) \rightarrow \mathbb{R}$, where $x' \in \forall X$, when $f(x')$ is the best among $f(\forall X)$.

Another property that heuristic search has is a fitness function. A fitness function is used in heuristic search to evaluate every solution being search in the search space. It is a function that measures the solution either to maximise or minimise the fitness value. A simple example of a fitness function for a maximisation optimisation problem is the

function to obtain the shortest route that a salesperson could travel among the cities in the TSP problem (a sum of distance of the TSP route). Meanwhile, this research work is a maximisation problem which the fitness function is the model accuracy for classifying the visual field data (further discussed in chapter 3 of this thesis). Four optimisation methods are used in this work: Hill Climbing, Simulated Annealing, and Generalised Simulated Annealing.

- **Hill Climbing**

  Hill Climbing is a simple greedy search method that iteratively searches (iteratively improves a solution (Kirkpatrick, Gelatt and Vecchi, 1983)). The analogy of hill climbing methods is when one walks upward to a highest point in the thick fog. The walk stops when the highest point is reached. However, there could be possibly other points that higher than the point has been reached. Similar to this process in hill climbing method, possible solutions are searched starting from a random point in a search space and only an improvement to a solution is accepted. Similar to climbing a hill, any downward from the current solution in the exploration is unacceptable. Once the search point reaches the highest state of solutions (peak of the hill), the search may stop. The inability of the method to accept worse solution during the search consequences the method results a local optimum solution (Selman and Gomes, 2006; Tovey, 1985). Nevertheless, there were numerous of research has devised this optimisation method due to its advantages such as faster in terms of time (Lang, 2016; Hoffmann, 2001). Two types of Hill Climbing methods used in this research are Random Mutation Hill Climbing (RMHC) and Random Restart Hill Climbing (RRHC). They are used in this research work due to simple and fast algorithms to implement. RRHC is an improvement on the RMHC algorithms which is simple and easy. However, RRHC generates few random points in the search space for the start of the search in order to avoid local optima. Owing to this capability of the search method, the RRHC method was chosen in this research (O'Neil and

Burtscher, 2015). A study performed by (Charnay, Lachiche and Braud, 2015) on the integration of complex aggregates in the construction of logical decision tree to address relational data mining tasks using RRHC has resulted a promising result on both real and artificial data. An empirical study (Jacobson and Yücesan, 2004) has proven that random restart local (RRHC) search can outperform Simulated Annealing method, which is also used in this research work, given a sufficiently large number of restarts executed. Algorithm 2.1 and 2.2 delineates the steps of RMHC and RRHC algorithms which designed in the context of this research.

Algorithm 2-1 : Random Mutation Hill Climbing Algorithm	
Input:	A random initial solution Number of iterations
1	For i=0 to iteration-1
2	Evaluate the current solution
3	Obtain a random new solution (small perturbation to the current solution)
4	Evaluate the new solution
5	If new solution > current solution
6	Keep the new solution configuration as the current solution
7	End if
8	End for
Output:	Return current solution

Algorithm 2-2: Random Restart Hill Climbing Algorithm	
Input:	A random initial solution Number of restarts Number of iterations
1	For i=0 to restart-1
2	Perform Algorithm 2.1 (RMHC) for a number of iterations
3	Obtain result for each restart
4	End for
Output:	Return current solutions of each restart

- **Simulated Annealing**

Simulated Annealing (SA), which was invented in 1983 (Busetti, 2003), is another optimiser appears to be the most common method used in complex optimisation problems to overcome local optima in search (Romeijn and Smith, 1994; Goffe, Ferrier and Rogers, 1994; Dekkers and Aarts, 1991). It is the third most popular metaheuristic technique by number of publications on Google Scholar (Varty, 2017) and commonly used to solve continuous problems. This research however applies SA in a discrete problem where the best clustering arrangements (combinatorial of elements in clusters) of visual field are searched.

This statistical mechanics method approach inspired by the process of metallurgy which annealing is referred to as tempering certain alloys of metal, glass, or crystal by heating above its melting point, holding its temperature, and then cooling it adequately slow until it solidifies into a perfect crystalline structure. Many other algorithms including SA were commonly tested their effectiveness and efficiency in the paradigm problem such as TSP (Kirkpatrick *et al.*, 1983). Meer K. (Meer, 2007) in a study has shown that SA outperformed the Metropolis Algorithm for solving a TSP problem. Another study (L. Fang, Chen and Liu, 2007) tested on TSP has improved the shortcoming of the Particle Swarm Optimisation (PSO) algorithm which suffers from local optima by combining SA approach with the algorithm. In this study, this hybrid algorithm (PSO-SA) solved the TSP problem very well as compared to other existing methods such as genetic algorithm. Meanwhile, a study on real data (K. Bryan, Cunningham and Bolshakova, 2005), which is bi-clustering gene expression data, has shown that SA algorithm performed better than the method used by Cheng et al. (Cheng and Church, 2000) on bi-clustering gene expression. A recent study (Z. Yang, Huo and Fang, 2017) has devoted SA to finding a number of clusters automatically in data with a consistent performance in experimental results.

SA behaves slightly similar to RMHC and RRHC which is always looking for a better solution to a problem in a search. However based on the assumption that accepting a worse solution at a certain tolerance level would help avoid local optima, SA algorithm computes probability for accepting the worse solution. Moreover, the algorithm has artificial temperature that used to control the acceptance probability. Acceptance probability and artificial temperature in SA are translated in Equation 2-8 and Equation 2-10 respectively (Swift *et al.*, 2004; S. P. Brooks and Morgan, 1995).

$$\Pr(f', f, T_i) = e\left(\frac{-\Delta E}{T_i}\right) \qquad \textbf{Equation 2-8}$$

In Equation 2-8, e corresponds to exponential constant with value approximately 2.718. T_i is the temperature at iteration i and ΔE is the different value of the new fitness and the current fitness in the search. Suppose f is defined as the fitness value for the current solution and f' is the fitness value for the new solution. Therefore, Equation 2-9 is the different of fitness value.

$$\Delta E = f - f' \qquad \textbf{Equation 2-9}$$

$$T_{i+1} = T_i \lambda \qquad \textbf{Equation 2-10}$$

The cooling rate (λ) is a constant that decreases the temperature gradually each iteration towards a very small value (0.001). If cooling is sufficiently slow, the global minimum will be reached (Brünger, Krukowski and Erickson, 1990). Algorithm 2.3 sketches the SA algorithms which specifically applied in the context of this research.

Algorithm 2-3: Simulated Annealing Algorithm	
Input:	A random initial solution Number of iteration Initial temperature Constant cooling rate
1	For i=0 to iteration-1
2	Evaluate the current solution
3	Obtain a random new solution (perturbation in the current solution)
4	Evaluate the new solution
5	If new solution > current solution
6	Keep the new solution configuration as the current solution
7	else
8	Compute probability
9	If probability > random number
10	Keep the new solution configuration as the current solution
11	End if
12	Compute new temperature (with cooling rate)
13	End for
Output:	Return current solution

- **Generalised Simulated Annealing**

Generalised Simulated Annealing (GSA) is another algorithm under the family of annealing methods. The algorithm is enhanced from the SA method, which is the improved version of Classical Simulated Annealing (CSA) and Fast Simulated Annealing (FSA) (Tsallis and Stariolo, 1996). The SA algorithm, which has been described in the previous section, is the CSA algorithm. The CSA algorithm is fairly slow in terms of convergence (Definition 2.5) (Xiang *et al.*, 1997). Slow convergence in CSA is attributed to the nature of visiting distribution, which uses Gaussian distribution (local search distribution). Therefore, another version of SA was invented: Fast Simulated Annealing (FSA). Szu and Hartley (Szu and Hartley, 1987) proposed FSA, which used Cauchy Lorentz visiting distribution (semi-local search distribution). The FSA method is quicker at finding an optimum solution compared to the CSA method (ibid.). This is because of the long jumps (high size of change in

perturbation of a current solution). However, the long jumps can occasionally be quite long. Moreover, cooling of the temperature in the method is much faster, which makes the search more efficient. Although both algorithms perform well in complex data in finding the global optimum, it is imperative to have a very efficient algorithm in terms of early convergence (Definition 2-6) in a search. Many studies on convergence of algorithms have been conducted to solve complex data (Nakamura and Hong, 2016; H. Lee *et al.*, 2016; Bonyadi and Michalewicz, 2016; Rudolph, 1994). One of the studies examines GSA algorithms, which are able to overcome the shortcoming of the aforementioned methods of annealing.

Definition 2-6 Convergence Point

The convergence point of an algorithm is the point (time or iteration number) where the fitness of the current solution no longer improves.

The convergence point is used in this research to measure the efficiency of an optimisation method applied in SMC. Reasonably early convergence point in a search of optimising visual field data indicates the efficiency of the method where the search is run with a very large number of iteration (100,000 iterations).

GSA is a successful continuous optimisation method in solving problems with multiple local optima (also known as non-convex) (Xiang *et al.*, 2017; Vizarim *et al.*, 2017; Taylor and Mildenberger, 2017; Menin, Martinez and Costa, 2016). A recent study (Fukui, Sato and Takahashi, 2016) has devised GSA to find an optimum value of the model they used in estimating style weights of mutual funds. Bagheri and colleagues (Bagheri *et al.*, 2017) applied a GSA algorithm in their study to find an optimum value (maximisation problem) for a function.

However, to date, there has not been much investigation of the GSA optimisation method in combinatorial or discrete problems especially in health-informatics and biomedical data. The high-dimensional data of visual field in this work present a challenge to any existing heuristic search methods to solve this combinatorial problem. The positive findings from the previous studies on continuous problems using GSA inspire this research to adopt the method in Simultaneous Modelling and Clustering (SMC).

The GSA method has the same properties as the SA method for acceptance probability and artificial temperature (Equation 2-11 and Equation 2-12 (Tsallis and Stariolo, 1996)). However, unlike SA, GSA takes into account a certain degree of change to the current solution in order to derive a new solution. This is called the visiting distribution (Equation 2-13 (Tsallis and Stariolo, 1996)). The visiting distribution allows a certain large scale of perturbation to the current solution in the early search, and the scale of perturbation reduces towards the end of the search. This behaviour in GSA supports an extensive exploration in a search space. Furthermore, GSA introduces two parameters which calibrate the GSA search. The parameter of acceptance index and visiting index are denoted by q_A and q_V respectively where the value of q_A and q_V are calibrated depending upon the nature of the problem. Similar to SA, the acceptance probability is a rule used to ascertain that a new worse solution in a search is acceptable as a stepping-stone to a global optimum solution.

$$p_{qA}(x_t \to x_{t+1}) = \frac{1}{[1+(q_A-1)(\Delta E)/T_{q_A}^{A}]^{1/(q_A-1)}} \qquad \textbf{Equation 2-11}$$

ΔE denotes the different of fitness and $T_{q_A}^{A}$ is the acceptance temperature which is computed using Equation 2-12. The temperature is also used in visiting distribution in Equation 2-14.

$$T_{q_v}^V(t) = T_{q_v}(1)\frac{2^{q_v-1}-1}{(1+t)^{q_v-1}-1} \qquad \textbf{Equation 2-12}$$

$$a = \frac{[T_{q_v}^V(t)]^{-D/(3-q_v)}}{\left\{1+(q-1)\frac{(\Delta x_t)^2}{[T_{q_v}^V(t)]^{2/(3-q_v)}}\right\}^{1/(q_v-1)+(D-1)/2}}$$

$$g_{qV}(\Delta x_t) = \left(\frac{q_V-1}{\pi}\right)^{D/2}\frac{\Gamma\left(\frac{1}{q_V-1}+\frac{D-1}{2}\right)}{\Gamma\left(\frac{1}{q_V-1}-\frac{1}{2}\right)}\times a \qquad \textbf{Equation 2-13}$$

The detailed procedure of the GSA method is described in Algorithm 2-4.

Algorithm 2-4: Generalised Simulated Annealing Algorithm	
Input:	A random initial solution Number of iteration Initial temperature Parameter qa, qv
1	For i=0 to iteration-1
2	Evaluate the current solution
	Compute visiting distribution for degree of change
3	Obtain a random new solution as Equation 2-15 (perturbation in the current solution by the degree of change computed)
4	Evaluate the new solution
5	If new solution > current solution
6	Keep the new solution configuration as the current solution
7	else
8	Compute probability
9	If probability > random number
10	Keep the new solution configuration as the current solution
11	End if
12	Compute new temperature
13	End for
Output:	Return current solution

2.4.5 Weighted Kappa Statistics

The Weighted Kappa Statistic (WK) is widely used in research to measure a level of agreement or disagreement between two opinions especially in medical studies (Cao *et al.*, 2016; Hill *et al.*, 2016; Viera and Garrett, 2005). This measurement is important in research to see whether opinions suggested by experts are within a standard or benchmark. In relation to this research, WK is devised to compute an agreement between two cluster arrangements, which is the solution found by the approach proposed in this study (SMC). Each of the visual field location numbered 1 to 52 are paired. Then the agreement and disagreement between the two clustering arrangements are checked for each pair of visual field locations. A number of agreements and disagreements from the

two clustering arrangements are counted, and WK is computed according to Equation 2-14 (Viera and Garrett, 2005). The WK agreement level is ranged from -1.0 to 1.0 as depicted in Table 2-5.

Kappa (k)	**Agreement Strength**
$-1.0 \leq k \leq 0.0$	Very Poor
$0.0 < k \leq 0.2$	Poor
$0.2 < k \leq 0.4$	Fair
$0.4 < k \leq 0.6$	Moderate
$0.6 < k \leq 0.8$	Good
$0.8 < k \leq 1.0$	Very Good

Table 2-5: Weighted Kappa Statistic Metrix

$$WK = \frac{P_{observed} - P_{expected}}{1 - P_{expected}} \qquad \textbf{Equation 2-14}$$

Table 2-6 shows the contingency table for observation and expected where agreement and disagreement of two opinions or decisions (in this study is algorithms and the 6NFB) are counted.

Decision	**Yes_{ex}**	**No_{ex}**
Yes_{ob}	$Yes_{ob}Yes_{ex}$	$Yes_{ob}No_{ex}$
No_{ob}	$No_{ob}Yes_{ex}$	$No_{ob}No_{ex}$

Table 2-6: Contingency Table for Observation and Expected

Where $P_{observed}$ is the proportion of similar decision of two clusters. While $P_{\exp ected}$ is the proportion of expected that counts the dissimilarity decision between two clusters.

$P_{observed} = \frac{Yes_{ob}Yes_{ex} + No_{ob}No_{ex}}{N}$, where N is number of pairs. While $P_{\exp ected} = \frac{a+b}{N^2}$ where,

$$a = (Yes_{ob}Yes_{ex} + Yes_{ob}No_{ex}) \times (Yes_{ob}Yes_{ex} + No_{ob}Yes_{ex})$$

$$b = (No_{ob}Yes_{ex} + No_{ob}No_{ex}) \times (Yes_{ob}No_{ex} + No_{ob}No_{ex}).$$

The example of computing WK is demonstrated as follows. Suppose five (5) objects are to be clustered by two algorithms. In order to find similarity between the two algorithms, WK is computed. The clusters formed by the two algorithms in this example are as follows:-

Clusters of Algorithm A: {1,2,5}{3,4} *Clusters of Algorithm B*: {1,2}{3}{4,5}

Each of the objects is paired and the decisions of the algorithm A and B are checked whether they agree the paired objects are in the same cluster.

Pair	Algorithm A	Algorithm B
1 and 2	Yes	Yes
1 and 3	No	No
1 and 4	No	No
1 and 5	Yes	No
2 and 3	No	No
2 and 4	No	No
2 and 5	Yes	No
3 and 4	Yes	No
3 and 5	No	No
4 and 5	No	Yes

Table 2-7: Count of Similarity and Dissimilarity of Algorithms Clusters

From the count of similarity and dissimilarity, the table is simplified as follows,

Decision	Yes_{ex}	No_{ex}
Yes_{ob}	1	3
No_{ob}	1	5

Table 2-8: Contingency Table for the Decision of Algorithm A and B

From Table 2-8, values for proportion observed and proportion expected are computed.

$$P_{observed} = \frac{1+5}{10}$$

$$P_{observed} = 0.6$$

$$a = (Yes_{ob}Yes_{ex} + Yes_{ob}No_{ex}) \times (Yes_{ob}Yes_{ex} + No_{ob}Yes_{ex})$$

$$a = (1+1) \times (1+3)$$

$$b = (No_{ob}Yes_{ex} + No_{ob}No_{ex}) \times (Yes_{ob}No_{ex} + No_{ob}No_{ex})$$

$$b = (1+5) \times (3+5)$$

$$P_{\exp ected} = \frac{a+b}{N^2}$$

$$P_{\exp ected} = \frac{8+48}{100}$$

$$P_{\exp ected} = \frac{56}{100}$$

$$P_{\exp ected} = 0.56$$

According to Equation 2-15, WK is computed as follows:-

$WK = \frac{0.6 - 0.56}{1 - 0.56}$ $WK = 0.091$. Therefore, in this example, algorithm A and algorithm B have poor agreement of clustering arrangement.

2.5 Summary

Chapter 2 provides a literature review of the domain of glaucoma, the second leading cause of blindness globally. Even though the cause of glaucoma is multifactorial, elevated intraocular pressure in the eyes is one of the common roots. Elevated intraocular pressure damages the optic nerve, which is the point where visual information is transmitted to the brain. Glaucoma is an irreversible disease and there is no cure established as confirmed by many clinical experts (Rizzo *et al.*, 2017; Sivapriya and Latha, 2017; Nayak *et al.*, 2009). Hence, early detection and prevention is the only way to avoid total loss of vision due to glaucoma. Therefore, research on this area is important to detect and predict the disease in patients in order to eradicate the disease at an early stage.

In clinical practices, visual field tests are commonly used to manage glaucoma. The data are collected by means of perimetry. With the advent of high digital technology, there is an abundance of visual field data available in clinics. However, to date, there is no recognised gold standard for analysing visual field data to manage glaucoma. A standard pattern of visual field locations that corresponds to the visual field deterioration has yet to be found. Therefore, finding hidden patterns in the data makes this research significant for clinical experts. In the current practice, the six nerve fiber bundles are used in managing glaucoma using visual field data. However, further research on finding the other possible significant bundles (clusters) is needed as there is no gold standard to predict glaucoma deterioration using visual field data. This research proposes a novel approach of finding significant clusters that might be available in the data. With other possible clusters to better predict the AGIS score, this research would benefit the experts

to comprehend the progression of glaucoma based on the patterns of visual field deterioration. This research can be materialised by devising machine learning techniques on the data to predict the progression of glaucoma in patients using the AGIS score.

The exponential growth of visual field data available on digital storage has resulted in the development of big data analysis and algorithms. Leveraging the benefits of big data analysis in health-informatics and biomedical changes a way of discovering new knowledge, diagnose a disease, and a new way of decision making (Zhao *et al.*, 2017; Y. Wang and Hajli, 2017; Gagneur *et al.*, 2017; McAfee and Brynjolfsson, 2012). With the visual field data in hand, the importance of applying machine learning on the data was discussed based on the literatures. The existing machine learning techniques available in the literatures such as clustering and classification, which are to be used in the proposed approach of this research, have shown promising results. Furthermore, the widely used optimisation techniques in research such as Hill Climbing, Simulated Annealing and Generalised Simulated Annealing which are the focal part of this research are to be explored in this work. Optimisation is one of the main processes entailed in the proposed approach. Developing efficient algorithms using the proposed novel approach in this study that find solution and patterns on big data is a contribution to the data science community

Chapter 3

Data Exploration

3.1 Overview

Data exploration is the first step in data analysis and typically involves summarising the main characteristics of a dataset. This process helps data analysts to comprehend the nature of dataset by getting statistical values and visualising the dataset. Two datasets are used in this research: real data (visual field) and synthetic data. The visual data are obtained with permission from Moorefield Eye Hospital London. There is no patients' information available in the data, and the data (patient identification number) are anonymised in line with the NHS and Brunel University research code of conduct. Meanwhile the synthetic data, which is multivariate normal generated, is used to validate the proposed approach (will be further discussed in Chapter 4).

3.2 Visual Field Data Exploration

Visual field data which are used in this research work were collected by means of computerised automated perimetry (24-2 test) and consist of 52 visual field locations. The data have 13,739 records of visual field test (the right eye) from 1,580 patients. Every patient has a few records of visual field test. Figure 3-1 shows a histogram of the number of visual field versus the number of patients (with the corresponding number of tests). The AGIS score, which indicates the severity of glaucoma progression, was provided with the data for each test record. As discussed in section 2.2, the AGIS score is classified into three categories: mild (coded as '0'), moderate (coded as '1') and severe (coded as '2').

There is one patient with maximum 42 records and 296 patients with a minimum of two records of visual field test. The patient with 42 records was classified as moderate defect of glaucoma in the last three recent tests. The patient has 34 (80.95%) tests with mild defects and 8 (19.05%) tests with moderate defects.

Meanwhile those 296 patients with only two records of visual field test accumulate 592 records. From these records, 376 (63.51%) records are with mild defect, 102 (17.23%) records with moderate defect, and 114 (19.26%) records with severe defect. Based on the percentage of AGIS score records from the highest and lowest number of tests in the data, the experiments in the next proceeding chapters use all data and sampled data. All data experiments include all the records from the data. Whilst, sampled dataset consists of one record from every patient in the data (totalling 1580 records). The reason being (sample datasets) is that to avoid bias that may exist in all data. Moreover, another approach of sampling is presented by splitting the data into three datasets which include early record, middle record and latest record. This approach of sampling is another way used in this study to avoid bias in simple sampled data. These split datasets are experimented in Chapter 7.

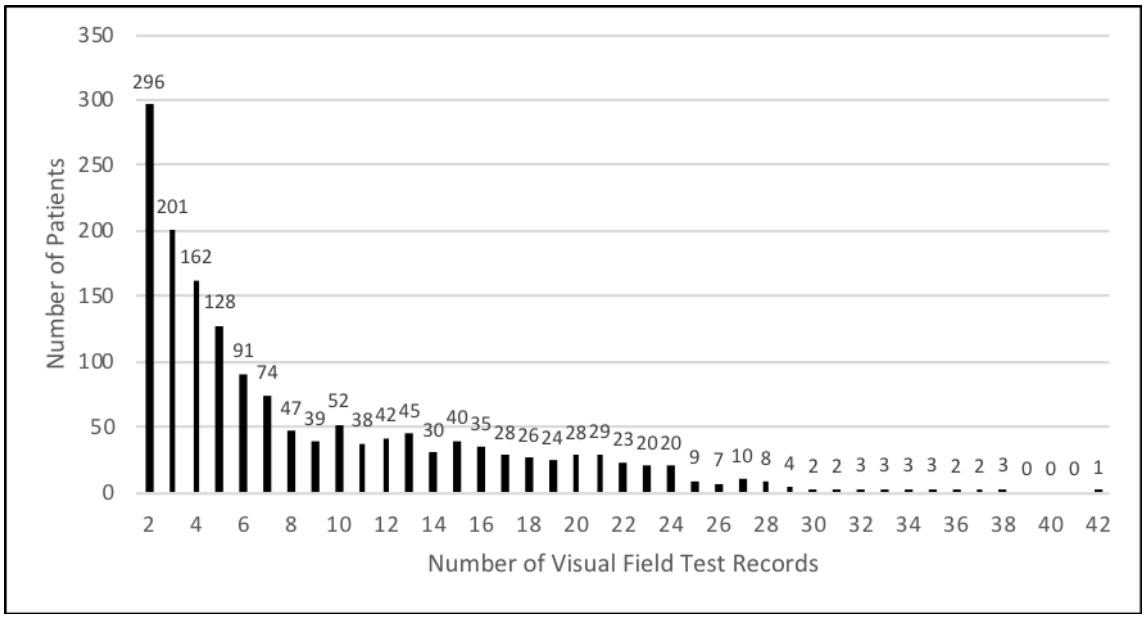


Figure 3-1: Distribution of Number of Visual Field Record

As tabulated in Table 3-1, nearly half of the data (41.4%) contains the AGIS score '0. Whereas in reclassified AGIS, which is used in this study, there is quite high proportionate of the data (69.1%) scored with 'none' and 'mild defect' of AGIS. This figure seems to suggest that the data suffer from the class imbalance problem, which is a problem particularly when there simply too few instances of a certain class, in the data (Chawla, Japkowicz and Kotcz, 2004). Imbalanced data is ubiquitous and is one of challenges faced by the big data community. The class imbalance problem is the problem in machine learning where the total number of a class of data (positive) is far less than the total number of another class of data (negative). In this research context, the number of instances of each class ('mild, 'moderate', and 'severe') are not equal. A number of solutions to this kind of problem can be proposed such as pre-processing techniques (re-sampling) aiming at rebalance data, and algorithmic techniques, which is allow an algorithm to learn from the imbalanced data (Fernández *et al.*, 2017). In this study, an initial experiment was conducted where the data were prepared in the 6NFB, and the performance of data classification was observed from the confusion matrix (H. He and

Garcia, 2009). A confusion matrix is a statistical tool used to measure the performance of classifiers. It is a table that tabulates the count of true positive, true negative, false positive and false negative (Fawcett, 2006). The confusion matrix also measures classifier performance by accuracy. Table 3-1 and Figure 3-2 exhibit the number of records by the AGIS score.

AGIS	AGIS Category	Number of Records	Percentage of Records	Reclassified AGIS Percentage
0	None	5692	41.43%	69.10%
1-5	Mild	3802	27.67%	
6-11	Moderate	2497	18.18%	18.18%
12-17	Severe	1382	10.06%	12.72%
18-20	End-Stage	366	2.66%	
Total		13,739	100%	100%

Table 3-1: Number of Visual Field Test Records by AGIS Category

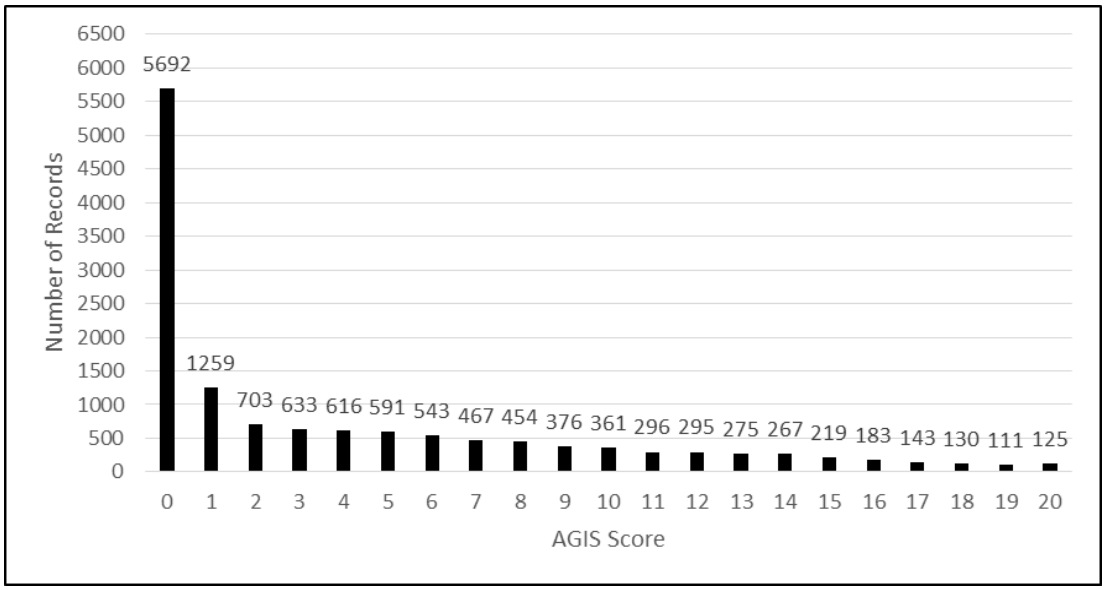


Figure 3-2: Number of Visual Field Test Records by AGIS Score

Table 3-2 to 3-6 tabulate the distribution of visual field records by the reclassified AGIS score. A random sampled dataset consists of 1,580 records where one visual field test from each patient is obtained randomly. Meanwhile, early, middle and latest datasets are sampled by getting one record from each patient with conditions. The early record dataset takes the first visual field test record and paired with the second record for the AGIS (considering *t+1*). Since the minimum number of visual field tests available in the raw data is two, all patients have early records. However, the middle record dataset accounts 1,083 records by excluding patients with two and three number of visual field tests (totalling 497). The assumption made in this study for middle record dataset is that patients with two and three records do not have middle record. They are only eligible for the early and latest datasets. In latest dataset, 1,284 records are used in this research by excluding those patients with two record of visual field test (totalling 296). From the tables (Table 3-3 to 3-6) the proportion of data with AGIS '0'is high (more than 63%) in all sampled datasets.

Class	No.	%
0	9494	69.1%
1	2497	18.2%
2	1748	12.7%
Total	13,739	100%

Table 3-2: Target Variable Class Proportion for Raw Data

Class	No.	%
0	1048	66.3%
1	291	18.4%
2	241	15.3%
Total	1,580	100%

Table 3-3:Target Variable Class Proportion for Random Sampled Dataset

Class	No.	%
0	1086	68.7%
1	274	17.3%
2	220	13.9%
Total	1,580	100%

Table 3-4: Target Variable Class Proportion for Early Records Dataset

Class	No.	%
0	744	68.7%
1	201	18.6%
2	138	12.7%
Total	1,083	100%

Table 3-5: Target Variable Class Proportion for Middle Records Dataset

Class	No.	%
0	818	63.7%
1	271	21.1%
2	195	15.2%
Total	1,284	100%

Table 3-6: Target Variable Class Proportion for Latest Records Dataset

In the data exploration process, the sampled datasets as discussed above were modelled using the Naïve Bayes classifier. Prior to modelling, the data with 52 visual field locations were prepared with the 6NFB. The sampled datasets consist of 1,580 records of visual field test (one record each patient). Whilst the 'all data' dataset consist of 12,159 records as the latest records from each patient are omitted due to pairing the visual field test with *t+1* of the AGIS (as discussed in section 2.3.1). The confusion matrix for classification of the datasets is retrieved to measure the classification performance. Tables 3-7 to 2-18 are the confusion matrix, which are the highlighted boxes indicate the true positive of the classes, for visual field data classification using the 6NFB. Although the data have disproportionate number of class instances, the performances of classification are good (with lower error rates and higher accuracy in each class) in all datasets. For 'all data' classification, the error rates of the classes are 6.5%, 5.6% and 2.5% for class '0', '1', and '2' respectively. Even though class '2' has least instances in the data, the true positive rate (10.2% - Table 3-7) is fairly good at a reasonable error rate (2.5%). Additionally, *G*-mean is computed to see the classifier performance. *G*-mean measures the balanced performance of a learning algorithm between the classes (Phung, Bouzerdoum and Nguyen, 2009; Sun *et al.*, 2007) that considers true negative rate and true positive rate. It is computed using true positive rate (*TP*) and true negative rate (*TN*) as follows:-

$$G-mean=\sqrt{TP_{rate}\times TN_{rate}} \qquad \textbf{Equation 3-1}$$

Where $TP_{rate}=\frac{TP}{TP+FN}$ and $TN_{rate}=\frac{TN}{TN+FP}$. With 85.5% accurate, the *G*-mean value for the classifier in all data classification are impressive with 0.89, 0.79, 0.88 (near to 1.0) for class '0', '1' and '2' respectively (Akosa, 2017). Therefore, this research work does not perform re-sampling dataset, which are over-sampling and under-sampling, before modelling the data (Yoon and Kwek, 2007).

Legend: Highlighted boxes indicate the true positive of the class.

Observed	**Predicted**					
	Classes	**0**		**1**		**2**	
	0	7649	62.9%	750	6.2%	31	0.3%
	1	387	3.2%	1507	12.4%	295	2.4%
	2	16	0.1%	286	2.4%	1238	10.2%
Accuracy	**85.5%**					

Table 3-7: Confusion Matrix for All Data

Observed	**Predicted**					
	Classes	**0**		**1**		**2**	
	0	944	59.7%	97	6.1%	4	0.3%
	1	63	4.0%	195	12.3%	42	2.7%
	2	3	0.2%	36	2.3%	196	12.4%
Accuracy	**84.5%**					

Table 3-8: Confusion Matrix for Random Sampled Dataset

Observed	Predicted					
	Classes	**0**		**1**		**2**	
	0	968	61.3%	113	7.2%	5	0.3%
	1	59	3.7%	171	10.8%	44	2.8%
	2	3	0.2%	39	2.5%	178	11.3%
Accuracy		**83.4%**					

Table 3-9: Confusion Matrix for Early Dataset

Observed	Predicted					
	Classes	**0**		**1**		**2**	
	0	681	62.9%	60	5.5%	3	0.3%
	1	44	4.1%	120	11.1%	37	3.4%
	2	2	0.2%	23	2.1%	113	10.4%
Accuracy		**84.4%**					

Table 3-10: Confusion Matrix for Middle Dataset

Observed	Predicted					
	Classes	**0**		**1**		**2**	
	0	730	56.9%	86	6.7%	2	0.2%
	1	46	3.6%	191	14.9%	34	2.6%
	2	3	0.2%	34	2.6%	158	12.3%
Accuracy		**84.0%**					

Table 3-11: Confusion Matrix for Latest Dataset

3.2.1 Synthetic Data

The use of synthetic data in experiments can be seen as a 'reverse engineering' technique to validate a new method. Yeung et al. (Yeung *et al.*, 2001) conducted a study on model-based clustering towards bioinformatics data and benchmarked the performance of model-based clustering on several synthetic data, which each class in the synthetic data was generated according to a multivariate normal distribution. Another study presented by He Yi and colleagues (Y. He, Pan and Lin, 2006) on model-based clustering used multivariate normal mixture models to create two synthetic datasets that mimic gene expression data. Much research has used synthetic data to evaluate algorithms such as (Bolón-Canedo, Sánchez-Maroño and Alonso-Betanzos, 2013; Metwally, Agrawal and El Abbadi, 2005; Ester *et al.*, 1996). The synthetic dataset used within this research is from a multivariate normal (MVN) data. The dataset length is 2500 (records) with 45 variables. The target variable (ordinal type) of the data is classed into the following three classes: '1', '2', and '3'. These classes indicate which cluster that has the highest average value over a set of variables. If the cluster 1 has the highest average value of the variables, the class of the target variable for the record is '1', so on and so forth. The proportion of the synthetic data by the target variable is balanced as shown in Figure 2-15. The details of the synthetic data are further discussed in Chapter 4 of this thesis.

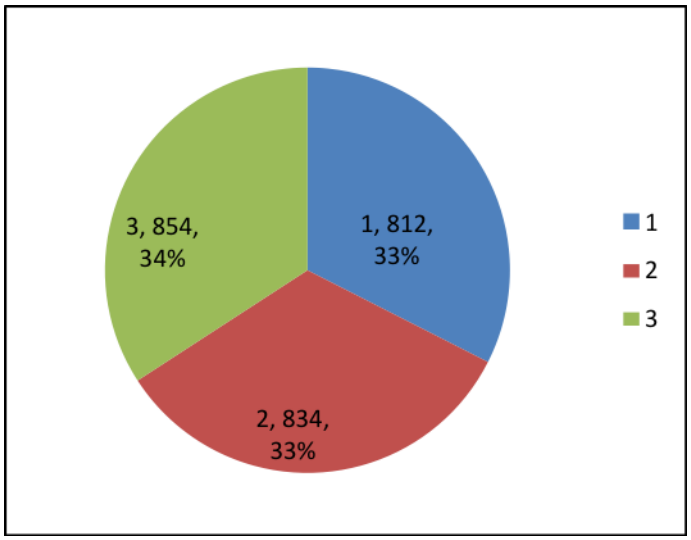


Figure 3-3: Number of Synthetic Record by Target Variable Classes

3.3 Summary

Data exploration (Jebb, Parrigon and Woo, 2017) is a precursor step of machine learning performed on the visual field data in this study. The data used in this study are good for empirical experiments to be carried out in the next chapters even though the data has imbalanced classes. In addition to the real data, a synthetic dataset is used in this research to validate the proposed approach of clustering and modelling as in many other studies (D. Zhang *et al.*, 2017; Enshaeifar *et al.*, 2017; Han and Abdelrahman, 2017; Brinkman *et al.*, 2006).

The main principle of the following chapter (Chapter 4) is to propose a model-based clustering technique namely, SMC. The practicability of the approach is examined by performing empirical experiments on the visual field data using the identified classification and optimisation methods discussed in this chapter. They are used to explore and validate the practicality of the approach on the real data.

Chapter 4

Simultaneous Modelling and Clustering

4.1 Overview

Chapter 4 presents the novel approach of modelling and clustering namely, Simultaneous Modelling and Clustering (SMC), which is based on the work presented at the Computer Based Medical System conference (M. Z. M. B. Jilani, A. Tucker and S. Swift, 2016). The goal of this chapter is to validate the proposed approach so that a better understanding of the relationship between visual field locations can be made, as well as the generation of models that can better predict glaucoma progression. The SMC method is tested on visual field data to cluster the 52 visual field locations. As discussed in the previous chapter, clustering and classification are common techniques used in big data analysis especially high-dimensional data such as visual field data. Applying these techniques enables the exploration of patterns and discovers some latent knowledge that available in data. In SMC, the spatial clusters over the visual field are determined by using heuristic search techniques which are scored based upon the prediction accuracy of glaucoma deterioration. The end result (resultant clusters) is compared to methods using the standard clusters that are based upon physiological traits (the six optic nerve fiber bundles – 6NFB). This chapter is structured into seven sections. Section 4.2 briefly presents introduction and background of Chapter 4, section 4.3 introduces the SMC method. The experimental method and results are presented in section 4.4 and 4.5 respectively. Section 4.6 discusses further detail about the results and section 4.7 concludes the chapter.

4.2 Introduction and Background

Visual field (VF) data is used to manage the condition of glaucoma. Analysis of visual field data has been widely undertaken in predicting glaucoma deterioration using a variety of techniques including data classification, clustering (Ceccon *et al.*, 2014) and statistical methods (S. R. Bryan *et al.*, 2013; Swift and Liu, 2002; Fitzke *et al.*, 1996). Additionally, machine learning techniques such as Bayesian Networks are commonly used in a number of studies (Ceccon *et al.*, 2014; Bowd and Goldbaum, 2008; Bizios, Heijl and Bengtsson, 2007). Discovering glaucoma deterioration patterns in visual field data helps with the early detection of glaucoma. Moreover, it helps physicians in providing appropriate treatments to patients who are suffering glaucoma, which is the second leading cause of blindness in the world (Quigley and Broman, 2006). The classification and clustering of visual field data can aid in the understanding of the multi-dimensional nature of the data which can be exploited in distinguishing healthy and diseased eyes. Applying a classifier to this time series data aids in discerning glaucoma deterioration progression in patients (Brusini and Johnson, 2007). Improving glaucoma deterioration prediction helps physicians to provide a more accurate early diagnosis of glaucoma thus better treatments for patients. Moreover, finding a cluster arrangement of visual field that have significant mapping to optic nerve head by means of evolutionary computation techniques is a major contribution to medical practitioners. The optic nerve head is a collection of fiber bundles (Bourne, 2006) which are connected to the brain. A number of visual field studies have used these 6NFB as the basis for their analysis. As presented in Chapter 2, (Garway-Heath *et al.*, 2000) proposed the mapping of visual field to the optic disc, which has motivated many clustering and classification studies based on these mappings (6NFB). (Ceccon *et al.*, 2012) have clustered the 52 visual field locations into the 6NFB. A recent study, (Sacchi *et al.*, 2014) which is grouped the 52 locations into six sectors and incorporated additional variables in the data, obtained up to 85% classification accuracy using the Naïve Bayesian classifier. Likewise, statistical methods have been used in predicting glaucoma deterioration, Swift *et al.* (Swift and Liu, 2002) in their study they used the

6NFB with short multivariate time series for prediction. This approach however fixes the number of visual field locations into six significant sectors to be used in modelling the visual field data. Even though the fixed significant sectors have been used as a standard approach in modelling, none of studies have had recorded a large prediction improvement in modelling. Thus, finding a new way in predicting glaucoma deterioration using the visual field data is an open opportunity for further research. Furthermore, improving prediction accuracy in modelling using a new method is a substantial contribution in this field. The hypothesis underpinning these experiments is that there exists a clustering arrangement, different (but possibly with some similarity) to the known 6NFB, that will result in a high classification/prediction accuracy (than the known bundles). This clustering arrangement can be determined by using an appropriate heuristic search technique. Therefore, this study proposes a novel approach of predicting glaucoma progression. In the novel approach, visual field data are simultaneously modelled and clustered to predict the stage of glaucoma. Unlike other classification and prediction studies, significant clusters of visual field are searched for based on the data and then these clusters are used to predict the level of glaucoma deterioration.

Optimisation methods are used as the basis for this search and a number of classification results are compared. The data are modelled based on the new clustering arrangement of visual field with the highest accuracy prediction.

4.3 Simultaneous Modelling and Clustering

As discussed in the previous chapter, clustering comprises a few techniques include model-based clustering. The SMC is a model-based clustering technique where clusters are identified based on objects shape and structure rather than on proximity between data points. Model-based clustering technique is extensively used both for continuous and discrete domain (Meilă and Heckerman, 2001). This research applies SMC in a discrete

problem where clustering arrangements of visual field locations are searched based on the classification accuracy (measure in percentage). As the name applies, the SMC comprises of these two main analysis techniques that is clustering and modelling. The diagram below shows the process of validating SMC using visual field data.

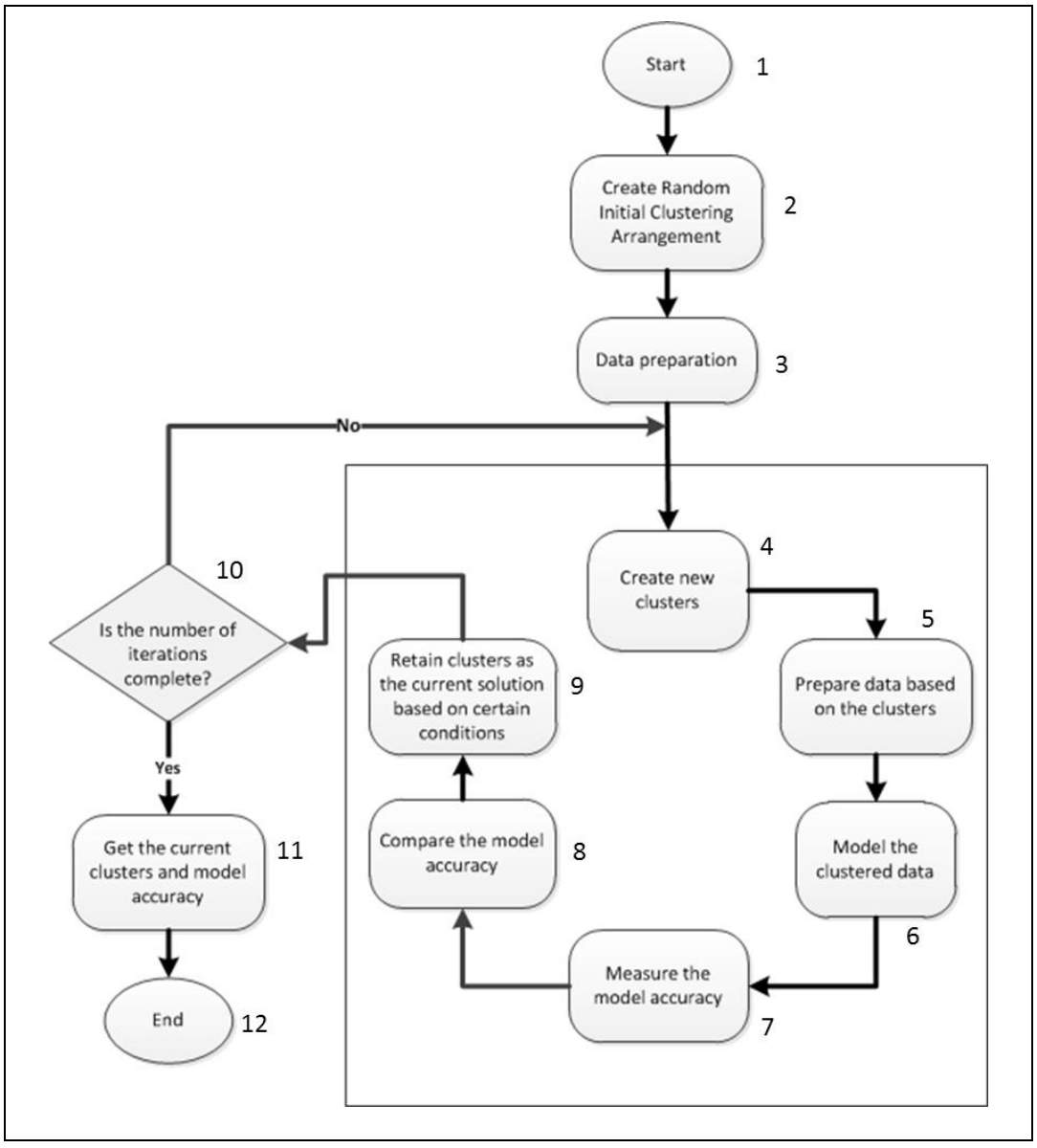


Figure 4-1: High Level Process of Simultaneous Modelling and Clustering

As depicted in the diagram, in process 2, the initial clusters of the visual field locations are created randomly. Based on the clustering arrangement of 52 visual field locations, the data are prepared (process 3). Data preparation is a process of obtaining the average values of VF locations in clusters as illustrated in Figure 4-2 (example of 15 variables). From this illustrated example, cluster 1 gets the average value of visual locations 1, 3, 5, and 9 (divided by 4).

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**AGIS (*T*+1)**
{1,3,5,9}	{2,6,14}	{4,7,10}	{8,15}	{11,12,13}	AGIS score

Figure 4-2 : Illustration of Data Preparation in the SMC

Following the data preparation stage, the looping process starting from process 4 to 9 is iteratively undertaken until the number of iterations is reached. From process 4 to 9, an optimisation method is applied to search for the best clustering arrangement of visual field locations. This heuristic process iterates for a large number of iterations where the clustering arrangement is improved in the search. Within the looping process, a new clustering arrangement is identified (process 4) by making a small change from the initial clustering arrangement (also referred to as current clustering arrangement). Continuing to process 5, data are prepared based on the new clustering arrangement. Next, the prepared data with the new clustering arrangement are modelled using classification in process 6. After classification, the model is measured by taking an accuracy value (predictive accuracy), which is measured in percentage (%). In this process, there are two solutions (clustering arrangements) with accuracy values: current accuracy and new accuracy for initial and new clustering arrangement respectively. Thereafter, the two solutions are compared in process 8. Whichever is the best solution based on certain conditions (such

as high accuracy, high acceptance probability), the clustering arrangement is retained as the current solution in the iteration (process 9). The hypothesis underpinning this approach is that the higher the accuracy, the better the quality of the associated clustering arrangement. The whole process of SMC is terminated in process 10 when the condition is met. Then, the latest clustering arrangement and corresponding accuracy value are captured for further analysis using Weighted Kappa statistics. The Weighted Kappa statistic between the clusters found by the method and the 6NFB is computed to measure the similarity of clustering arrangement between them. In other words, this work involves dimensionality reduction, which is the process of reducing the number of variables.

4.4 Method

The method of these empirical experiments entails data pre-processing, which is the preliminary task involved in machine learning, classifiers and optimisation methods, and experiment strategy.

4.4.1 Data Pre-Processing

Anonymised visual field data provided by Moorefield Eye Hospital London were used in these experiments. In the SMC process, data are prepared into a time series record before the experiments are executed. Each patient's test record is paired with the AGIS score of next visit test (t+1) as illustrated in Figure 4-3. As such, the most recent test record of every patient was excluded in the experiments. Therefore, the data used in the experiments after data cleaning consisted of 12,159 of visual field test records. The AGIS scores are classified into five categories that indicate the severity of the glaucoma condition, however in this study, the stages were reclassified into three stages for an efficient prediction following the same procedure detailed in Sacchi's work (Sacchi *et al.*,

2014). The results presented in Sacchi's work showed that prediction of visual field data (AGIS score) was as high as 85% accurate utilising additional variables which are not used in this study.

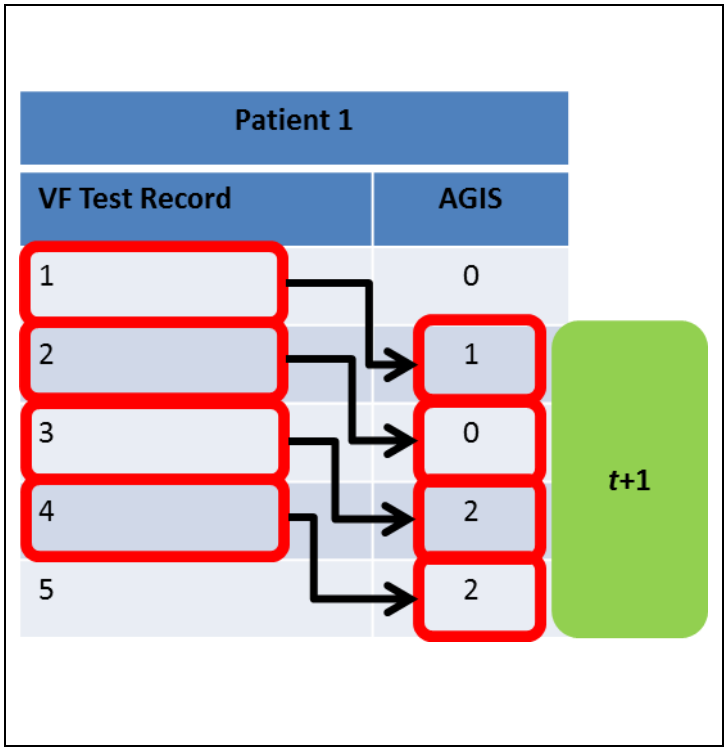


Figure 4-3 : Preparation of VF record with AGIS

4.4.2 Classifiers and Optimisation Methods

There are three classifiers and three optimisation methods used in this work. The classifiers are Decision Tree (J48), Naïve Bayes (NB) and Multinomial Naïve Bayes (MNB). Selection of the classifiers was made due to their efficiency and track record in data classification (N. Friedman, Geiger and Goldszmidt, 1997; J. H. Friedman, 1997). Meanwhile Multinomial Naïve Bayes is an efficient classifier (in terms of runtime), although the prediction performance may not be as good as NB (Tao and Wei-hua, 2010;

Rennie *et al.*, 2003). In SMC, the data are modelled using 10-fold cross validation (10FCV). Cross validation is the commonest way in data mining and data modelling (Fushiki, 2011; Kohavi, 1995; P. Zhang, 1993), and in this research context it is used to measure the predictive performance (predictive accuracy percentage value). Predictive accuracy of visual field data is obtained in the SCM process where it works as a fitness value to measure the clustering arrangements of visual field locations.

The optimisation methods used in the experiments are Random Mutation Hill Climbing (RMHC), Random Restart Hill Climbing (RRHC) and Simulated Annealing (SA). RMHC is the simplest and most straight forward optimisation method that searches for a solution in the data space based on improvement of a fitness value. Due to this simplicity, the method tends to get stuck in a locally optimal solution in the search space (Basseur and Goëffon, 2013; Tovey, 1985). For that reason, RRHC is also employed which is believed to be able to avoid local optimum solutions (Lim, Rodrigues and Zhang, 2006). Additionally, SA is also employed to avoid local optimum in the search due to the possible acceptance of worse fitness values under certain conditions (Rutenbar, 1989). Empirical experiments are carried out using these methods and the performance of the classifiers is captured.

Hill Climbing

Random Mutation Hill Climbing (RMHC) is a simple and straightforward optimisation method and is easy to implement and understand. As such it is often used as a benchmark technique when testing search methods (Mitchell, Holland and Forrest, 1994). However due to this simplicity, the method tends to get stuck in a locally optimal solution in the search space (Basseur and Goëffon, 2013; Tovey, 1985). The SMC algorithm of RMHC and RRHC are detailed in Algorithm 4-1 and 4-2.

Algorithm 4-1 : Simultaneous Modelling and Clustering of Random Mutation Hill Climbing Algorithm	
Input:	D = visual field data iterations = Number of iterations fd = 10-fold cross validation Model$_{\in}${J48, NB, MNB}
1	Let $C_{current}$ = random clusters of visual field points
2	Let $D_{current}$ = visual field data of the $C_{current}$
3	Let $fitness_{current}$ = prediction accuracy of the $C_{current}$ with fd
4	For i=0 to iterations-1
6	C_{new} = re-arrange elements in $C_{current}$ for small_change
7	D_{new} = D of the C_{new}
8	$Fitness_{new}$ = prediction accuracy of D_{new} with fd
9	if $fitness_{new}$ > $fitness_{current}$
10	$fitness_{current}$ = $fitness_{new}$
11	$C_{current}$ = C_{new}
12	end if
13	end for
Output:	$C_{current}$, prediction accuracy = $fitness_{current}$

Algorithm 4-2 : Random Restart Hill Climbing Algorithm	
Input:	D = visual field data iterations = Number of iterations fd $_{\in}${10-fold, 2-fold, no-fold cross validation} Model$_{\in}${J48, NB, MNB} Restart = 10
1	For i=0 to restart-1
2	Perform an RRHC algorithm as Algorithm 3-1
3	$C_{restart\_i}$ = Current clusters of restart of i
4	End for
Output:	C = The best cluster from restarts Best prediction accuracy of D of $C_{restart\_i}$

In the RMHC and RRHC search, two options small change (Definition 4.1) is used to obtain a new solution. Clusters are shuffled by moving two variables from its own cluster to another cluster. With this perturbation to the current solution, it would give a little change to the current fitness value.

Definition 4-1 Small Change

A small change in this work is defined as a number of moves that changes variable (visual field locations) position from one cluster to another cluster. It is a perturbation made to the current solution for a new solution.

Simulated Annealing

The SA method was used to advance the search from the HC methods. The heuristic search based on metallurgy process enables to avoid from local optima. From the literature, the SA method has two properties such as acceptance probability and artificial temperature that make the method becomes efficient in searching a solution. In the SA method the artificial temperature is cooled using a constant cooling rate (λ) and the final temperature in SA system is set to as small as possible ($T_n \approx 0.001$). With a guess value for the initial temperature and the known value of T_n, the rate of λ can be computed (Equation 4-3) as the following explanation:-

Computation of temperature for iteration 1 is denoted $T_1 = T_0\lambda$. Subsequently for iteration 2, the temperature is $T_2 = T_1\lambda$ with a note that $T_1 = T_0\lambda$ then the equation is further expanded to $T_2 = (T_0\lambda)\lambda$. Similar to the preceding temperature equation, with $T_3 = T_2\lambda$ the equation is further expanded to $T_3 = [(T_0\lambda)\lambda]\lambda$.

Therefore from these equations, in order to obtain the final temperature T_n (*n* is a number of iterations), the equation is defined as:-

$$T_n = T_0\lambda^n \qquad \textbf{Equation 4-1}$$

From Equation 4-1, the equation is expressed to get an equation for λ as the following steps:-

$$\ln(T_n) = \ln(T_0 \lambda^n)$$
$$\ln(T_n) = \ln(T_0) + \ln(\lambda^n)$$
$$\ln(T_n) = \ln(T_0) + n\ln(\lambda\ )$$
$$\ln(T_n) - \ln(T_0) = n\ln(\lambda\ )$$
$$n\ln(\lambda\ ) = \ln(T_n) - \ln(T_0)$$

$$\ln(\lambda\ ) = \frac{\ln(T_n) - \ln(T_0)}{n} \qquad \textbf{Equation 4-2}$$

With the assumption $x = \frac{\ln(T_n) - \ln(T_0)}{n}$, Equation 4-2 is simplified as:-

$$\lambda\ = e^x \qquad \textbf{Equation 4-3}$$

However the initial temperature (T_0) in this work is derived from another simulation as successfully applied by Swift *et al.* (Swift *et al.*, 2004) in clustering gene expression data. The simulation is run to get the accumulation of small change of fitness value and to be divided by the number of iteration. In the simulation five percent (5%) from the total iterations of the SA experiment is allotted to compute the initial temperature. In this work, 500 (5% of 10,000) was used in the simulation experiment to determine the initial temperature. This resulted the SA experiments were run for 9500 (95% of 10,000). Algorithm 4-3 delineates the simulation of getting an initial temperature for SA.

Algorithm 4-3 : Initial Temperature Simulation Algorithm	
Input:	D = visual field data iterations = Number of iterations temperature = 0 fd $\in$ {10-fold, 2-fold, no-fold cross validation} Model $\in$ {J48, NB, MNB}
1	For i=0 to iterations-1
2	Let $C_{current}$ = a random cluster of visual field points
3	$D_{current}$ = D of $C_{current}$
4	Let $fitness_{current}$ = prediction accuracy of $D_{current}$ classification with fd
5	C_{new} = re-arrange $C_{current}$ for small_change
6	D_{new} = D of the C_{new}
7	$Fitness_{new}$ = prediction accuracy of D_{new} classification with fd
8	Δfitness = |new fitness – current fitness|
9	temperature = temperature + different fitness
10	$fitness_{current}$ = $fitness_{new}$
11	$C_{current}$ = C_{new}
12	end for
15	temperature = temperature /iterations
Output:	Initial temperature

Figure 4-4 illustrates the example of temperature is cooled in a SA method experiment.

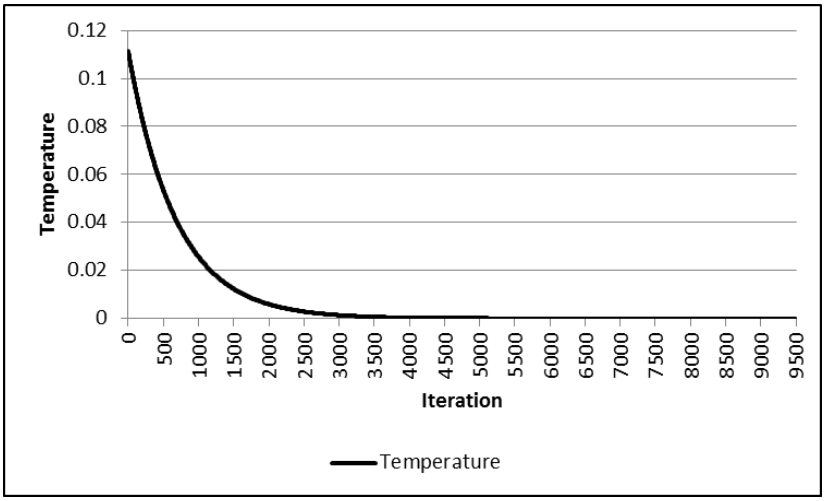


Figure 4-4 : Cooling of Temperature in SA Method Experiment

The SMC algorithm of SA is described in Algorithm 4-4.

Algorithm 4-4 : Simultaneous Modelling and Clustering of Simulated Annealing Algorithm	
Input:	D = visual field data iterations = Number of iterations initial temperature = Algorithm 3.3 temp = initial temperature fd = 10-fold cross validation Model$_{\in}${J48, NB, MNB}
1	Let $C_{current}$ = a random cluster of visual field points
2	Let $D_{current}$ = D of $C_{current}$
3	Let $fitness_{current}$ = prediction accuracy of $D_{current}$ with fd
4	For i=0 to iterations-1
6	C_{new} = re-arrange $C_{current}$ for small_change
7	D_{new} = D of C_{new}
8	$fitness_{new}$ = prediction accuracy of D_{new} with fd
9	if $fitness_{new}$ > $fitness_{current}$
10	$fitness_{current}$ = $fitness_{new}$
11	$C_{current}$ = C_{new}
13	else
14	$\Delta fitness$ = $fitness_{current}$ – $fitness_{new}$
15	Let Pr = Compute acceptance probability (Equation 2-9)
16	Let random = UR(0,1)
17	If Pr > random
18	$fitness_{current}$ = $fitness_{new}$
19	$C_{current}$ = C_{new}
20	end if
21	end if
22	temp = compute temperature (Equation 3-1)
23	end for
Output:	$C_{current}$, prediction accuracy = $fitness_{current}$

4.4.3 Experiment Strategy

In an initial data exploration, there exists a high negative correlation (-0.725) between number of tests and the corresponding AGIS score (**Appendix 4A**). The high negative correlation indicates that the more tests that patients have, the lower the AGIS. The

negative correlation between the number of tests and the AGIS score makes sense since the more tests undertaken by the patients, immediate clinical intervention can be undertaken and thus the deterioration could be avoided.

Some patients have very few records and some patients with severe glaucoma have many records. To avoid this inherent bias in the dataset, experiments were run using two strategies: one using all of the data and the second using sampled data. The "all data" experiments consists of all of the 12,159 records whilst sampled data strategy consists of 1,580 visual field test records. A random test record is sampled from each patient for experiments. However, the latest test record from each patient will not be selected due to pairing with the test record with t+1 of AGIS. Each experimental strategy was run using the three optimisation methods where each method employed the three classifiers. Experiments were run for 10 times with 10,000 iterations each. For RRHC method, 10 restarts were used in the experiment and each of the restart has 1,000 iterations (equivalent 10,000 iterations in all repeats). The choice of 10,000 iterations was decided upon due to an initial exploration of the data. Few experiments were run with 100,000 iterations using RMHC method with MNB classifier to observe the optimum solution and iteration. It was found that the algorithm has reached to the optimum solution at 3,785th. Therefore, choosing 10,000 as the iterations in formal experiments is adequate to reach the optimum solution. There are nine experiments are carried out for this work, which each of the optimisation method (RMHC, RRHC, and SA) is run for three individual experiments using the J48, NB and MNB classifiers. The nine experiment strategies are summarised as in Table 4-1.

Method	Classifier	Iteration	Experiment Repeat
RMHC	J48	10,000	10
RRHC	NB	1,000 (10 restarts)	10
SA	MNB	10,000	10

Table 4-1 : Summary of the Nine Experiment Strategies

4.5 Results

The results are presented as three individual experiments: initial experiments, all data experiments and sampled data experiments.

4.5.1 Initial Experiments

In the initial experiments, data were classified (with 10FCV) to predict the AGIS score at test (t+1) using the 6NFB. The full dataset was used for the classification without finding the best clusters and optimising the accuracy. As shown in Table 4-2, the results found that J48 is the best classifier with a prediction accuracy of 86.11%. Meanwhile NB and MNB have an accuracy of 85.17% and 76.29% respectively. The results above are consistent with (Sacchi *et al.*, 2014) for NB. Thus, these accuracies are used as the benchmark in this study. The experimental results in all data and sampled data were analysed and discussed on the prediction accuracy and Weighted Kappa statistic to determine the overall improvement over these benchmark accuracies.

Classifier	Accuracy (%)
J48	86.11
NB	85.17
MNB	76.29

Table 4-2 : Initial Experiments Prediction Accuracy using the 6NFB

4.5.2 All Data Experiments

Predictive Accuracy

Out of the 10 runs of the experiments, the best results of the models for each method were tabulated. Table 4-3 to 4-5 present the best results of prediction and the resulting clusters (with 10FCV). An average of the accuracy is over all of the 10 runs of the methods. Overall, the models improved the prediction from the initial experiments where the best accuracy recorded was 86.99% by J48 using SA (Table 4-5). The model also proposed the same size of clusters with the 6NFB. With NB, the results appear to have larger cluster sizes for the optimum solution (high accuracy) in all of the methods, these being 15, 17 and 10 clusters respectively. Meanwhile, the NB model results are consistent in all methods (86.16% - 5 clusters, 86.39 – 4 clusters, 86.45 - 7 clusters).

Model	Best Accuracy (%)	VF Clusters	Mean (%)
J48	86.84	5	85.74
NB	86.16	5	85.83
MNB	83.24	15	82.94

Table 4-3 : Model Prediction Accuracy of RMHC

Model	Best Accuracy (%)	VF Clusters	Mean (%)
J48	86.94	4	86.38
NB	86.39	4	85.95
MNB	83.27	17	83.15

Table 4-4 : Model Prediction Accuracy of RRHC

Model	Best Accuracy (%)	VF Clusters	Mean (%)
J48	86.99	6	86.59
NB	86.45	7	86.26
MNB	83.33	10	83.09

Table 4-5 : Model Prediction Accuracy of SA

Figure 4-5 and 4-6 exhibit the convergence graph of the experiments for the RMHC and the SA method respectively. The RMHC search has a smooth convergence graph as the method does not accept worse solutions in the search. However, the convergence graph for SA (Figure 4-6) clearly has fluctuations during the early iterations of the search. The explanation for this is that the method is accepting worse solutions in the search by computing the acceptance probability as discussed in Chapter 2. It can be seen that the fluctuations become less extreme as the iterations increase, due to the temperature cooling. This behaviour is exactly as expected and allows SA to avoid becoming stuck in local optima.

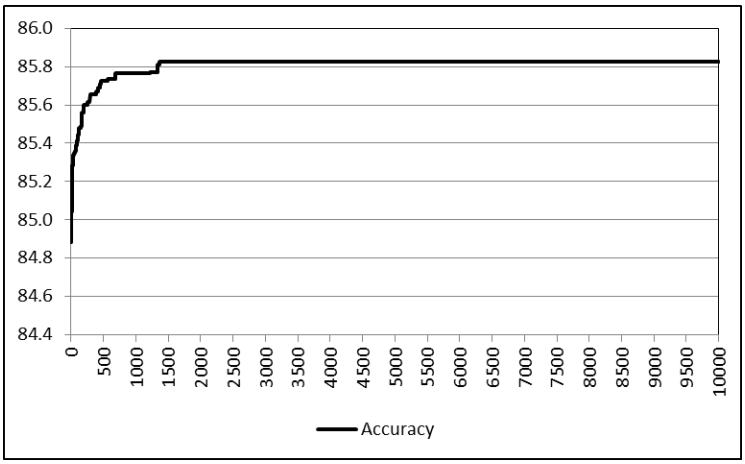


Figure 4-5 : Convergence Graph for RMHC with NB Classifier

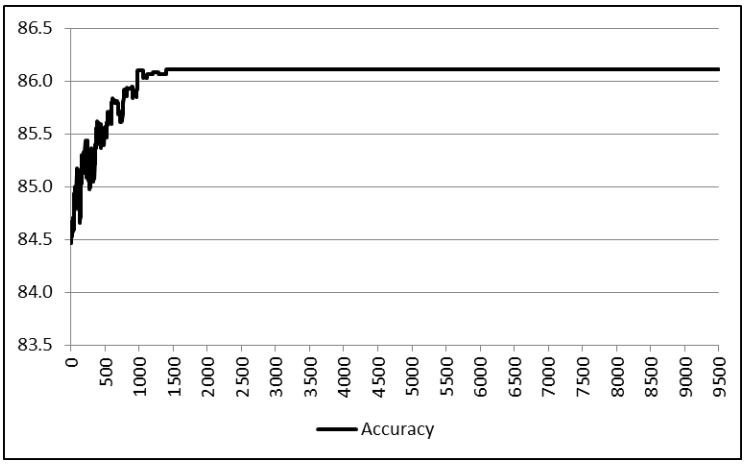


Figure 4-6 : Convergence Graph for SA with NB Classifier

Weighted Kappa Statistics

The clusters of the best experiments results based on the highest accuracy were used in the Weighted Kappa calculation. Weighted Kappa is calculated to see the agreement between the resulting clusters and the 6NFB. The Weighted Kappa results for the best accuracies of the HC, RRHC and SA methods are presented in Table 4-6.

Method/Model	**J48 (#cluster)**	**NB (#cluster)**	**MNB (#cluster)**
RMHC	0.018 (5)	0.080 (5)	0.013 (15)
RRHC	-0.001 (4)	0.009 (4)	0.001 (17)
SA	-0.007 (6)	0.002 (7)	-0.054 (10)

Table 4-6 : Weighted Kappa of Resulting Clusters

The results found that the resulting clusters using the SMC method with a random initial solution in the search has weak agreement with the 6NFB. Weighted Kappa in the experiment recorded between -0.054 to 0.080. This shows that the resulting clusters have very poor agreement with the 6NFB. Both the J48 and NB classifiers proposed smaller clusters (less than 8 clusters) in all methods compared to MNB (10, 15 and 17 clusters). Figure 4-7 shows the number of clusters in the search from the SA method experiment using the NB classifier. The number of clusters is decreasing during the search, where the clusters corresponding to high predictive accuracy in modelling.

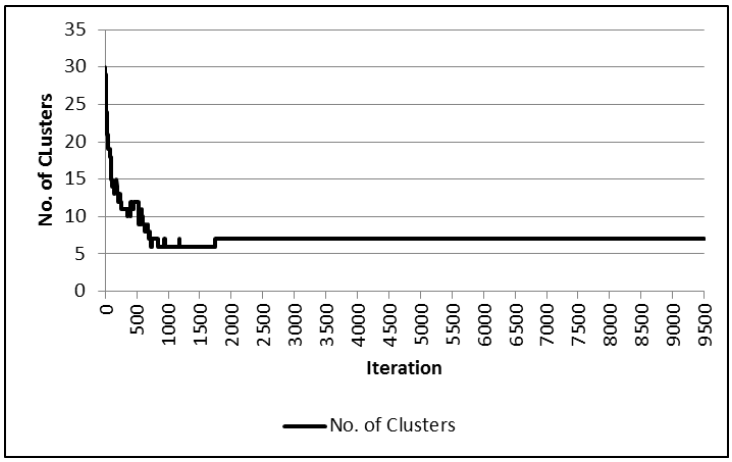


Figure 4-7 : Number of Cluster in the SA Method (NB)

4.5.3 Sampled Data Experiments

As discussed in the previous section, sampled data experiments use a single random visual field record from every patient. Therefore, the sampled dataset consists of 1,580 records. Each run of the experiments used a different sampled dataset. However, the RRHC method, which has 10 restarts in each run experiment, uses the same datasets within each restart.

Predictive Accuracy

Predictive accuracy results for sampled dataset are presented in Table 4-7 to 4-9. The sampled data experiments have slightly improvement compared to the all data experiments. In these experiments, the NB classifier was the best model at predicting the AGIS score (t+1). The model accuracies were 88.49% in SA, 88.35%, in RMHC and 87.38% in RRHC. The J48 classifier also has a significant accuracy improvement where the results were more than 87%. However, this is different to the performance of the MNB classifier. The classifier seems to perform better in the all data experiments when compared with the sampled data. As for the resulting clusters, it was found that the resulting clusters in the sampled data experiments are larger than the clusters in the all data experiments.

Model	Best Accuracy (%)	VF Cluster	Mean (%)
J48	87.27	17	85.99
NB	88.35	16	86.28
MNB	82.67	19	80.85

Table 4-7 : Model Prediction Accuracy of RMHC

Model	Best Accuracy (%)	VF Cluster	Mean (%)
J48	87.18	5	86.36
NB	87.38	8	85.96
MNB	82.45	12	81.68

Table 4-8 : Model Prediction Accuracy of RRHC

Model	Best Accuracy (%)	VF Cluster	Mean (%)
J48	87.15	7	86.35
NB	88.49	12	86.73
MNB	82.48	12	81.06

Table 4-9 : Model Prediction Accuracy of SA

Figure 4-8 and 4-9 exhibit the convergence graph for the HC and SA methods (NB classifier) respectively. From the sampled data experiments, it can be clearly seen that the convergence graphs are slightly earlier than the all data experiments.

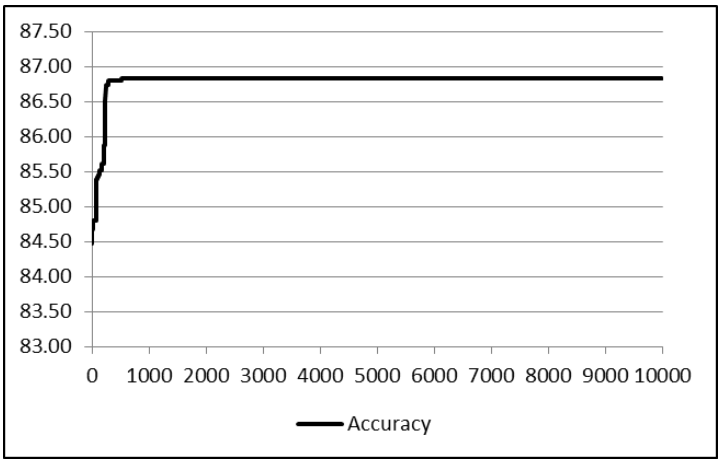


Figure 4-8 : Convergence Graph for RMHC with NB Classifier

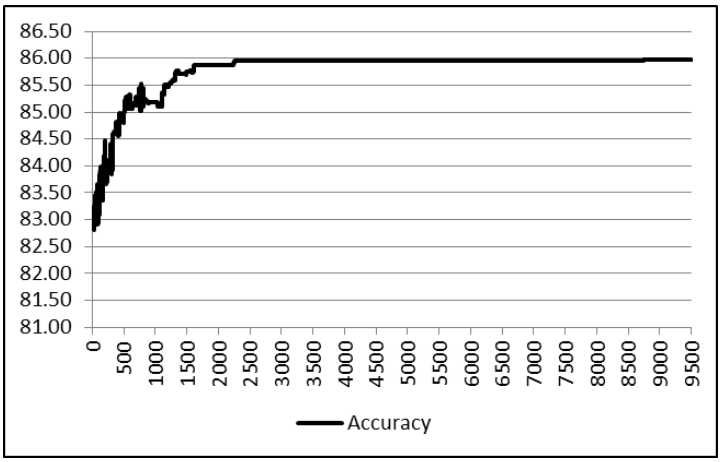


Figure 4-9 : Convergence Graph for SA with NB Classifier

Weighted Kappa Statistics

The Weighted Kappa results for the RMHC, RRHC and SA methods are shown in Table 4-10.

Method/Model	J48 (#clusters)	NB (#clusters)	MNB (#clusters)
RMHC	0.025 (17)	0.080 (5)	-0.029 (19)
RRHC	0.013 (5)	0.006 (8)	-0.019 (12)
SA	0.013 (7)	0.022 (12)	-0.021 (12)

Table 4-10 : Weighted Kappa of Resulting Clusters

There is no significant difference between the all data and sampled data as far as the Weighted Kappa score is concerned. Weighted Kappa in the sampled data experiments are very poor with the range of value being between -0.029 to 0.025. In these experiments, it appears that the larger the resulting clusters size the poorer

the Weighted Kappa value. Figure 4-10 shows number of clusters in the SA search method using the NB classifier.

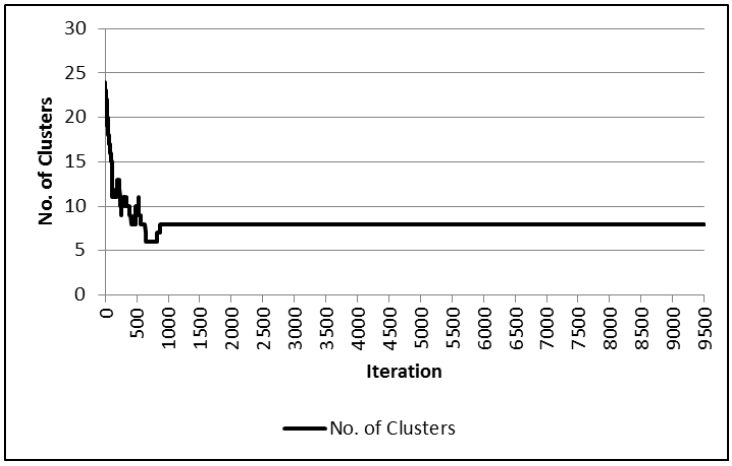


Figure 4-10: Number of Cluster in the SA Method (NB)

4.6 Discussion

Based on observation of the experimental results, SMC has proven an effective way in improving predictive accuracy for visual field data. The best accuracy recorded in the experiments was 88.49% by the SA method, with the NB classifier in sampled dataset. The accuracy was improved by 3.89% from the initial experiment result (using the 6NFB). The sampled data strategy was found to be more effective than using all of the data due to afore mentioned bias with the data. The strategy is also a more efficient strategy being significantly computationally quicker. On average, predictive accuracy of the sampled data experiments is better than the 'all data' experiments. Although the sampled data experiments have higher accuracy than the 'all data' experiments, the resulting clusters are variables and have low Weighted Kappa. This is likely to be due the nature of the sample data. Overall, based on the predictive accuracy result, SMC on the

sampled dataset experiments (best with 89.49%) outperformed the all data experiments (best with 86.99%).

The MNB classifier largely improves (9.23%) the prediction in all data experiment compared to other classifier. However, the best accuracy recorded by MNB was far less accurate (83.33%) compared to other models. The best model was NB (88.49%) with 3.89% prediction improvement in sampled dataset using the SA method. However, for all data experiment, J48 recorded the best accuracy (86.99%) in the SA method.

The resulting clusters appear to have larger size of cluster in sampled data experiments. The HC method proposed more than 15 clusters with all of the classifiers in the sampled data. Even though SMC has appeared to have an improved predictive accuracy, the resulting clusters have shown a disagreement in clustering arrangement with the 6NFB (best WK recorded 0.080) in both experiments (all data and sampled data).

Analysis in Chapter 4 is extended to see the mapping of the resulting clusters to the 54 locations of visual field. To visualise the best clusters result, "The Normal Approximation for the Binomial Approximation of the Hypergeometric distribution" (NBH) metric was used to locate the best resulting clusters of the experiments. This metric, proposed by (Swift *et al.*, 2004) (used in gene-expression data analysis), identifies the significance of the overlap between an individual cluster and a known set or function grouping of genes. This metric can be used to determine the overlap between any two individual collections of objects, or in this work case visual field points and the benchmark clustering arrangement (6NFB). The NHB analysis found that the resulting clusters from sampled data, RRHC, model J48, run 6 was the highest. Thus, the best clusters (with 19 clusters) restart 8 of the experiments was chosen to visualise the clusters on the 54 visual field locations. The NBH ranking analysis shows that RRHC, model J48, run number 6, had the most agreement (overlap) with the 6NFB, which is shown in Figure 4-11. From the

visualisation of the clusters, noted that the method found many small clusters size rather than larger clusters size. The larger clusters appear on the periphery of vision, which agrees with medical evidence that glaucoma first start in the periphery near the blind spot, corresponding to cluster numbers 15, 14 and 4. Cluster number 10 appears to be the only central set of points of a significant size.

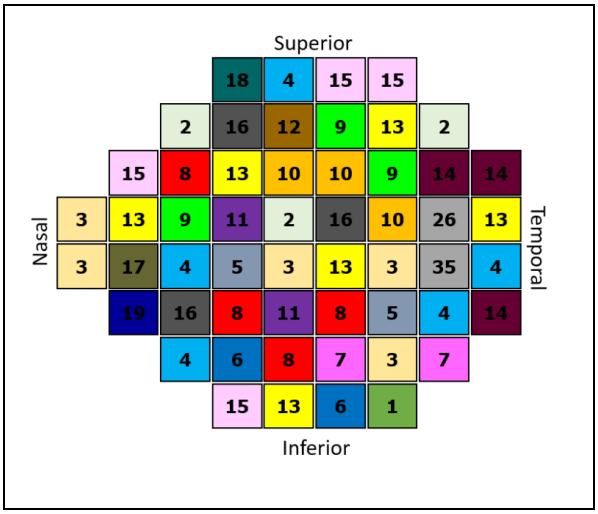


Figure 4-11 : Visual Field Map for Resulting Clusters RRHC, J48

4.7 Summary

The empirical experiments carried out in this work explore the new clustering and classification technique (SMC) to find the optimal visual field clusters for predicting the AGIS score for patient's suffering from glaucoma. The motivation of exploring this new method of visual field classification is to improve glaucomatous deterioration prediction and indirectly this could provide clinicians with accurate diagnosis of glaucoma towards the end of providing better treatment to patients. Additionally, model-based clustering is a significant contribution in analysing medical data that can widely be applied to other medical data and other domains.

The experiments results support the hypothesis that there exists a clustering arrangement of visual field, which is different (or can be slightly similar) to the known 6NFB, as well as improving the prediction accuracy of AGIS. The work in Chapter 4 concludes that SMC can be applied to visual field data to predict glaucomatous progression as all methods and models improve prediction accuracy.

From the positive results in this work, it opens more research opportunity to explore the technique especially involving other optimisation methods and synthetic data. Ideally, developing an algorithm that searches clusters efficiently and effectively in terms of convergence is a research opportunity to be addressed. Furthermore, since the result on the real data cannot guarantees the effectiveness of SMC (obtaining the right clusters), devising the technique on synthetic data could validate the effectiveness of the technique to search the optimum solution. Additionally, it is also believed that the SA method is not providing as efficient a search as would be expected. Therefore, extending the number of iterations could possibly improve prediction accuracy.

The gaps found in Chapter 4 will be bridged in the next chapters. Chapter 5 will be focusing on applying SMC on synthetic data to validate SMC.

Chapter 5

Simultaneous Modelling and Clustering on Synthetic Data

5.1 Overview

The work in the previous chapter successfully demonstrated that SMC improves prediction accuracy applied to the real world data. However, the resultant clusters are still uncertain since the Weighted Kappa results against the six nerve fiber bundles (6NFB) were very poor. The poor Weighted Kappa between the resultants clusters and the 6NFB in the previous work does not mean that SMC is not a viable clustering technique even though accuracy is improved. Therefore, Chapter 5 presents the application of SMC on synthetic data. Synthetic data is generated to have the same nature and properties with the real data (visual field). Devising synthetic data is a *'reversed engineering'* approach to validate SMC. In addition to this, Chapter 5 extends the work by conducting an analysis on both datasets using another clustering technique: *K*-means. The objective of this extension analysis is to see performance of *K*-means clustering technique as a benchmark for this study. This chapter consists of six sections: Overview, Introduction, Synthetic Data, Experiments and Results, Discussion, and Summary.

5.2 Introduction

It has been widely accepted that efficient modelling in classification requires a large amount of training dataset. A recent study (Özgür and Erdem, 2017) on the impact of using a large training dataset on classification accuracies has shown performance benefits even though it is high computation cost to algorithms. However, this is impossible to obtain due to expensive and many other reasons such as class imbalance (Nonnemaker and Baird, 2009). Additionally, the use of synthetic data is an acceptable approach to

validate and simulate algorithms. Tsai and a colleague (Tsai and Chatterjee, 2017) in their recent study have devised synthetic data to evaluate the performance of an algorithm. Similarly in this work, the novel model-based clustering technique SMC is tested on a synthetic dataset. This was motivated by the experiments results from the real data conducted in Chapter 4 are still uncertain to justify the effectiveness of the technique in terms of WK. The WK metric for the resultant clusters against the 6NFB in the previous chapter were found very poor. However, the poor results of WK cannot conclude that SMC is not viable. Moreover, the 6NFB is commonly used in clinical practices and is not a gold standard for representing the collection of visual locations on the optic disc. Therefore this chapter is essential to validate SMC, which is expected to find the known clustering arrangement from a multi-dimension synthetic dataset. Synthetic dataset is generated to have similar data properties with the real data, visual field. The resultants clusters are measured using the WK metric with the known clusters arrangement of the dataset.

5.3 Synthetic Data

A synthetic dataset was constructed as a verification tool for the SMC technique. This dataset was constructed to have similar properties to that of the visual field data, namely, number of variables, length, and number of classes. The synthetic dataset used in this work is based on a number of multivariate normal (MVN). The multivariate normal distribution is a generalisation of the one-dimensional (univariate) normal distribution to higher dimensions. The distribution of MVN is demonstrated in **Appendix 5-A** (Casabianca and Junker, 2016). The synthetic dataset was generated from three multivariate normal distributions (15 variables each) with 2,500 records. Each distribution has a different mean vector and different positive semi definite covariance matrix. The samples are then concatenated to form a single 45 variable dataset of 2,500 samples in length. Each row in the dataset is given a class label defined by which 15 variables (from the original datasets) have the highest average. Figure 5-1 exhibits

labelling the target variable class for the synthetic dataset. For example in record 1, the target variable is labelled as '2' as the highest average of 15 variables is from cluster two or variable 16 to 30. In record 2, the highest average is from variable 1 to 15 (cluster 1), thus the class label is '1'. The expectation of using this synthetic dataset is that the SMC approach will be able to garner the original structure of the underlying data generation process, i.e. three clusters of 15 variables each.

Record	Cluster 1	Cluster 2	Cluster 3	Target Variable
	Variable 1- Variable 15	Variable 16- Variable 30	Variable 31 – Variable 45	
1	Average value is 55.5	Average value is 100.5	Average value is 5.0	2
2	Average value is 27.6	Average value is -15.2	Average value is 12.8	1
3	Average value is 17.9	Average value is 6.1	Average value is 85.5	3
2499	Average value is 5.3	Average value is 35.9	Average value is 17.0	2
2500	Average value is 33.3	Average value is -21.4	Average value is 192.8	3

Figure 5-1 : Visualisation of Synthetic Dataset and its Target Variable

It can be summarised that the synthetic dataset has three clusters for the 45 variables as follows:-

$$N = 15 \qquad S = 1500$$

$$X_1 \in \Re^{S \times N}$$

$$X_2 \in \Re^{S \times N}$$

$$X_3 \in \Re^{S \times N}$$

$$D = [X_1, X_2, X_3]$$ **Equation 5-1**

Where *N* is number of variables, *S* is sample size and *D* is the synthetic dataset. Algorithm 5-1 shows the generation of the synthetic dataset.

Algorithm 5-1 : Generation of Synthetic Data	
Input:	N = a vector of sizes M = the length of each sample
1	Repeat
2	D = NULL (an empty matrix)
3	For i = 1 to |N|
4	n = N_i
5	S = RPSD(n)
6	T = MVN(0,S,m)
7	D = [D T]
8	End For
9	Add a blank column to the end of D
10	For each row of D (excluding the last column), compute the average of the columns grouped according to N, set the last column value of each row to whichever group (index of N) which has the highest average
11	Until the class variable (last column) is approximately balanced (equal counts)
Output:	D, a synthetic dataset

Where RPSD(*n*) is a function that generates a random positive semi-definite (covariance) matrix of size *n* by *n*.

MVN(μ,σ,m) generates an *m* length (rows) multivariate normal sample with mean μ, and covariance σ. The number of variables (columns) is implicit from the dimensionality of μ and σ.

The operator [..] concatenates matrices and/or vectors by column.

5.4 Experiments and Results

Experiments carried out in this chapter include the SMC experiments on the synthetic dataset. The assumption in the SMC experiments is that higher WK in the synthetic dataset experiments indicates the effectiveness of SMC (getting the right clusters in the synthetic dataset with WK = 1.0) since the clustering arrangements of the data are known. However, getting higher classification accuracy remains the main objective of SMC in this experiment, which is the underlying assumption that higher accuracy represents the best solution (clustering arrangement). The optimisation methods used in SMC are SA and RRHC owing to the high accuracy results in the work carried out in Chapter 4.

In addition, supplementary experiments were performed in this work using *K*-means clustering on both datasets. The choice of *K*-means as the benchmark to SMC is that the technique is widely used in research (Khanmohammadi, Adibeig and Shanehbandy, 2017; Capó, Pérez and Lozano, 2017) for clustering problems where data are clustered based on the proximity of data points. Also, comparison between a clustering technique that uses data point distance and model-based clustering technique can substantiate the efficiency and effectiveness of SMC. The *K*-means clustering experiments were conducted to see the performance of the common technique of clustering as a benchmark for the SMC. The *K*-means algorithm is outlined in **Appendix 5-B**.

5.4.1 The SMC Experiments

Preliminary experiments on the synthetic data were conducted using the RRHC and SA algorithms as these two optimisation methods have good track records with high predictive accuracy in the previous experiments. Also, the NB classifier (best classifier with high accuracy in Chapter 4) is used in this work to model the datasets with 10-fold

cross validation. As shown in the previous chapter, to run a larger iteration could probably give a better result (high accuracy). As such, the experiments were run 50,000 iterations. These experiments were run 10 times and statistical values such as average, minimum and maximum are presented. The level of agreement (WK) between the resultant clusters and the known clusters are also presented.

Interesting results with high classification accuracy were obtained from the experiments on the synthetic data as shown in Table 5-1. The SMC technique appears to have good improvements in classifying the synthetic data with 96.85% accuracy (best result) in the SA method. There is not much difference in the RRHC method (94.86%). The average accuracies of 10 runs of the experiments are 92.51% and 95.77% for RRHC and SA respectively.

Method	**Average Accuracy (%)**	**Maximum Accuracy (%)**	**Minimum Accuracy (%)**
RRHC	92.51	94.86	88.64
SA	95.77	96.85	94.81

Table 5-1 : Prediction Accuracy for Synthetic Data

As for the agreement level (WK) between the resultants clusters and the known clusters of the synthetic data, the average results are fair and good strength for RRHC (0.34) and SA (0.77) respectively (Table 5-2). Surprisingly, the SA method found the expected clusters of the synthetic data with WK 1.00 (best result recorded) in one of the experiments.

Method	Average WK	Maximum WK	Minimum WK
RRHC	0.34	0.46	0.26
SA	0.77	1.00	0.63

Table 5-2 : Weighted Kappa Statistics for Synthetic Data

5.4.2 *K*-Means Clustering

Experiments using *K*-means clustering technique were conducted on both datasets and run for 25 times. Since the *K*-means clustering technique requires for practitioners to specify a number of clusters (*K)* (Kodinariya and Makwana, 2013), in these experiments therefore the *K* is set to six and three for VF and the synthetic data respectively. The reason being is that the 6NFB for the VF data and the known three clusters for the synthetic data. The 25 resultants clusters from *K*-means clustering were then used for modelling and classification accuracies are obtained. In these experiments, three modelling strategies were used to classify the datasets. The three modelling strategies are 10-fold, 2-fold with 10 repeats and no-fold cross validation (P. Zhang, 1993). These modelling strategies are to experiment the effect of the number of partitions (*N*-fold) in cross validation towards the SMC's performance which is measured in classification accuracy as well as WK. A smaller *K*-fold in cross validation would result in a larger test dataset.

The visual field data were modelled using all data and sampled dataset whilst the synthetic data were experimented using all data. Modelling the VF data in the all dataset has slightly better results (high accuracy) as compared to the sampled dataset (Table 5-3 and Table 5-4). Both datasets results have high prediction in the no-cross validation modelling strategy (84.85% and 83.89% for all data and sampled data respectively) as shown in Table 5-3

and 5-4. On average, the resultant clusters of the visual field data experiments have fair agreement WK (0.362) with the 6NFB (Table 5-5).

Modelling Strategy/Result	**10FCV Accuracy - %**	**2FCV Accuracy - %**	**NoFCV Accuracy - %**
Minimum	82.65	83.01	83.07
Maximum	85.80	85.77	85.75
Average	84.77	84.80	84.85

Table 5-3 : Prediction Accuracy of Visual Field Data using *K*-Means Resultants Bundles (All Data)

Modelling Strategy/Result	**10FCV Accuracy - %**	**2FCV Accuracy - %**	**NoFCV Accuracy - %**
Minimum	81.76	82.63	83.89
Maximum	86.38	85.45	85.00
Average	83.82	83.70	83.89

Table 5-4 : Prediction Accuracy of Visual Field Data using *K*-Means Resultants Clusters (Sampled Data)

Average	0.362
Max	0.432
Min	0.265

Table 5-5 : *K*-Means Resultants Clusters Weighted Kappa Statistic for Visual Field Data (With the 6NFB)

On the contrary, the resultant clusters of the synthetic dataset experiments have very poor agreement (WK) with the known clusters (average of 0.055 - Table 5-6). The synthetic dataset experiments are found poor accuracy using the *K*-means resultant clusters with an average 61.41% accurate (Table 5-7) in the no-cross validation modelling strategy.

Average	0.055
Max	0.185
Min	-0.022

Table 5-6: *K*-Means Resultants Clusters Weighted Kappa Statistic for the Synthetic Data (With the known 3 clusters)

Modelling Strategy/Result	**10 FCV Accuracy - %**	**2 FCV Accuracy - %**	**No FCV Accuracy - %**
Minimum	53.90	52.90	53.88
Maximum	65.58	64.64	65.80
Average	60.91	60.72	61.41

Table 5-7: Prediction Accuracy of the Synthetic Data using *K*-Means Resultants Clusters

5.5 Discussion

SMC on the Synthetic Dataset

On average, over the 10 run experiments, the performance of SMC on the synthetic dataset is more than 90% accurate in both methods using the NB classifier. The best result

of the RRHC method was from experiment 1, in repeat 3 (94.86% accurate) with 0.46 WK. Figure 5-2 shows the convergence graph for the RRHC method in the synthetic dataset (iteration 0 to 5000). As RRHC search does not accept worse solutions in the search, thus the graph has no fluctuation.

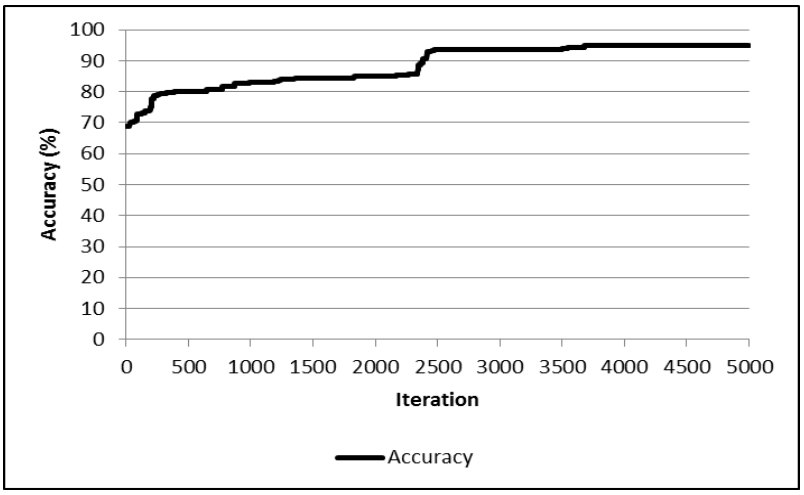


Figure 5-2 : Convergence Graph for RRHC Method on Synthetic Dataset

Meanwhile SA has shown interesting results with high classification accuracy and WK. Figure 5-3 shows the convergence graph for the best result in SA, which obtained from experiment 10. The SA search from the graph looks extremely fluctuate in the beginning of the search due to the nature accepting worse solution. Then, it becomes stable toward the end of the search as the temperature and the acceptance probability are decreasing.

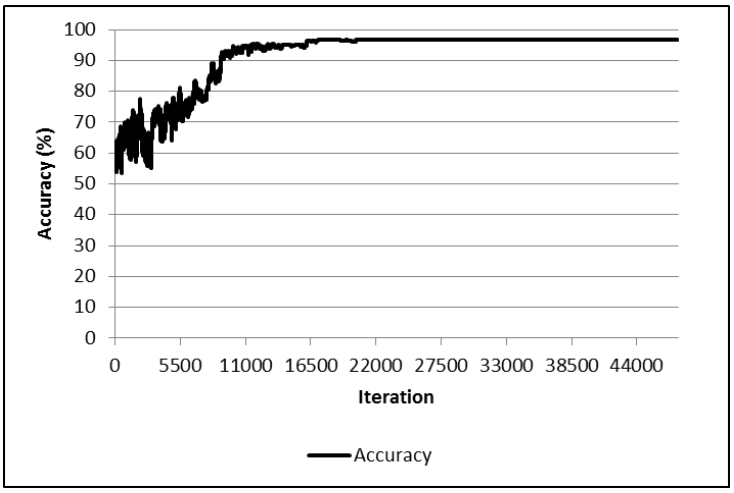


Figure 5-3 : Convergence Graph for SA Method on Synthetic Dataset

From the results, it was found that the highest accuracy in SA does not correspond to the highest WK. The WK value for the highest accuracy (96.85%) in the SA experiment is 0.88, with four clustering arrangement. The resultant clusters for experiment 10 of SA were retrieved and the arrangement of clusters is as follows:-

Cluster 1: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}

Cluster 2: {16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30}

Cluster 3: {31, 32, 33, 38, 41, 42, 43, 44}

Cluster 4: {34, 35, 36, 37, 39, 40}

Based on this observation, cluster one and two are the prefect clustering arrangement as per the known clusters. However, cluster three and four are the third expected clustering arrangement split in half (approximately). It can be concluded that high accuracy does not guarantee to produce the best clustering arrangement. This might be due to noisy

fitness convergence (Definition 5.1) and a random dataset in cross validation. Unlike a non-probabilistic optimisation method such as HC (presented in Chapter 4 and Figure 5-2) this is not affected by noisy fitness convergence.

Definition 5-1 : Noisy Fitness

A noisy fitness function is defined as one where when given two solutions that have the same genotype (Kojima, 1971), the fitness evaluation is not the same. More formally there exists solutions: *S1* and *S2* where *S1* = *S2* and $F(S1) \neq F(S2)$. This phenomenon in optimisation problem relates to the notion of fitness landscape which originated from theoretical biology. It requires analysis on the problem and data to draw an analogy with the real landscape to gain a better understanding of how and where an algorithm operate for a problem (M. Wang *et al.*, 2017).

The experiment with the best WK value was retrieved and plotted. The best WK (1.00) recorded by the SA experiment was from experiment 6 with accuracy 95.59%. Figure 5-4 shows the convergence graph of WK for the SA method (from experiment 6). From Figure 5-4, it indicates that the search has found the known cluster in the synthetic data (with 1.00 WK) slightly early in the search before it converged (as circled in Figure 5-4). Discovery of clusters with 1.00 WK in the synthetic data in the search before its convergent indicates the inherence of noisy fitness. This means that high fitness values (classification accuracy) in the search do not guarantee the best solution (clustering arrangement).

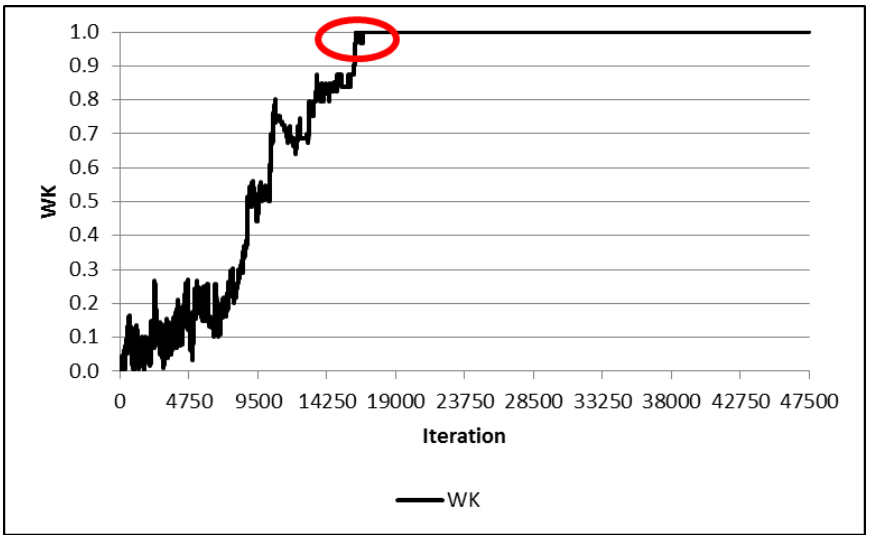


Figure 5-4 : WK Convergence Graph for SA Method on Synthetic Dataset

K-means Experiment

K-means clustering was performed as a benchmark for SMC. The results were found that the *K*-mean clustering technique did not improve the accuracy in both VF data and synthetic data. In this technique the best accuracy for the VF data was from the 'all data' experiments with the no-fold cross validation modelling strategy (84.85%). Meanwhile the best accuracy recorded by SMC in the previous chapter was 88.49% which is improved by 4.29%. However, surprising results were found that WKs in *K*-means are much higher (the best 0.432) than SMC in the VF data experiments. On average, the *K*-means clustering technique has fair WK (0.362) with the 6NFB.

The *K*-means experiments results on the synthetic data were found very poor in accuracy with less than 66% as compared to SMC (best accuracy: 96.85%). The *K*-means results were found not to improve the classification accuracy when applied to the synthetic data. The best average accuracy was 61.41% from the no-fold cross validation modelling strategy. SMC largely improves the accuracy in the synthetic dataset by 55.95% (best

average 95.77% from the SA method). Corresponding to the lower accuracy in the *K*-means experiments, the WK results of *K*-means in the synthetic data experiments are somewhat counterintuitive. On average, the resultant clusters' agreement level with the known cluster is very poor (WK = 0.055). SMC appears to have much higher WK values than *K*-means as the method is advantaged from a heuristics search finding the right clusters based on accuracy.

5.6 Summary

The results from the synthetic dataset experiments have ascertained that SMC effectively clusters high-dimension data as well as improving classification accuracy in both datasets (visual field and synthetic). The SA method in SMC has shown promising results searching clusters (with maximum WK 1.00) in the synthetic data. Henceforth, the search method will be used in the next chapter's experiments. Meanwhile *K*-means clustering, which is set as a benchmark for SMC, appears to be a viable clustering technique for visual field data having a higher WK (average 0.362 - against the 6NFB) than SMC. However, using the *K*-means' clustering arrangement as a basis for predicting deterioration of glaucoma/modelling the synthetic dataset results in poor predictive/classification accuracy.

Even though SMC has shown promising results toward the synthetic data, there is a need to explore the technique further on the real data. There is still lack of evidence in terms of efficiency in SMC (this related to convergence). In the next chapter, a set of experiments are presented which looks at search method convergence as a measure of algorithmic efficiency. Noted in this work that both datasets suffer from noisy fitness which is high accuracy of modelling does not guarantee to produce the best clustering arrangement and there potentially to get a different fitness value with the same solution.

Therefore, noisy fitness tolerance analysis will be introduced in the next chapter as a part of convergence point analysis.

Given the computational overheads of the search methods, there is a need to look into more advanced and faster heuristic search techniques such as Generalised Simulated Annealing (Tsallis and Stariolo, 1996) and other faster classifier such as Multinomial Naïve Updateable. As the increased number of iterations in the experiments in this chapter was really significant to improve accuracy, a larger iteration of search (Chang *et al.*, 2016) in Chapter 6 is to be considered.

Chapter 6

Generalised Simulated Annealing in Simultaneous Modelling and Clustering

6.1 Overview

The SMC technique was proven effective in searching the known clusters in the synthetic data with high WK as demonstrated in Chapter 5. Building on the findings from the synthetic data experiments in the previous chapter, there is a need to further explore the effectiveness of SMC on the real data by employing an advanced optimisation method. Chapter 6 presents a novel application of an algorithm namely, Generalised Simulated Annealing (GSA) tailored to solve the discrete optimisation problem (SMC). The expectation of applying this method within SMC is to find clusters arrangement of visual field in more efficient way than SA and RRHC. The focus of this chapter is testing and observing the GSA method which is believed to have better performance in terms of accuracy, quality of cluster arrangements (high WK) and efficacy than other heuristic search techniques as reported in literature. As such, convergence point analysis is performed in this work to capture the efficiency of the method. Chapter 6 is divided into six sections. Section 6.2 gives a short brief on GSA algorithm. Section 6.3 outlines the experiments setup where the experiments are carried out on both datasets. Section 6.4 presents the results of the experiments. Section 6.5 discusses the results. Finally, section 6.6 provides the summary of the chapter.

6.2 Generalised Simulated Annealing

Broadly speaking, there is “no free lunch” in optimisation problems where one could apply a generic method to solve many problems. However, much in the literature (Menin

and Bauch, 2017) has shown that the GSA search method is effective and efficient when applied to continuous problems. Generalised Simulated Annealing (GSA) is an improved version of the Simulated Annealing (SA) algorithm, proposed by Tsallis (Tsallis and Stariolo, 1996). The algorithm generalises both types of SA, i.e. Classical Simulated Annealing (CSA) and Fast Simulated Annealing (FSA). This family of stochastic algorithms was inspired by the metallurgy process for making a molten metal reach its crystalline state by employing an artificial temperature (Xiang *et al.*, 1997). Unlike Hill Climbing (HC) (Selman and Gomes, 2006; Tovey, 1985), SA and GSA methods are able to avoid local optimum in the search due to the inherent statistical nature of the method (Bohachevsky, Johnson and Myron L. Stein, 1986). Worse solutions found in the search are accepted when certain probabilistic criteria are met, thus enabling the methods to escape local optima.

CSA is likely to find a global optimum solution in the search. However, the convergence is fairly slow (Xiang *et al.*, 1997). This is attributed to the nature of the visiting distribution which uses a Gaussian distribution (local search distribution). Thus in 1987, Szu and Hartley (Szu and Hartley, 1987) proposed FSA which uses a Cauchy Lorentz visiting distribution (semi-local search distribution). The FSA method is quicker at finding the optimum solution compared to CSA since the jumps are frequently local, but can occasionally be quite long. The cooling of the temperature in this method is much higher than in CSA which can make the search more efficient.

Later in 1988, a generalisation of the Boltzmann-Gibb statistics was introduced (Tsallis, 1988). GSA was invented for generalising both CSA and FSA methods according to the Tsallis statistics. GSA uses a distorted Cauchy Lorentz visiting distribution where the distribution is controlled by the visiting index parameter (q_v). This method (GSA) was believed to be more efficient in terms of convergence and global optimum in nonconvex problems (multiple extrema) compared to the precedence methods of annealing. Because

of these advantages, there are many studies have applied the method in many fields. Application of GSA in the field of biology, chemistry, physic and mathematics (dos R Correia *et al.*, 2005; Andricioaei and Straub, 1996; Sutter, Dixon and Jurs, 1995; D. G. Brooks and Verdini, 1988) are commonly involved in the determination of the global optimum of multidimensional continuous functions (Xiang *et al.*, 2013). The GSA approach was proven faster than the other simulated annealing algorithms (CSA and FSA) in the study of mapping minima points of molecular conformational energy surfaces (Moret *et al.*, 1998). Xiang and Gong (Xiang and Gong, 2000) have shown that the GSA algorithms are relatively efficient in Thomson's model and nickel clusters compared to CSA and FSA. It was also claimed that the more complex the system, the more efficient the GSA method. A recent study (Mojica and Bassrei, 2015) on the simulation of 2D gravity inversion of basement relief with synthetic data has found that the GSA method produces better results with the calibrated parameters.

The positive findings from the previous studies on GSA have inspired this study to investigate the method in SMC of glaucomatous progression using visual field (VF) data (M. Z. M. B. Jilani *et al.*, 2016). Note that most of the studies have applied GSA towards solving continuous problems; hence within this chapter presents a GSA algorithm for discrete optimisation. The hypothesis underpinning this work is that the GSA method finds the optimum solution more efficient compared to the SA. This is determined by observing the prediction accuracy (with high accuracy) and the convergence (early convergence) of the search.

In GSA, two parameters are introduced: acceptance index (q_a) and visiting index (q_v). These parameters regulate the methods behaviour such as acceptance probability, search convergence and cooling rate. As discussed in Chapter 2, the GSA equations for acceptance probability, acceptance temperature and visiting distribution are as follows:-

$$p_{qA}(x_t \to x_{t+1}) = \frac{1}{[1+(q_A-1)(E(x_{t+1})-E(x_t))/T_{q_A}^{A}]^{1/(q_A-1)}} \quad \textbf{c.f. Equation 2-11}$$

$$T_{q_v}^{V}(t) = T_{q_v}(1)\frac{2^{q_v-1}-1}{(1+t)^{q_v-1}-1} \quad \textbf{c.f. Equation 2-12}$$

$$a = \frac{[T_{q_v}^{V}(t)]^{-D/(3-q_v)}}{\left\{1+(q-1)\dfrac{(\Delta x_t)^2}{[T_{q_v}^{V}(t)]^{2/(3-q_v)}}\right\}^{1/(q_v-1)+(D-1)/2}}$$

$$g_{qV}(\Delta x_t) = \left(\frac{q_V-1}{\pi}\right)^{D/2}\frac{\Gamma\left(\dfrac{1}{q_V-1}+\dfrac{D-1}{2}\right)}{\Gamma\left(\dfrac{1}{q_V-1}-\dfrac{1}{2}\right)}\times a \quad \textbf{c.f. Equation 2-13}$$

Algorithm 6-1 shows the application of the GSA search method in SMC where visiting distribution and the artificial temperature are computed in line 5 and 22 respectively.

Algorithm 6-1 : Simultaneous Modelling and Clustering of Generalised Simulated Annealing	
Input:	D = visual field data iterations = Number of iterations fd $\in$ {10-fold, 2-fold, no-fold cross validation} temp = initial temperature qa = acceptance index qv = visiting index Model= MultinomialNaïveBayesUpdatable
1	Let $C_{current}$ = a random clusters of visual field points
2	Let $D_{current}$ = visual field data of the $C_{current}$
3	Let $fitness_{current}$ = prediction accuracy of the $C_{current}$ with fd
4	For i=0 to iterations-1
5	Calculate newvisit (c.f Equation 2-13)
6	Calculate small_change = newvisit × number of variable
7	C_{new} = re-arrange $C_{current}$ for small_change
8	D_{new} = D of the C_{new}
9	$fitness_{new}$ = prediction accuracy of D_{new} with fd
10	if $fitness_{new} > fitness_{current}$
11	$fitness_{current} = fitness_{new}$
12	$C_{current} = C_{new}$
13	Else
14	$\Delta fitness = fitness_{current} - fitness_{new}$
15	Calculate Pr (c.f Equation 2-11)
16	Let random = UR(0,1)
17	if Pr > random
18	$fitness_{current} = fitness_{new}$
19	$C_{current} = C_{new}$
20	end if
21	end if
22	Calculate temp (c.f Equation 2-12)
23	end for
Output:	$C_{current}$, prediction accuracy = $fitness_{current}$

Unlike the hill climbing and SA methods, the GSA algorithm may have more than two moves (perturbation to a current solution), which is derived from visiting distribution as the small change. Helsgaun (Helsgaun, 2009) demonstrated in his study on TSP (10,000

to 10,000,000 cities) problem using *k*-opt has shown both effective and scalable to solve the problem even though it was evidence time consuming if *k* is larger than 4.

The small change for GSA which is used in this study has a range between 2 and 52 as shown in the example below (Figure 6-1):-

The *new_visit* value (from c.f Equation 2-14) for visual field data in iteration 550 is 0.625.

Visual field data consists of 52 variables.

The small change which also known as a number of moves is computed as:-

$small\_change = new\_visit \times$ number of variables

$= 0.625 \times 52$

$= 32.5 \approx 33$

Therefore, in iteration 550, the current clusters will have perturbation with random 33 moves to produce a new solution.

Figure 6-1 : Example of Small Change Computation in GSA Algorithm

The GSA method is distinct from SA in determining the temperature, the acceptance probability value, and the selection of a next neighbouring point in the search space. The

acceptance probability (c.f Equation 2-12) and artificial temperature (c.f Equation 2-13) use an acceptance index (q_a) and visiting index (q_v) parameters (Tsallis and Stariolo, 1996). Furthermore, it is used in the visiting distribution equation (c.f Equation 2-14) (Tsallis and Stariolo, 1996) to determine the size of change (in this study this is a number of visual field locations to be re-arranged between clusters) for the next potential solution in the GSA search. Figure 6-2 shows the graph of number of moves in a GSA algorithm. The algorithm applies a large number of moves at early iterations for extensive search of solutions and it decreases towards the end of the search. Having a decreasing number of moves in the search allows a good current solution to have a tiny perturbation in order to improve the solution. This process in GSA supports to search a quality solution (clusters) in an efficient way where a good solution obtained at later iterations is unlikely to have a large magnitude of change.

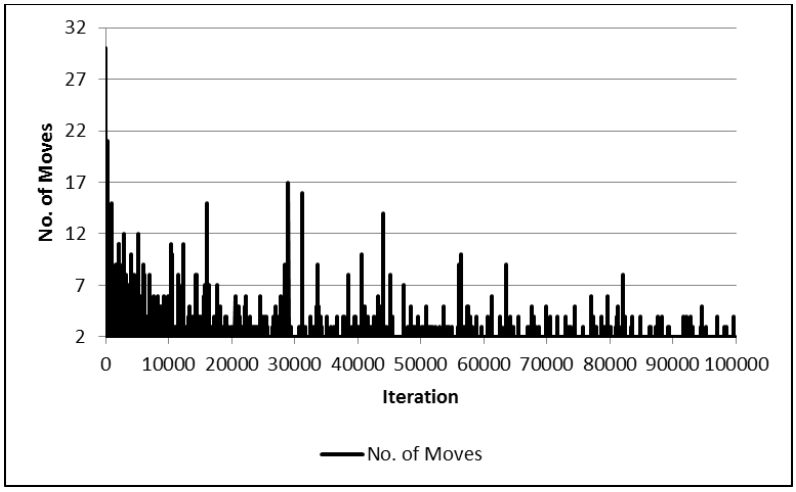


Figure 6-2 : Example of Number of Moves in the GSA Algorithm

Similar to the SA and HC method, each new better (in terms of fitness function) solution found in the GSA search is accepted. However, if the new solution is worse, a criterion of accepting the new worse solution is derived by computing the acceptance probability according to Equation 6-1. Then the acceptance probability (p) value is compared with

a random number (r) which obtained from a uniform distribution (0,1). The new worse solution is accepted if $p > r$. This process continues until the specified number of iterations is complete.

6.3 Experiment Setup

The experimental setup defines the properties of the experiments and the GSA algorithm used in this work such datasets, modelling strategies and parameter values. Two datasets are used in this work: visual field data and synthetic data. In order to avoid bias in visual field data experiment, the data are re-sampled every iteration resulting in 1,580 records. Additionally, the results obtained in Chapter 4 found that the sampled experiments of visual field data produced higher prediction accuracy in visual field data. Meanwhile all records of the synthetic data (2,500 records) are used in the experiments. As larger iteration experiments have shown positive results from the previous chapter, the GSA experiments are run for 100,000 iterations with the three modelling strategies: 10-fold (10FCV), 2-fold with 10 repeats (2FCV), and no-fold cross validation (NoFCV). Whilst the Naïve Bayes Multinomial Updateable (NBU) classifier is used to model the data owing to the most efficient method in terms of runtime (M. Z. M. B. Jilani *et al.*, 2016; Sundar, 2013; Tao and Wei-hua, 2010; Rennie *et al.*, 2003).

Since the GSA algorithm is more complex than the SA algorithm, setting the parameters values in the GSA algorithm needs careful investigation. Otherwise the method may not work well. Additional simulations are needed in order to ensure the values for the parameters are suitable for the problem by running preliminary small scale experiments on the data. Once these values are assured appropriate by the simulation experiments, formal experiments are run on both datasets.

Additionally, a separate set of experiments are conducted using *K*-means clustering method (both datasets) and the 6NFB clustering arrangement (visual field) as the benchmark for the methods used in this chapter. The *K*-means and 6NFB experiments use the same modelling strategies, classifier and number of iterations as the formal experiments.

6.3.1 GSA Parameter Exploration

Many studies have devised GSA algorithms and calibrate the parameters in GSA to suit the nature of problems. Determination of the best parameters value (q_a and q_v) is the crucial part in developing the GSA algorithm. A study applied GSA to protein folding (Agostini *et al.*, 2006) has explored the best parameters values range and discovered the best values ranges are between 1.10 to 2.60 and 1.50 to 2.60 for q_a and q_v respectively. The study also introduced a new parameter (q_t) for the cooling function to better control the temperature decreasing in the GSA system.

Obtaining the right values for q_a and q_v is a challenge. Inappropriate values for the parameters may get the GSA method becomes inefficient or may not work. Correct parameters values are depending on the nature of problems. Therefore, a mathematical approach is introduced within this work to obtain the right parameters values. This work uses the Newton Raphson mathematical technique to compute the suitable parameters values (q_v and T_0). The equations, which have three parameters values to be determined, are simplified to a single parameter. With the knowledge of the q_a value and a few assumptions, the fitness value being between 0 to 1, equations are manipulated using the

Newton Raphson method to get q_v. As such, Equation 5-1 and 5-2 are simplified to derive T_0, and thus q_v is obtained.

6.3.2 Newton Raphson

The Newton Raphson technique (Akram and ul Ann, 2015; Kelley, 2003) is a powerful mathematical technique to solve numeric equations. It involves finding a value for the root of a function. Finding a root in an equation is an iterative process by guessing the initial value of x from the function ($f(x)$) and the derivative of the function (tangent line - $f(x)$) is used to obtain the intercept of the tangent line. The *x*-intercept will be the enhanced approximation to the functions root. This iterates until $x: f(x) = 0$. The Newton Raphson method is as follows:-

Let $f(x)$ be a continuous function.

$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$ geometrically $(x_1, 0)$ is the intersection with the x-axis of the tangent to the $f(x)$ of $(x_0, f(x_0))$. This process iteratively repeats as $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ until a sufficient accurate value $(f(x) \approx 0)$ is reached. Further illustration with graphs is detailed in **Appendix 6-A**.

This method is used to solve the GSA equations in order to get the appropriate value of parameter q_v and T_0. This study manipulates the acceptance probability equation (c.f Equation 2-12) and temperature equation (c.f Equation 2-12) to derive the parameters values. From Equation 2-12, the acceptance probability is simplified as:-

$$P_0 = \frac{1}{[1+(q_a-1)(\Delta f)/T_o]^{1/q_a-1}}$$ **Equation 6-1**

In order to get the initial temperature (T_0) from this equation, some assumptions have been made. The acceptance probability is assumed to have value 0.4 at iteration 1 in the search.

This is possible when the worst random solution which is obtained at iteration 1 would have different of fitness 0.5 (50% deviates from the initial fitness value). This means that when P_0 is set to 0.4, there is a 40% chance of accepting a worse solution when the prediction accuracy is 0.5 (50%) worse than the current accuracy. Thus the Equation for the initial temperature is derived as follows (Equation 6-2):-

Assumptions $q = q_a - 1$ $\Delta f = 0.5$ $P_0 = 0.4$ $P_0 = \frac{1}{(1+q(\Delta f / T_0))^{1/q}}$	$(1+q(\Delta f / T_0))^{1/q} = \frac{1}{P_0}$ $(1+q(\Delta f / T_0)) = \left(\frac{1}{P_0}\right)^q$ $q(\Delta f / T_0) = \left(\frac{1}{P_0}\right)^q - 1$	$\frac{q\Delta f}{T_0} = \left(\frac{1}{P_0}\right)^q - 1$ $0.5 = \frac{1}{2}$ $\frac{1}{0.4} = 2.5$ $\frac{q}{2T_0} = 2.5^q - 1$
$T_0 = \frac{q}{2\left(\left(\frac{5}{2}\right)^q - 1\right)}$		**Equation 6-2**

Then the temperature equation is simplified as follows:-

$T(N) = T_0\left[\frac{2^{qv-1}-1}{(1+N)^{qv-1}-1}\right]$ where $s = q_v - 1$, N is number of iteration $T_N\left((1+N)^s - 1\right) = T_0\left(2^s - 1\right)$	$T_N(1+N)^s - T_N = T_0 2^s - T_0$ $T_N(1+N)^s - T_0 2^s = T_N - T_0$
$\frac{T_N(1+N)^s}{T_N - T_0} - \frac{T_0 2^s}{T_N - T_0} = 1$	**Equation 6-3**

Within Equation 6-3, the T_0 value is obtained from Equation 6-2 by selecting q_a (values ranged between -0.5 to 1.6). With T_0 and N (number of iterations, 100,000) in hand, Equation 6-3 is used to derive the correspondent value of q_v where T_N is the final temperature set as 0.001. Equation 6-3 is solved by means the Newton Raphson method.

It is known that $ab^x + cd^x = 1$ is non-linear. Therefore, Equation 6-3 is written as $a(1+N)^s + c2^s = 1$, where $a = \frac{T_n}{T_N - T_0}$ and $c = \frac{-T_0}{T_N - T_0}$. Thus, the non-linear function of Equation 6-3 is established as:-

$$f(x) = a(1+N)^s + c2^s - 1 \qquad \textbf{Equation 6-4}$$

$$f(x) = ab^x + cd^x - 1 \qquad \textbf{Equation 6-5}$$

And the derivative of function $f(x)$ is

$$f'(x) = ab^x \ln(b) + cd^x \ln(d)$$ **Equation 6-6**

Thus with $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$ iteratively repeats until $f(x) \approx 0$ the value of x is obtained to derived the value of q_v.

Figure 6-3 shows the convergence graph of the GSA temperature. The GSA temperature has drastic convergence as compared to the SA (Chapter 4, Figure 4-4).

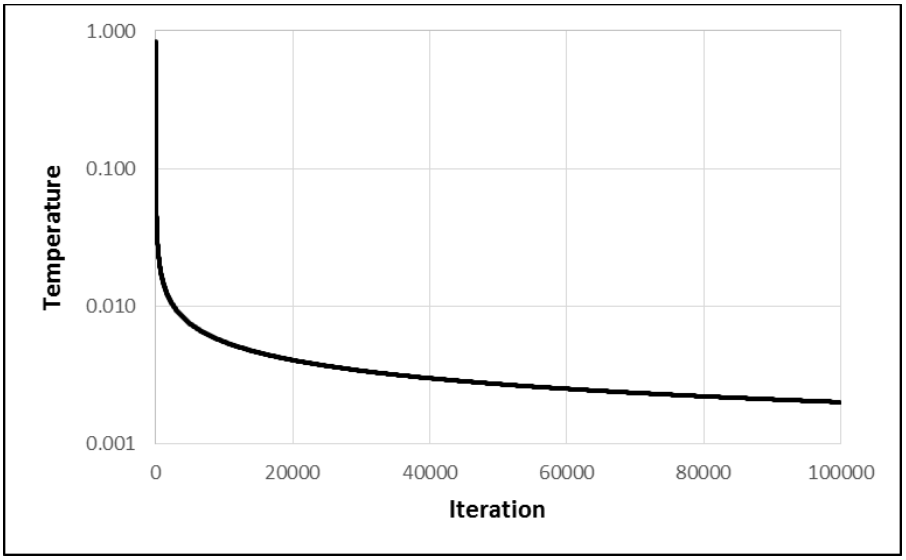


Figure 6-3 : Convergence Graph of the GSA Temperature

6.3.3 Simulation Experiments

From the literature, the value range for the parameter q_a is known. Therefore, preliminary experiments were conducted to find the suitable value for q_a and corresponding q_v. The preliminary experiments were run for three levels of the parameter range: level 1 between 0.001 to 0.009 (incremental by 0.001), level 2 between 0.01 to 0.09 (incremental by 0.01) and level 3 between 0.1 to 0.5 (Tsallis and Stariolo, 1996) (incremental by 0.1). The

experiments were run for 10,000 iterations (10 experiments) on both visual field and synthetic data with 10-fold cross validation. Table 6-1 and Table 6-2 show the corresponding fitness value (predictive accuracy) and convergence point of the search to the q_a for visual field and synthetic data respectively. The tables show the list of q_a values and its corresponding fitness values and convergence points, which is ranked by convergence point. Full results are tabulated in **Appendix 6-B**

q_a	Fitness	Convergence Point
0.010	87.25	3934.0
0.300	87.27	4609.4
0.050	87.52	4716.9
0.200	87.29	4731.8
0.040	87.48	4747.6
0.003	87.55	5064.8
0.080	87.46	5208.1
0.004	87.39	5221.4
0.100	87.43	5233.4
0.090	87.46	5411.0
0.009	87.38	5463.2
0.006	87.37	5551.9
0.020	87.32	5621.9
0.400	87.38	5744.1
0.030	87.44	5808.2
0.005	87.37	6243.0
0.008	87.58	6300.2
0.001	87.32	6507.3
0.002	87.25	6729.4
0.060	87.36	6881.6
0.070	87.31	7136.7
0.007	87.38	7230.4
0.500	83.45	10000.9

Table 6-1 : Visual Field Data Parameter Fitness Value

q_a	Fitness	Convergence Point
0.009	97.69	7416.1
0.100	97.52	7452.4
0.040	97.78	7906.0
0.005	97.78	8018.7
0.080	97.42	8032.8
0.008	97.17	8075.6
0.003	97.66	8178.7
0.007	97.79	8269.1
0.020	95.75	8365.5
0.001	96.37	8389.5
0.004	97.07	8448.5
0.200	97.80	8693.5
0.030	95.94	8697.6
0.050	96.38	8720.3
0.060	97.23	8807.7
0.006	96.79	8874.5
0.070	97.71	8980.4
0.090	97.47	9028.0
0.300	97.78	9054.8
0.400	96.70	9267.5
0.010	97.21	9292.8
0.002	97.65	9344.9
0.500	49.81	10000.7

Table 6-2 : Synthetic Data Parameter Fitness Value

6.4 Results

Experiment results present the model prediction accuracy, Weighted Kappa statistic (WK), convergence point, algorithm runtime, and resultant clusters for both datasets experiments. Statistical values of the results for each modelling strategy are tabulated for comparison.

6.4.1 *K*-means Results

Table 6-3 and 6-4 tabulate the results for *K*-means experiments for both datasets visual field and synthetic data respectively. The results show that classification accuracy using *K*-means clustering in the synthetic data experiment are less than 50% accurate. Whilst in visual field data experiment, the best average prediction accuracy is 84.47% with NoFCV.

Modelling Strategy/Result	**10FCV**	**2FCV**	**NoFCV**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	82.49	82.91	83.17
Maximum	85.89	85.45	86.27
Average	84.30	84.22	84.47

Table 6-3 : Predictive Accuracy of *K*-means for Visual Field Data

Modelling Strategy/Result	10FCV Accuracy% (WK)	2FCV Accuracy% (WK)	NoFCV Accuracy% (WK)
Minimum	39.90	41.19	42.08
Maximum	57.19	57.14	57.52
Average	48.21	48.01	48.61

Table 6-4 : Classification Accuracy of *K*-means for Synthetic Data

6.4.2 The 6NFB Results

There are slightly lower accuracies found in the 6NFB experiment on visual field data compared to the *K*-means experiment. On average, the predictive performance using 6NFB are 83.27%, 83.90%, and 83.65% with 10FCV, 2FCV and NoFCV respectively.

Modelling Strategy/Result	10FCV Accuracy%	2FCV Accuracy%	NoFCV Accuracy%
Minimum	80.65	82.98	82.41
Maximum	85.31	84.67	85.00
Average	83.27	83.90	83.65

Table 6-5 : Predictive Accuracy of the 6NFB (Visual Field Data)

6.4.3 Predictive Accuracy: Visual Field Data

Both methods predict the visual field data better using the 10-fold cross validation strategy than the other modelling strategies with 88.48% accurate (Table 6-6 - SA) and 88.54% accurate (Table 6-7 - GSA). The results of visual field experiments show that the

GSA method improves the prediction accuracy better than the SA method with all modelling strategies where the highest accuracy is 87.89% (average) in 10FCV (Table 6-7). From the result, it is noted that the average accuracy of 2FCV strategy is slightly lower than 10FCV and NoFCV. The WK values (average) in all experiments present poor agreement with the 6NFB that is less than 0.005. The best average WK value is recorded by the GSA method in 2FCV (0.004 - Table 6-7).

Modelling Strategy/Result	**10FCV** **Accuracy% (WK)**	**2FCV** **Accuracy% (WK)**	**NoFCV** **Accuracy% (WK)**
Minimum	86.32	85.94	86.27
Maximum	88.48	87.09	87.47
Average	87.76 (0.002)	86.45 (0.002)	86.63 (0.001)

Table 6-6 : Prediction Accuracy of SA for Visual Field Data

Modelling Strategy/Result	**10FCV** **Accuracy% (WK)**	**2FCV** **Accuracy% (WK)**	**NoFCV** **Accuracy% (WK)**
Minimum	87.40	86.23	85.89
Maximum	88.54	87.17	87.28
Average	87.89 (0.001)	86.63 (0.004)	86.65 (-0.005)

Table 6-7 : Prediction Accuracy of GSA for Visual Field Data

Confidence intervals (95%) of the accuracy were computed for these 25 experiments. Tables 6-8 and Table 6-9 show the confidence intervals for SA and GSA respectively.

The 10FCV strategy has a higher accuracy range (upper and lower limits) in both methods. With the 10FCV modelling strategy, the GSA experiments show with a 95% confidence that prediction accuracy is between 88.00% and 87.77%. Computation of confidence interval is discussed in **Appendix 6-C**.

Confidence Interval	**Modelling Strategy**		
	10FCV	**2FCV**	**NoCV**
Upper Limit	87.92	86.56	86.74
Lower Limit	87.59	86.34	86.52

Table 6-8 : Confidence Interval of SA for Visual Field Data

Confidence Interval	**Modelling Strategy**		
	10FCV	**2FCV**	**NoCV**
Upper Limit	88.00	86.74	86.79
Lower Limit	87.77	86.53	86.52

Table 6-9 : Confidence Interval of GSA for Visual Field Data

6.4.4 Predictive Accuracy: Synthetic Data

Table 6-10 and 6-11 show the prediction accuracy of the synthetic data by modelling strategy for the SA and GSA method respectively. As opposed to the visual field data experiments, classification accuracy on the synthetic data is highly accurate in the NoFCV modelling strategy for both methods (98.55% - SA and 98.59% - GSA). The synthetic results do not show any pattern in classification accuracy in both methods and modelling strategies. However, the SA method with 10FCV outperforms the GSA method with 10FCV. Whilst with 2FCV and NoFCV, the GSA method accuracies are higher than the SA method. The significant difference between these two methods on synthetic data can

be seen in the WK values. It can be concluded that GSA recorded almost perfect WK (near 1.0) which agrees with the expected synthetic clusters of the 45 variables.

Modelling Strategy/Result	**10FCV**	**2FCV**	**NoFCV**
	Accuracy% (WK)	**Accuracy%(WK)**	**Accuracy% (WK)**
Minimum	98.04	96.76	98.24
Maximum	98.46	97.04	98.60
Average	98.36 (0.868)	97.02 (0.987)	98.55 (0.940)

Table 6-10 : Prediction Accuracy of SA for Synthetic Data

Modelling Strategy/Result	**10FCV**	**2FCV**	**NoFCV**
	Accuracy% (WK)	**Accuracy %(WK)**	**Accuracy% (WK)**
Minimum	97.93	96.87	98.36
Maximum	98.56	97.04	98.60
Average	98.33 (0.884)	97.04 (0.999)	98.59 (0.965)

Table 6-11 : Prediction Accuracy of GSA for Synthetic Data

Table 6-12 (SA) and Table 6-13 (GSA) show the 95% confidence intervals of the accuracy for the synthetic dataset. The confidence interval (CI) in the synthetic experiments was found to have a higher range in NoCV strategy in both methods (SA and GSA). The higher CI in the accuracy of modelling is potentially due to overfitting, and this can be prevented with a number of fold cross validation (Hsu, Chang and Lin, 2003).

Confidence Interval	Modelling Strategy		
	10FCV	**2FCV**	**NoCV**
Upper Limit	98.40	97.05	98.59
Lower Limit	98.31	97.00	98.51

Table 6-12 : Confidence Interval of SA for Synthetic Data

Confidence Interval	Modelling Strategy		
	10FCV	**2FCV**	**NoCV**
Upper Limit	98.39	97.05	98.61
Lower Limit	98.26	97.02	98.57

Table 6-13 : Confidence Interval of GSA for Synthetic Data

Full results of the prediction accuracy are presented in **Appendix 6-D**.

6.4.5 Convergence Point

GSA algorithms are often described as having very fast convergence in the literature (Tsallis and Stariolo, 1996; Penna, 1995; Bohachevsky *et al.*, 1986). Thus, the iteration point at which the search has converged was captured. Since the fitness function of both datasets is noisy (as depicted in Figure 5-4), a rule was established to determine the convergence point by calculating a noisy fitness tolerance. Here a noisy fitness function (Definition 5-1) is defined as one that returns different fitness values each time it is evaluated on the same solution. Noisy fitness is occurred in this work due to both cross validation (different cross validation folds) and to sampling (a random pair of visual field records is selected for each patient for each fitness evaluation). This is not uncommon and other noisy fitness functions can arise due to measurement limitations or the nature of training datasets used in modelling (Varma and Simon, 2006). From the graph shown

in Figure 6-4 (with 10FCV), it clearly shows that the search has converged before reaching the 20,000th iteration, however within the convergence line, the fitness value (classification accuracy) varies.

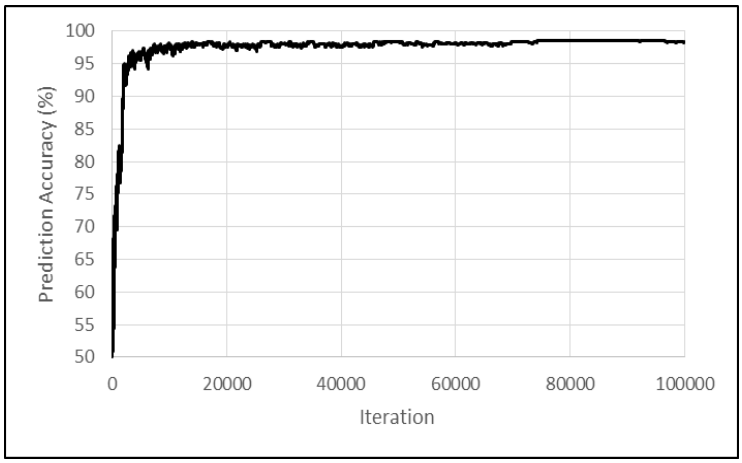


Figure 6-4 : Convergence Graph of the Synthetic Data with the GSA Method

This shows that the GSA rule (acceptance probability), which is applied in the search, has resulted poorer solutions with a certain degree of change are still being accepted. Meanwhile a sample (from experiment number 10) convergence graph for visual field data of the GSA method (with 10FCV) as shown in Figure 6-5 is rather stable from noisy fitness. From the graph (Figure 6-5) it can be understood that an extensive search was happening during the early iterations since the graph has extreme fitness fluctuations. Figure 6-6 shows the close-up of Figure 6-5 in the fluctuation region (iteration 1 to 10,000), which indicates the search converged before reaching iteration 4,000.

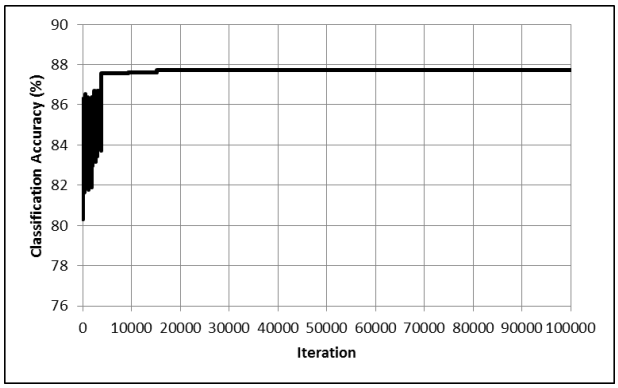


Figure 6-5 : Convergence Graph of Visual Field Data with the GSA Method

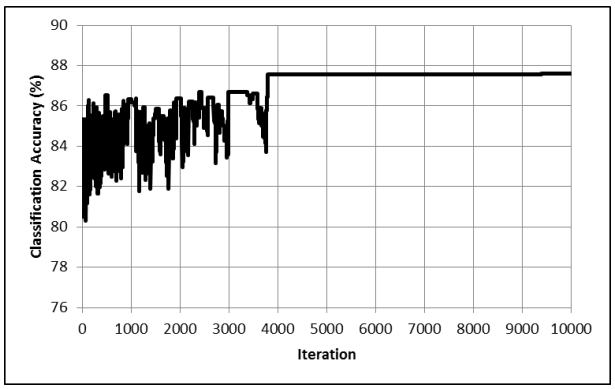


Figure 6-6: The Close-Up of Convergence Graph for Figure 5-5

In order to deal with the inherent noisy nature of the fitness function, an investigation into tolerance limits was conducted. The fitness values in both datasets were found be to normally distributed. To demonstrate this, a simulation was run on both data using the two modelling strategies (10FCV and 2FCV). The simulation experiments were run using the SA and GSA method with 10,000 iterations. Each iteration calls 100 fitness values using the same solution (clusters) to capture the distribution of the fitness. The NoFCV modelling strategy was excluded in this simulation due the same dataset used in training and testing process of data modelling which results the same fitness values in 100 samples of fitness. Algorithm 6-2 shows the simulation steps for the GSA algorithm.

From a simulation in both datasets, it was shown that the fitness values are normal distributed (with p-value 0.49 using Lilliefors test (Lilliefors, 1967)) and Table 6-14 shows the standard deviation of the data by methods and strategies. It is found that standard deviation of fitness within 10FCV experiment is higher than 2FCV. This strongly indicates that the fitness for this strategy is highly noisy. Using these standard deviation values, standard score (*z*-score) for each of the modelling strategy were calculated in order to derive a noisy fitness tolerance limit.

Method/Modelling Strategy	**10 FCV**		**2 FCV**	
	VF	**Syn.**	**VF**	**Syn.**
SA	0.77	0.59	0.18	0.23
GSA	0.79	0.54	0.23	0.20

Table 6-14: Standard Deviation of the Data

Algorithm 6-2: Simulation for Fitness Distribution (GSA)	
Input:	D = {Visual Field Data, Synthetic Data} iterations = Number of iterations fd $\in${10-fold, 2-fold} temp = initial temperature qa = acceptance index qv = visiting index Model= MultinomialNaïveBayesUpdatable
1	Let $C_{current}$ = a random cluster of visual field points
2	Let $D_{current}$ = D of $C_{current}$
3	Let $fitness_{current}$ = prediction accuracy of $D_{current}$ classification with fd
4	For i=0 to iterations-1
5	Calculate newvisit (Equation 6-3)
6	Calculate small_change = newvisit × number of variable
7	C_{new} = re-arrange $C_{current}$ for small_change
8	D_{new} = D of the C_{new}
9	$fitness_{new}$ = prediction accuracy of D_{new} classification with fd
10	if $fitness_{new} > fitness_{current}$
11	$fitness_{current} = fitness_{new}$
12	$C_{current} = C_{new}$
13	else
14	$\Delta fitness = fitness_{current} - fitness_{new}$
15	Calculate Pr (Equation 5-1)
16	Let random = UR(0,1)
17	if Pr > random
18	$fitness_{current} = fitness_{new}$
19	$C_{current} = C_{new}$
20	end if
21	end if
22	For i=0 to 99
23	D_{new} = D of the $C_{current}$
24	fitness = prediction accuracy of D_{new} classification with fd
25	end for
26	Calculate temp (Equation 5-2)
27	end for
Output:	$C_{current}$, prediction accuracy = $fitness_{current}$

The noisy fitness tolerance limits were calculated with z-score value 1.98 which equivalent to 97.61% of the data lies under the defined limits (Table 6-12). **Appendix 6-E** shows the calculation of the noisy fitness tolerance. The noisy fitness tolerance limits tabulated in Table 6-15 are used to determine whether the change of fitness value is significant change or not. At any point of the search that has change in the fitness value (ΔF) within the limit are considered has no significant change. The different of fitness value is calculated using Equation 6-7.

$$\Delta F = F_{t+1} - F_t \qquad \textbf{Equation 6-7}$$

Method/Modelling Strategy	**10 FCV**		**2 FCV 10 Repeats**	
	VF	**Synthetic**	**VF**	**Synthetic**
SA	1.539	1.182	0.347	0.464
GSA	1.569	1.065	0.453	0.392

Table 6-15 : Noisy Fitness Tolerance

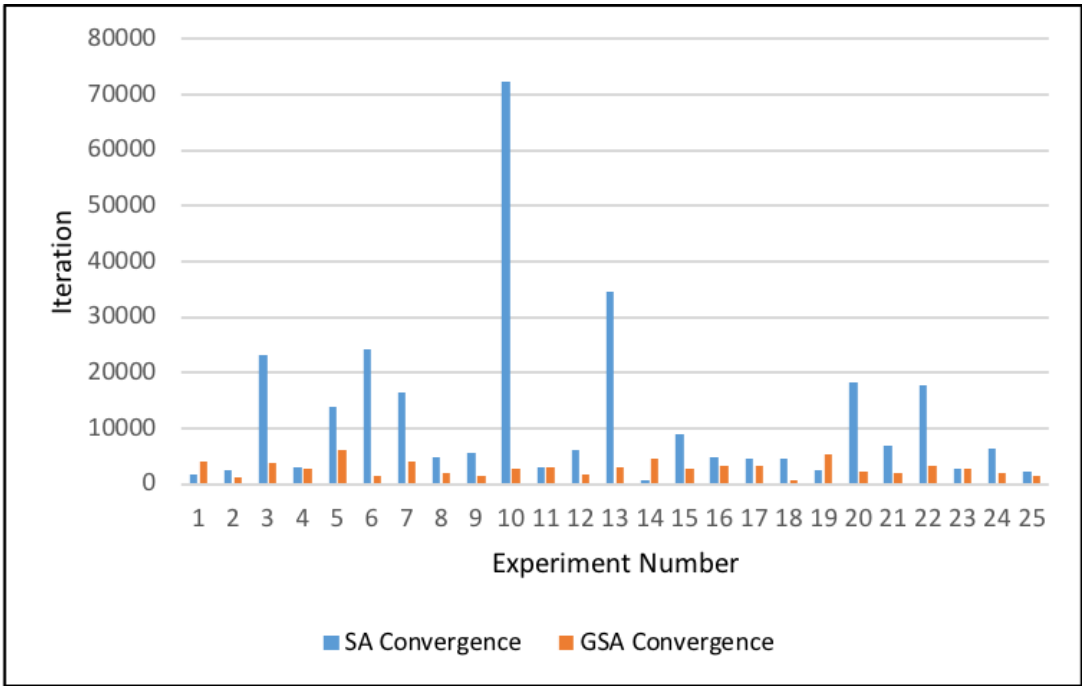
Table 6-16 tabulates the average of convergence point of the individual experiments (visual field data). It shows that the GSA method has converged fastest with 10FCV that is in average at iteration 2,839. This also can be seen in Figure 6-7 for the 25 experiment repeats convergence point (GSA) in comparison with the SA method. However, with the 2FCV modelling strategy in the visual field data, the SA method is far outperformed the GSA method with average 19,938 (GSA: 34,889). The synthetic data convergence point results (Table 6-17) are found consistent with the visual field data. The best convergence point for the synthetic data was the GSA method with 10FCV (7,328).

Modelling Strategy/Method	10 FCV	2 FCV	No FCV
SA	11,643.16	19,938.04	47,304.00
GSA	2,839.88	34,889.04	73,928.08

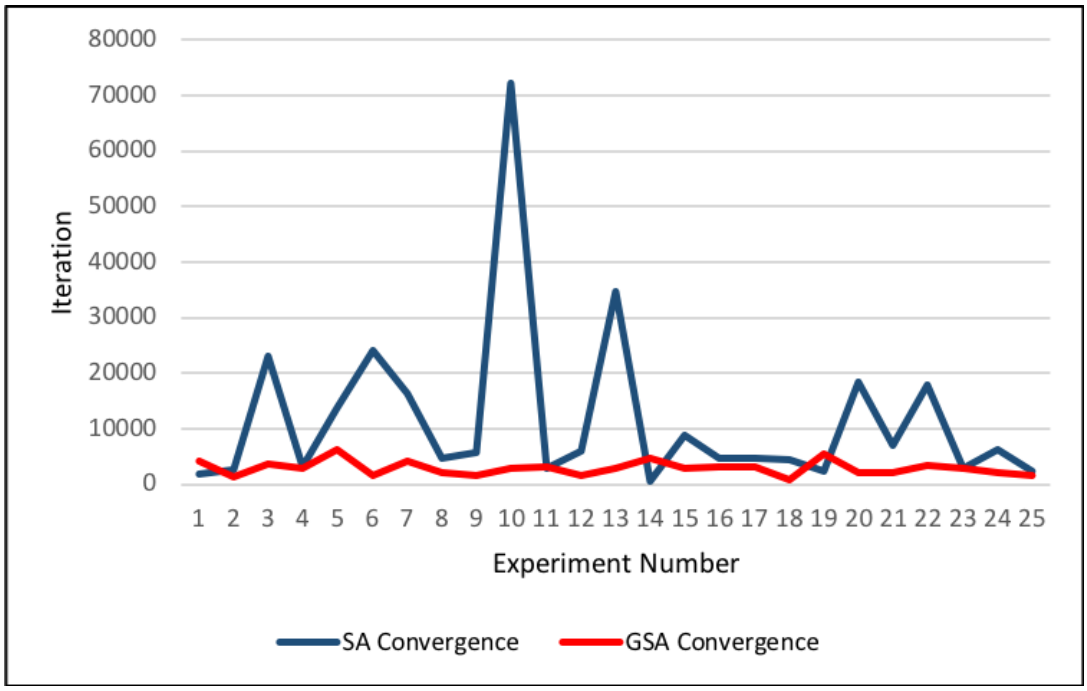
Table 6-16 : Average Convergence Point of Visual Field Data

Modelling Strategy/Method	10 FCV	2 FCV	No FCV
SA	19,582.4	30,365.04	44,843.68
GSA	7,328.00	45,244.56	71,007

Table 6-17 : Average Convergence Point of Synthetic Data



(A)



(B)

Figure 6-7 : Convergence Point of Visual Field Data for 10FCV (A: Histogram, B : Line Graph)

6.4.6 Algorithm Runtime

The algorithms' runtimes for the SA and GSA experiments are computed to support the convergence point results. The runtimes of the algorithms are captured in second which are the total effort of the experiment and the effort of classification. Classification effort, which is measure in second, includes data preparation within the iteration loops of the experiment. The algorithm effort (SA and GSA) in terms of runtime is obtained by subtracting the total runtime and classification runtime. The reason being is that based on experimental observation, classification effort in SMC is not part of the algorithms (SA and GSA) process and classification effort (also includes data preparation) is dependent with number of moves which is more complex with high number of moves. Table 6-18

shows GSA algorithm effort is 1.66% out of the entire SMC runtime. This indicate that a vast amount of the algorithmic computation time is taken up by the fitness function and thus the overheads of the specific search technique have little impact.

Method	Total Experiment Runtime	Classification Runtime	Algorithm Effort (%)
SA	41,860.86	39,180.37	6.40%
GSA	128,120.47	125,985.66	1.66%

Table 6-18 : Algorithm Runtime

6.4.7 Resultant Clusters

The clusters size of each experiment was captured to see the range of clusters size searched by the algorithms. It can be seen that the visual field data experiments with the SA method has a tendency to get lower clusters size when the number of fold cross validation is decreasing. There are 12-18 (minimum-maximum), 7-17, and 6-15 clusters sizes with the mode 15, 12 and 10 for 10FCV, 2FCV and NoFCV respectively. However, smaller cluster size ranges are found in GSA than SA between 3-12 (minimum-maximum) for both 10FCV and 2FCV, and 3-14 for NoFCV. All modelling strategies of the GSA method have the same mode value of clusters size that is 7 clusters.

For the synthetic data, the ranges of cluster size are 4-6 (minimum-maximum), 3-4, and 4-7 in SA method for 10FCV, 2FCV, and NoFCV respectively. Knowing that the correct cluster size for the synthetic data is 3 clusters, it can be claimed that the GSA method searches the clusters better than the SA method with the ranges of cluster size: 4-5,

(minimum-maximum), 3-4, and 4-4 for 10FCV, 2FCV, and NoFCV respectively (with mode: 4, 3, and 4). Also, 2FCV strategy is the efficient strategy finding the right cluster in the synthetic data when the WK value is very near to 1.0.

6.5 Discussion

The GSA algorithm accuracy results (average accuracy) appear to be the highest accuracy even though there are tiny differences between GSA and SA (Table 6-19). Based on Table 6-19, it can be seen that SMC with SA and GSA optimisation method improves predictive accuracy from the benchmark methods.

Modelling Strategy/Method	**10FCV**	**2FCV**	**NoFCV**
	Accuracy%		
K-means	84.30	84.22	84.47
6NFB	83.27	83.90	83.65
SA	87.76	86.45	86.63
GSA	87.89	86.63	86.65

Table 6-19 : Average Predictive Accuracy Comparison

Empirical observation on the three modelling strategies found that 2FCV appears to be the most efficient modelling strategy compared to the others in both methods (SA and GSA). Even though the average prediction accuracies of the strategy are not as best as 10FCV and NoFCV, the WK value in the synthetic data is substantial evidence. Due to the noisy fitness in the data, the 10FCV may have a bias element whereas NoFCV may have overfitting in modelling (Varma and Simon, 2006).

However the best method to produce high accuracy in the real data is 10FCV and this corresponds to (Kohavi, 1995). Also in the synthetic data, with the knowledge about the correct elements of 3 clusters in hand, the accuracy value is 97.04% with WK value 1.0. Therefore, accuracy values that are more than 97.04% are potentially overfitting and have an element of noisy fitness. However there is a slight inconsistency in the results for the convergence point of the search in both methods and data. The GSA method appeared to converge very fast with 10FCV in both datasets. In the visual field data experiment, the GSA method was 8.80% faster than the SA, and 12.25% faster in the synthetic data. In contrast, the SA method is faster than the GSA in 2FCV in both data (14.95% and 14.88% visual field and synthetic respectively). Additionally, a positive result of GSA was found in the runtime analysis on the algorithms that GSA is 4.74% more efficient than SA.

Figure 6-8 and 6-9 show the convergence graph for the SA and GSA search in 10FCV experiment respectively. From the graphs, it clearly shows that GSA has a smooth graph and converges earlier than SA. The convergence graph for SA has very high fluctuation compared to GSA. However, the GSA search is affected by noisy fitness as there is tiny fluctuation in the graph toward the end of the search.

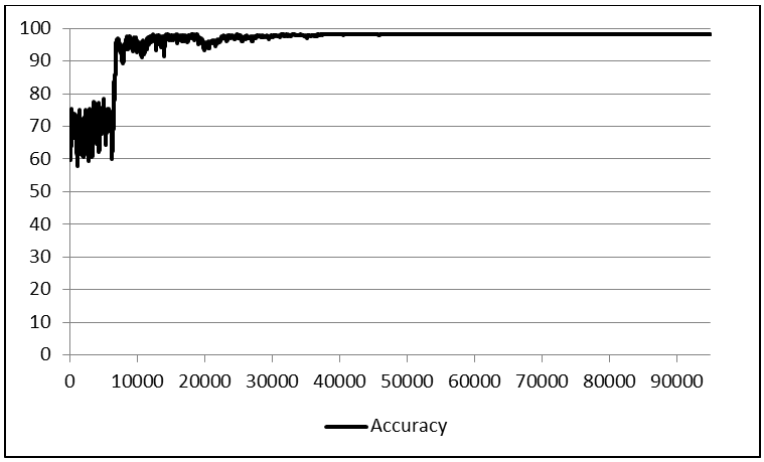


Figure 6-8 : Convergence Graph for SA Method in Synthetic Data (10FCV)

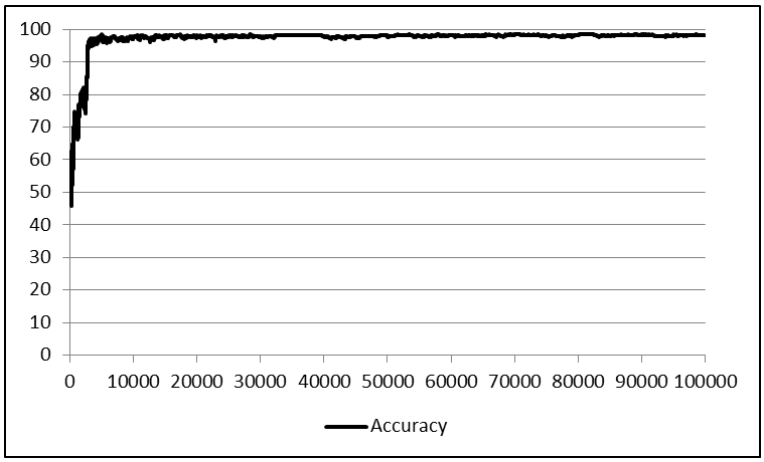


Figure 6-9 : Convergence Graph for GSA Method in Synthetic Data (10FCV)

Another concern in this work is that of degeneracy (Definition 6-1). From the results it was found that the GSA method suffered from degeneracy (Table 6.20). In biology systems, degeneracy occurs when distinct structure of a solution can perform similar functions. The GSA method search is found less efficient due to getting the same clusters quality in the search. Observations to the experiments outputs also found that the method tends to get the same solutions in the search.

Definition 6-1 : Degeneracy

Degeneracy is defined as a mechanism of different characteristic or the ability of elements that structurally different to perform the same function or yield the same output (Edelman and Gally, 2001).

From the table below (Table 6-20), it shows that with 5 clusters which have 88.888% accurate been searched two times in early and middle iteration of the search (8,171 and 33,067). Other records with 5 clusters which have 97.677% accurate been searched three times at iteration 8,501, 8,525, and 8,532. Interesting result was found that with the GSA method, high accuracy (97.765%) was searched in early iteration (7,817), but due to

acceptance probability was high in the early search, the new worse solution been accepted. Then the other solution with different configuration (4 clusters) was found at iteration 54,820 with the same value of accuracy as iteration 7,817.

Iteration #	New Fitness (Accuracy %)	No. of Clusters
8,171	88.888	5
33,067	88.888	5
8,501	97.677	5
8,525	97.677	5
8,532	97.677	5
7,817	97.765	6
54,820	97.765	4

Table 6-20: Excerpt of the Experiments Output of Synthetic Data Affected by Degeneracy

In addition to the analysis of the experiments results on prediction, convergence and resultant clusters size, extended analysis was performed to visualise the resultant clusters by mapping the visual field locations on the visual field grid map. Mapping the visual field locations on the grid is to comprehend the pattern of visual loss suggested by the algorithms. The resultant clusters with the highest prediction accuracy of the GSA were selected for this analysis. Therefore, the resultant clusters from the GSA method with 10FCV (88.54% accurate with 8 clusters) is visualised in the 54 locations visual field grid map (Figure 6-10). From the visualisation of the clusters, the larger clusters size appears on the periphery of vision. This can be seen from Figure 6-10 that clusters 2 and 4 are on the periphery of the visual field grid. Also, it exhibits that cluster 2 locations are near to the blind spot as well as cluster 5 which only in the center of the grid. These findings positively correspond to the clinical evidence that glaucoma first starts at the periphery and near to the blind spot.

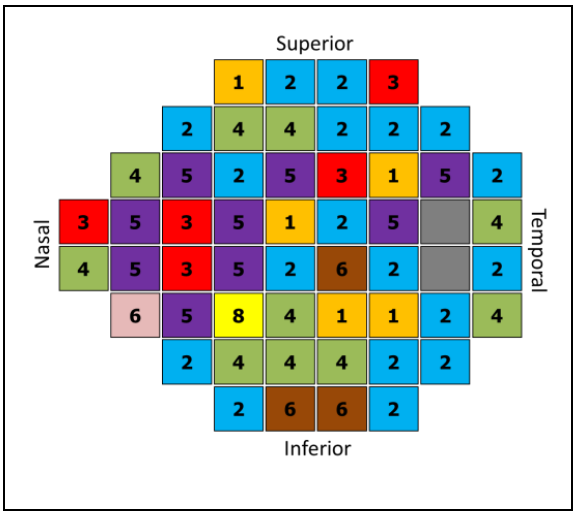


Figure 6-10 : Visual Field Map For Resulting Clusters GSA 10 FCV

6.6 Summary

Chapter 6 presented the application of GSA method on the discrete optimisation problems, SMC. The aim of this work was to further explore SMC on the real data by investigating the arguably efficient search method, the GSA. With the GSA algorithm, SMC has appeared more efficient than the SA method. This was attributed to the extensive search which allowing the method to have a large number of moves in early iterations of the search. Other than that, the parameters used in GSA calibrate the acceptance probability and artificial temperature which makes the method becomes more efficient. These properties in the GSA method make the method distinct from the other annealing methods to produce good results. The GSA method has shown great search results on the synthetic dataset with high WK. As for accuracy, the method was found significantly improves the prediction accuracy in the real data compared to SA and the benchmarks methods (*K*-means and the 6NFB). In terms of efficiency, the GSA method has proven more efficient compared to the SA method with early convergence in the search. Besides the convergence analysis, the GSA algorithm runtime was found more efficient than SA.

Interesting results were found with the three modelling strategies implemented in this work. The 2FCV strategy has shown outstanding results with high WK. Therefore, these modelling strategies will be used in the next chapter of this study. Moreover, the visual field locations mapping of the SMC result was found coincides with the clinical evidence on glaucoma deterioration.

This work was successfully implemented the GSA method in SMC with notable improvement results. However, besides noisy fitness, SMC is still deficient in performance when applying the GSA search method in the discrete problem. Optimising discrete problems such as SMC involves a combinatorial elements arrangement which this task can be affected by degeneracy. There is sufficient evidence shown in this work that the method has suffered from the degeneracy. Degeneracy can lead to inefficient exploration of a search space as the same clustering arrangements are repeatedly revisited. To overcome this problem, a certain representation of getting a new solution in a search is to be used in the algorithm as a vehicle for removing degeneracy. Therefore, Chapter 7 will present a fine-tuned algorithm where an improvement is to be made on the selection of a new solution in the search.

Chapter 7

Restricted Growth Function based Generalised Simulated Annealing to Predict Glaucoma using Simultaneous Modelling and Clustering

7.1 Overview

Since the application of the GSA method towards the SMC problem was found less efficient in the previous chapter due to degeneracy, enhancing the algorithm is thus essential. Chapter 7 introduces an algorithm that incorporates Restricted Growth Functions (RGFs), which are capable of vastly reducing degeneracy (Swift *et al.*, 2007), in the search procedure. An RGF can be used to represent a clustering arrangement such that there is a one to one mapping between the clustering arrangement and representation, hence no degeneracy. The main expectation is that, SMC would be able to find the known clusters in the synthetic data much faster using the RGFGSA method than the existing methods as well as improving prediction accuracy in the real data. Moreover, RGFGSA is expected to have a less noisy fitness than the existing methods in the search.

The algorithms used in the preceding chapters obtain a new solution (cluster arrangement) by implementing random moves exchanging variables among clusters based on a small change value (a number of moves). However, this process may lead to the repetition of the same solutions in the search (degeneracy) due to the representation being used. Therefore, this work adopts RGFs in the GSA search procedure for selecting a new solution.

This work is a continuation from the previous chapter where the same datasets and experiment's setup are used to test the enhanced algorithm. Moreover, additional experiments are carried out such as manipulating the RGFs procedure (namely, *kN* experiment) and testing SMC (using RGFs) on subsets of visual field data: early record, middle record and latest record. These supplementary experiments and analyses are aimed

at discovering any other latent knowledge within the data and carrying out more exploration on the SMC technique.

The chapter is presented into six sections. Section 7.2 demonstrates combinatorial optimisation and highlights the need of this work. Section 7.3 introduces a novel algorithm namely, RGFGSA, which is the improved version of GSA, with the adaption of RGFs. This section also presents results of experiments using the RGFGSA algorithm. Additional experiments on the algorithm are presented in section 7.4 and 7.5. Finally, Chapter 7 summarises the work in section 7.6.

7.2 Combinatorial Optimisation

The application of SMC in visual field data presented in the earlier chapters involves a process where visual field locations are re-arranged to form a good combination of the variables (in clusters) that accurately predict the AGIS score. In relation to this process, it refers to combinatorial optimisation (Papadimitriou and Steiglitz, 1982) which searches for the best elements arrangement of some finite sets of discrete items such as a travel salesman problem (TSP). Solving these problems is typically NP-hard (Klein and Young, 2010). Hence, many researchers focus on inventing a heuristic algorithm that efficiently garners a good solution to problems. An efficient and effective algorithm requires a good representation of solution in a search. To solve the Bin Packing problem (Garey and Johnson, 1979) for instance, a group of objects are used to represent the partitions of objects (Falkenauer and Delchambre, 1992). Meanwhile for the TSP problem, a list of permutations of distinct objects is used to represent the sequence of cities (Skiścim and Golden, 1983; Lin, 1965).

In developing an algorithm for solving combinatorial optimisation problems, one must be aware of possible solutions that suffer from degeneracy. In the genetic algorithm, multiple chromosomes that have the same solution are termed as degeneracy (Radcliffe and Surry,

1994). This issue consequences an algorithm becoming less efficient as the same configurations of solutions are repeatedly obtained in the search due to the inherent stochastic nature of an algorithm. Avoidance of degeneracy in a search might help to improve algorithm's search performance with early convergence. This can be attributed to quality solutions with high accuracies being searched for by an algorithm as well as removing repeated solutions. A few studies, particularly on the genetic algorithm method and combinatorial optimisation such as (Jiao and Wang, 2000), (Yau *et al.*, 2003) and (Tucker *et al.*, 2005), have been done to alleviate the degenerative issue. Tucker et al. specifically have used the RGF technique effectively removed degeneracy in a search with promising results.

The motivation of this work is to develop an efficient algorithm for solving a discrete combinatorial problem by removing degeneracy in a search. This is possible to achieve by employing RGF in SMC (M. Z. M. B. Jilani *et al.*, 2016) to find the best cluster arrangement of visual field (VF) points. The RGF representation (Falkenauer, 1998) is a grouping technique used towards the high dimension of VF points clustered in order to classify the glaucoma progression. With this approach, classification accuracy is observed and an early convergence point is captured to determine the effectiveness and the efficacy of the algorithm. Obtaining faster convergence as well as higher classification accuracy of the search is the essential contribution of the RGFGSA algorithm in SMC.

7.3 Restricted Growth Function Generalised Simulated Annealing

Restricted Growth Function

Unlike the approach of representing a solution (clustering arrangement) in the previous algorithms (Chapter 4, 5 and 6), the algorithm used in this work applies Restricted growth functions to represent a clustering arrangement of visual field in a list of cluster indices

in a search. RGF is a procedure used in the GSA algorithm where a new clustering arrangement is obtained.

An RGF is a function $f : I_n \to I_n$, where I_n is a list of cluster indices such that:-

$$f(1) = 1,$$

$$f(i+1) \leq \max\{ f(1),..., f(i)\} + 1 .$$

For example, $v_1 = 11213224$ is an RGF, but $v_2 = 11214322$ is not an RGF since $4 > 1 + \max\{1,1,2,1\}$ (Campbell *et al.*, 2016).

RGFs can be used in heuristic search techniques to vastly reduce the degeneracy inherent in other representations (Tucker *et al.*, 2005). Adopting RGFs in the GSA algorithm advantages a search to obtain quality clusters (non-degeneracy clusters) from granularity solutions. An RGF represents clusters in a list of integers that the elements in the list indicate the cluster of its index belong to. The length of the list is the number of objects in clusters and the maximum integer value in the list indicates the number of clusters. Figure 7-1 exhibits the example of RGFs representation for clusters.

Function	Number of Clusters	Clusters
Index 0 1 2 3 4 5 6 7 8 9 RGF 1 1 2 3 1 4 4 3 5 5	5	[[0,1,4],[2],[3,7],[5,6],[8,9]]
Index 0 1 2 3 4 5 6 7 8 9 RGF 1 1 1 1 2 1 2 2 1 3	3	[[0,1,2,3,5,8],[4,6,7],[9]]

Figure 7-1 : RGFs Representation of Clusters for 10 Elements

In this work RGFs are used to represent a candidate of solution in the search. As delineated in Algorithm 6-1 (in Chapter 6), the current solution is perturbed (small change) to get a new solution based on the GSA new visit (denoted by g_{qv}) value. Then the current clusters which is denote by f and the new clusters (g) are represented in the RGF format. Henceforth, hamming distance (Tucker *et al.*, 2005; Hamming, 1950) is used to get the distance between the two RGFs (f and g). The hamming distance between the two RGFs is also termed as a path ($f \rightarrow g$). If the hamming distance between the two RGF points is 5, this means that there are 5 different solutions in the RGF format in the $f \rightarrow g$. Each of the points in the $f \rightarrow g$ has a possibility to be selected based on the g_{qv} value as the next solution. The schematic diagram of potential new solutions in the $f \rightarrow g$ is well explained in Figure 7-2.

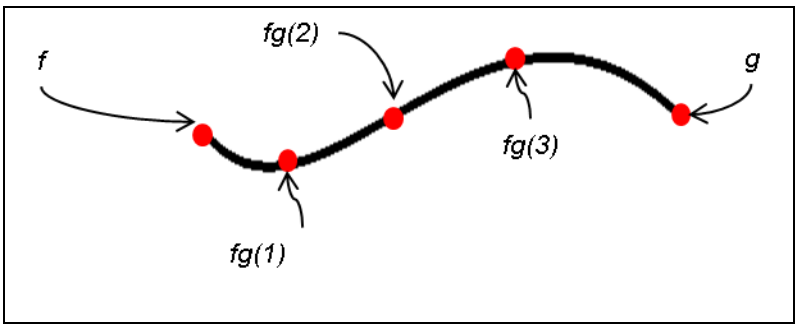


Figure 7-2 : The RGF Path of the New Solution

With the figure above, a new solution which is derived from RGF ($fg(i)$), is obtained by the following simple mathematical calculation:-

Suppose the GSA new visit (from Equation 2-13) is $g_{qv} = 0.75$ and the RGF path length is $f \rightarrow g = 5$. Then the selected RGF point is computed as, $fg(i) = 0.75 \times 5$. In this example, $fg(i) = 3.75 \approx 4.0$ therefore, the RGF number 3 ($fg(3)$) is chosen as the new

solution. In case of 0 value is obtained for g_{qv}, hence RGF solution g is chosen. In Figure 7-3, which is clearly can be seen that the RGF point becomes smaller towards the end of iteration, shows the distribution of selection points in the RGF path for the new solution. This pattern is caused by the new visit (g_{qv}) of the GSA algorithm. This process is defined in Algorithm 7-2.

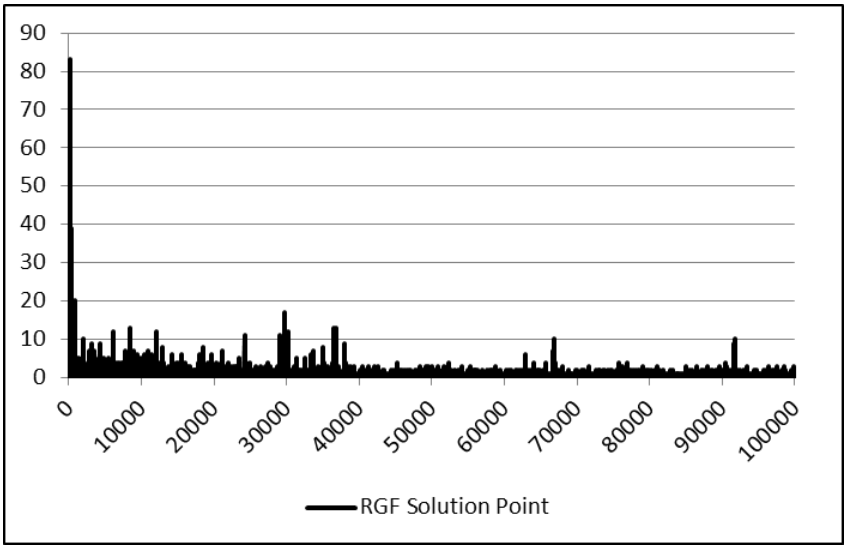


Figure 7-3 : Distribution of Selection Point in RGF Path (VF Data Experiment)

Restricted Growth Function Generalised Simulated Annealing

The RGFGSA algorithm incorporates RGFs in the procedure for obtaining a new solution. As discussed in Chapter 6, the GSA method may have the minimum 2 number of moves and maximum 52 (for visual field data) and 45 (for synthetic data) number of moves in order to generate a new solution. This procedure is remained applied in the RGFGSA method. However, in this procedure, a new solution is obtained from the list of solutions which are generated from the RGF operator as presented in example Figure 7-2. The selection of a new solution from the list of solutions in the RGF path is based upon

the new visit value of GSA (Equation 2-13). Algorithm 7-1 exhibits the RGFGSA method which the procedure of RGFs in illustrated in Algorithm 7-2.

Algorithm 7-1 : Simultaneous Modelling and Clustering of Restricted Growth Function Generalised Simulated Annealing	
Input:	D = visual field data iterations = Number of iterations fd $\in$ {10-fold, 2-fold, no-fold cross validation} temp = initial temperature qa = acceptance index qv = visiting index Model= MultinomialNaïveBayesUpdatable
1	Let $C_{current}$ = a random clusters of visual field points
2	Let $D_{current}$ = visual field data of the $C_{current}$
3	Let $fitness_{current}$ = prediction accuracy of the $C_{current}$ with fd
4	For i=0 to iterations-1
5	C_{new} = get C_{new} using the RGF operator (Algorithm 6-2)
6	D_{new} = D of the C_{new}
7	$fitness_{new}$ = prediction accuracy of data classification with fd
8	if $fitness_{new} > fitness_{current}$
9	$fitness_{current} = fitness_{new}$
10	$C_{current} = C_{new}$
11	else
12	$\Delta fitness = fitness_{current} - fitness_{new}$
13	Calculate Pr (Equation 5-1)
14	Let random = UR(0,1)
15	if Pr > random
16	$fitness = fitness_{new}$
17	$C_{current} = C_{new}$
18	end if
19	end if
20	Calculate temp (Equation 5-2)
21	end for
Output:	$C_{current}$, prediction accuracy = $fitness_{current}$

Algorithm 7-2 : RGF Representation in Small Change of GSA Algorithm	
Input:	$C_{current}$ V = No of variables of visual field data qv temp
1	Let C_{new} = clone of $C_{current}$
2	newvisit = Compute GSA small change (Equation 6-3)
3	Let no. of moves = newvisit × V
4	$clusters_{new}$ = re-arrange $clusters_{new}$ for no. of moves
5	Let $C_{current}$ and C_{new} in RGF format
6	Let f = $C_{current}$ in RGF format
7	Let g = C_{new} in RGF format
8	Generate RGF path from f to g
9	Let RGF length = get the f→g length
10	Let RGF Point = newvisit × f→g length
11	If RGF Point = RGF length or Solution Point = 0
12	RGF solution = g
13	else
14	Let RGF solution = get RGF at RGF Point
15	end if
16	C_{new} = Convert the RGF solution in clusters format
Output:	C_{new} of visual field locations

7.3.1 Experiments

The datasets (real data and synthetic data) and experiment's setup in Chapter 6 are used in this work. As such, the same q_a and q_v values for the GSA method from Chapter 6 are used. Also, the three modelling strategies: 10FCV, 2FCV and NoFCV are implemented in this work. The *K*-means and 6NFB results are also presented in this chapter for comparison. This is an "*apples to apples*" comparison between the results from Chapter 6 and Chapter 7.

7.3.2 Results

Result section presents the predictive accuracy (in percentage), Weighted Kappa (WK), convergence point, algorithm runtime and resultant clusters of the experiments for both VF and synthetic data. The average accuracy and average WK are derived from the 25 set experiments in each method and modelling strategy. The results are tabulated for comparing the three optimisation methods (SA, GSA and RGFGSA) using three modelling strategies.

Predictive Accuracy: Visual Field Data

Table 7-1 shows the results for the 10-fold cross validation (10FCV) modelling strategy. In the average accuracy, the three methods do not have a significant difference. The GSA average accuracy (87.89%) is slightly higher than RGFGSA and SA. However, it was recorded that the best accuracy was from RGFGSA with 88.83% accurate. Note that WK values, which is less than 0.005, for the resultant clusters have very poor agreement with the 6NFB for the three methods in all modelling strategies (Table 7-1-Table 7-3).

Result	**Modelling Strategy**		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	86.32	87.40	86.01
Maximum	88.48	88.54	88.83
Average	87.76 (0.002)	87.89 (0.001)	87.84 (-0.003)

Table 7-1 : 10-fold Cross Validation Experiment Results

With the 2-fold cross validation (2FCV) modelling strategy (Table 6-2), the results are consistent with 10FCV where the average accuracy of the GSA (86.63%) is slightly higher than RGFGSA and SA methods (0.04 and 0.18 respectively). However, the best accuracy was recorded by the RGFGSA with 87.19% accurate.

Result	**Modelling Strategy**		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	85.94	86.23	85.74
Maximum	87.09	87.17	87.19
Average	86.45 (0.002)	86.63 (0.004)	86.59 (-0.008)

Table 7-2 : 2-fold Cross Validation Experiment Results

Meanwhile with the no cross validation strategy (NoFCV) as shown in Table 7-3, the RGFGSA method is slightly better in predictive accuracy with 86.80% accurate than the other two methods. The best accuracy was found by the SA method with 87.47% accurate.

Result	**Modelling Strategy**		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	86.27	85.89	86.33
Maximum	87.47	87.28	87.22
Average	86.63 (0.001)	86.65 (-0.005)	86.80 (-0.0136)

Table 7-3 : No Cross Validation Experiment Results

Predictive Accuracy: Synthetic Data

Classification accuracies in the synthetic data experiments are higher (above 90%) as compared to the visual field data. Table 7-4 shows the 10FCV modelling strategy results for the synthetic data experiments. Within this strategy, the highest accuracy was recorded by the RGFGSA method with 99.16% accurate. Also note that in average accuracy, RGFGSA is the best method in improving the classification accuracy with 98.62% accurate as compared to the other two methods.

Result	**Modelling Strategy**		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	98.04	97.93	93.27
Maximum	98.46	98.56	99.16
Average	98.36 (0.868)	98.33 (0.884)	98.62 (0.915)

Table 7-4 : 10-fold Cross Validation Experiment Results

Likewise with the 2FCV modelling strategy, RGFGSA obtained higher accuracy than the other methods as exhibits in Table 7-5. The RGFGSA method outperformed the other two methods with the best accuracy value 98.04% and the average accuracy 97.87%.

Result	Modelling Strategy		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	96.76	96.87	97.68
Maximum	97.04	97.04	98.04
Average	97.02 (0.987)	97.04 (0.999)	97.87 (1.000)

Table 7-5 : 2-fold Cross Validation Experiment Results

Meanwhile results from the experiments with NoFCV, there are inconclusive results as GSA and RGFGSA have the same average accuracy with 98.59% (Table 7-6). The best accuracy was 98.60% for all three methods.

Result	Modelling Strategy		
	SA	**GSA**	**RGFGSA**
	Accuracy% (WK)	**Accuracy% (WK)**	**Accuracy% (WK)**
Minimum	98.24	98.36	98.48
Maximum	98.60	98.60	98.60
Average	98.55 (0.940)	98.59 (0.965)	98.59 (0.961)

Table 7-6 : No Cross Validation Experiment Results

As opposed to the visual field data results, the resultant clusters in the synthetic data have high agreement of WK (near 1.0) with the known clusters that is three clusters of the 45 variables. The RGFGSA method outperformed the other method with the 10FCV and 2FCV modelling strategies where 2FCV perfectly obtained WK 1.0 in all 25 experiments. However with NoFCV, the GSA method slightly surpassed the RGFGSA method by 0.004 with WK=0.965 (Table 7-6).

Convergence Point

The convergence point is defined as the iteration point at which there are no further fitness improvements. This means that after the convergence point, the search could not found any better improvement solution (maximising prediction accuracy). The earlier iteration the convergence point is, the more efficient the method. In this work, 100,000 iterations are used and the smallest iteration point indicates the best convergence point. Similar to the work carried out in Chapter 6, due to noisy fitness, fitness tolerances were computed for every experiment. Thus, the convergence points for the search are the fitness tolerance convergence points.

- Visual Field Data

Table 7-7 tabulates the average (of 25 experiments) convergence point for the visual field data experiments. The results indicate that the RGFGSA method does not have the best convergence points with the three modelling strategies compared to GSA and SA. GSA recorded the best average convergence with 10FCV (2,839.88) and whilst SA with 2FCV (19,938.04) and NoFCV (46,304.00). However, in this comparison, RGFGSA is the second best method to converge in the experiments.

Method	**Modelling Strategy**		
	10FCV	**2FCV**	**NoFCV**
SA	11,643.16	19,938.04	47,304.00
GSA	2,839.88	34,889.04	73,928.08
RGFGSA	3,590.52	30,859.68	68,446.84

Table 7-7 : Average Fitness Tolerance Convergence Point for Visual Field Data

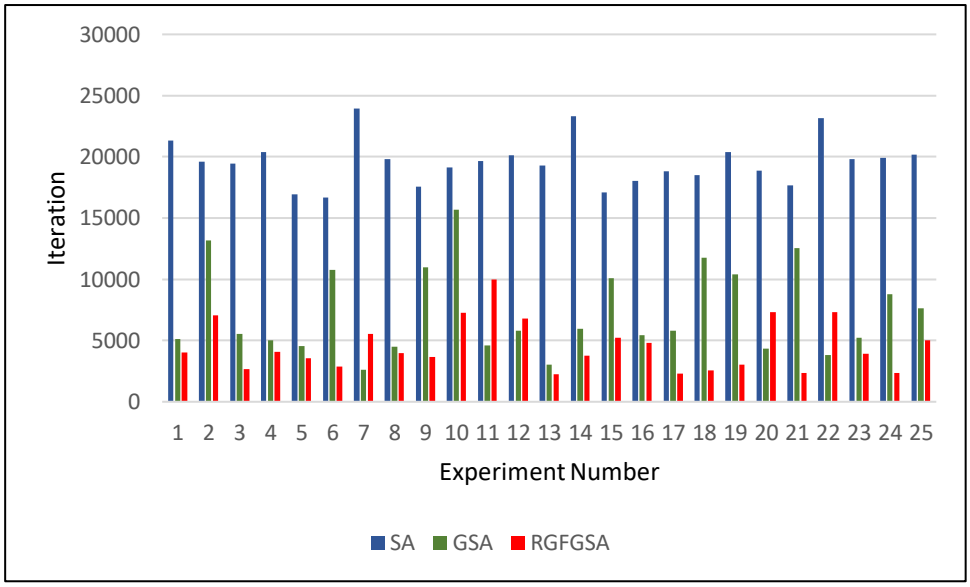
- Synthetic Data

Meanwhile in the synthetic data experiments, the RGFGSA method was found very efficient with 10FCV and 2FCV modelling strategies (Table 7-8). However, the SA method has better convergence with the NoFCV modelling strategy than GSA and RGFGSA.

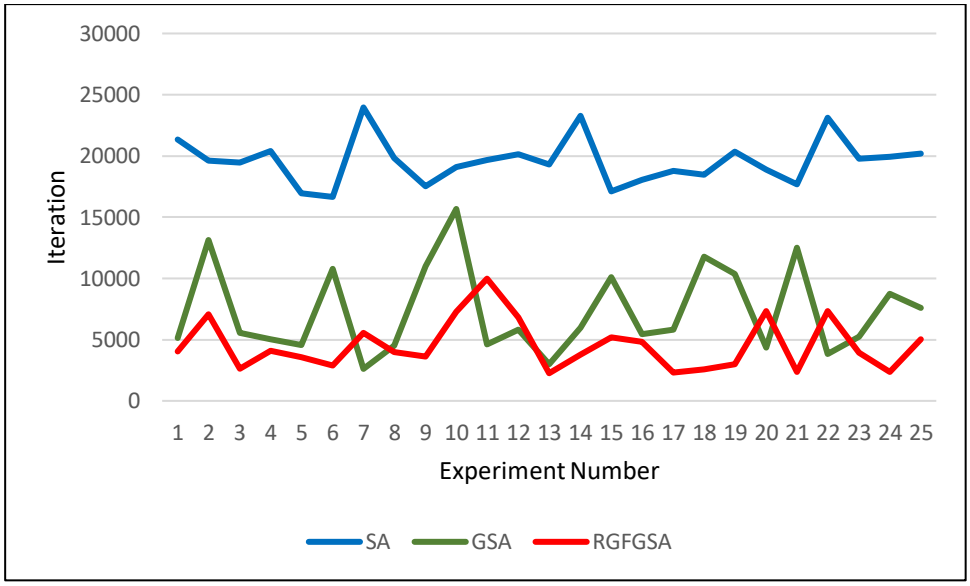
Method	**Modelling Strategy**		
	SA	**GSA**	**RGFGSA**
10FCV	19582.40	7328.00	4551.64
2FCV	30365.04	45244.56	20924.08
NoFCV	44843.68	71007.00	54554.08

Table 7-8 : Average Fitness Tolerance Convergence Point for Synthetic Data

Figure 7-4 exhibits the plot of convergence points for the three methods (with 10FCV) in 25 experiments. From the graph it clearly points out that the RGFGSA method is the fastest algorithm (early convergence) as the plots' line is lower than the other two methods as exhibit in the line graph. The SA method convergence plots are far separated from the GSA and RGFGSA methods.



(A)



(B)

Figure 7-4 : Convergence Point of Synthetic Data for 10FCV

(A: Histogram, B: Line Graph)

- Algorithm Runtime

As presented in Chapter 6, the algorithm runtime for RGFGSA is captured for analysis. Table 7-9 shows that the total runtime (average) of RGFGSA experiments is higher than SA and GSA. As discussed in Chapter 6, classification effort in the experiments took so much time due to the classification process in SMC that dependent on number of moves and data preparation. However, the RGFGSA algorithm's effort in the experiments is only 0.38% from the total runtime.

Method	Total Experiment Runtime	Classification Runtime	Algorithm Effort (%)
SA	41,860.86	39,180.37	6.40%
GSA	128,120.47	125,985.66	1.66%
RGFGSA	584,188.19	581,956.10	0.38%

Table 7-9 : Algorithm Runtime

Resultant Clusters

Resultant clusters of the search were captured to see the significant compound variables of the visual field locations besides the 6NFB. The resultant clusters from the 25 set experiments note that the maximum clusters proposed by SMC were 18 clusters which from SA with 10FCV, and the minimum clusters (2 clusters) were from RGFGSA with 10FCV (Table 7-10). These results show that GSA and RGFGSA have tendency to get smaller clusters in the search in all modelling strategies.

Modelling Strategy/Method	SA		GSA		RGFGSA	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
10-FCV	12	18	3	12	2	10
2-FCV	7	17	3	12	4	11
No FCV	6	15	3	14	4	11

Table 7-10 : Range of Resultant Clusters for Visual Field Data

For the synthetic data (Table 7-11), as the known clusters size of the data is three, the experiments on the synthetic data was seemed to be more effective to search the expected clusters. Note that with all methods and modelling strategies, the resultant clusters obtained by the methods and strategies were less than 8 clusters. The 2FCV modelling results stand out from the other strategies in all three methods, where the minimum clusters were three and the maximum clusters were four. The 2FCV modelling strategy with the RGFGSA method was the most effective method as the known clusters (three clusters) were found in all 25 set of experiments.

Modelling Strategy/Method	SA		GSA		RGFGSA	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
10-FCV	4	6	4	5	3	7
2-FCV	3	4	3	4	3	3
No FCV	4	7	4	4	4	5

Table 7-11 : Range of Resultant Clusters for Synthetic Data

7.3.3 Discussion

Visual Field

In the visual field data experiments, it can clearly be seen that the 10FCV modelling strategy is the most suitable approach (with higher average accuracy) for improving prediction accuracy. This coincides with (Kohavi, 1995) that 10FCV is a good modelling strategy to produce high accuracy of real data. The average GSA accuracy (from 25 experiments) is higher than RGFGSA with 10FCV and 2FCV as presented in Table 7-12. However, the accuracy variance between GSA and RGFGSA is very small (0.05 and 0.04) and RGFGSA results are better than the benchmarks (*K*-means and the 6NFB). Also it can be noted that the RGFGSA method recorded the best accuracy (88.83%) in experiments with 10FCV.

Modelling Strategy/Method	**10FCV**	**2FCV**	**NoFCV**
	Accuracy%		
K-means	84.30	84.22	84.47
6NFB	83.27	83.90	83.65
SA	87.76	86.45	86.63
GSA	87.89	86.63	86.65
RGFGSA	87.84	86.59	86.80

Table 7-12 : Average Predictive Accuracy Comparison

The agreement level of the resultant clusters with the 6NFB was found to be very poor in the experiments with all methods and modelling strategies. As for the method's efficiency, RGFGSA did not record the best convergence (average) in visual field

experiments. However, algorithm runtime analysis found that the method was 1.28% and 6.02% more efficient than GSA and SA respectively.

WKs among the resultant clusters were computed, in order to see similarity and consistency among the resultant clusters that have been suggested by SMC within the 25 set experiments. With the 25 set of the experiment results, there are 300 pairs ($^{n}C_{r}$: $n = 25$, $r = 2$ **Appendix 7-A**) of the resultant clusters for computing WK. This means that WK is computed for resultant clusters 1 and 2, resultant clusters 1 and 3, and so on and so forth. Then the average value is obtained from those 300 WKs.

The WK (average) results among clusters were appeared to very poor in visual field data experiments. This poor WK confirms that there are inconsistency resultant clusters in the 25 set experiments.

However, the RGFGSA's resultant clusters have slightly higher WK than the other methods with 0.00561 and 0.01138 for 2FCV and NoFCV respectively (Table 7-13).

Method/Modelling Strategy	**SA**	**GSA**	**RGFGSA**
10FCV	0.00149	0.00298	0.00126
2FCV	0.00273	0.00205	0.00561
NoFCV	0.00415	0.00497	0.01138

Table 7-13 : Average of WK among the Resultant Bundles for VF Data

Synthetic Data

Meanwhile in the synthetic data experiments, SMC performed very well with the RGFGSA method. RGFGSA has appeared to be the best method with high average accuracy in all modelling strategies. As opposed to the visual field data (with 10FCV), the best modelling strategy to improve classification accuracy in the synthetic data is 2FCV.

The WK results in the synthetic data have shown promising results with high average WK with all methods and modelling strategies. The 2FCV modelling strategy was found the most efficient strategy. This was supported by the results that RGFGSA has found the known clusters in all 25 set experiments. As all methods were effective to search the known cluster in the synthetic data, the average WK results among the resultant clusters are very good as shown in Table 7-14. The RGFGSA method has produced high consistent resultant clusters with the 2FCV within the 25 set experiments (WK= 1.0).

Method/Modelling Strategy	**SA**	**GSA**	**RGFGSA**
10FCV	0.8354	0.8540	0.8530
2FCV	0.9735	0.9974	1.0000
NoFCV	0.9275	0.9903	0.9893

Table 7-14 : Average of WK among the Resultant Bundles for Synthetic Data

In terms of efficiency, it can be concluded that the RGFGSA method is more efficient than the other methods with 10FCV and 2FCV (as shown in Table 7-8). However with NoFCV, RGFGSA was found the second best in convergence point results after

SA. Note that the RGFGSA method was 7.33% and 24.32% more efficient than the GSA method in the synthetic data with 10FCV and 2FCV modelling strategy respectively. Whilst in the comparison with the SA method, RGFGSA was 15.03% and 9.44% more efficient in 10FCV and 2FCV respectively.

From the NoFCV experiments' results, it can be seen that the average accuracy in the synthetic data are high (all methods). This may due to overfitting in modelling (Varma and Simon, 2006). Moreover, it was noticed that the search have suffered from noisy fitness in all modelling strategies. Noisy fitness is illustrated in Figure 7-5 and Figure 7-6 for RGFGSA and GSA respectively. However, the fitness of the RGFGSA method was found very less noisy as compared to the GSA method. Noisy fitness happens when a worse solution (with a worse fitness value) is still being accepted in the search even though the search has converged. This phenomenon affects the search performance. As RGFGSA was found less noisy in fitness, the method is thus able to obtain higher WKs in the experiments.

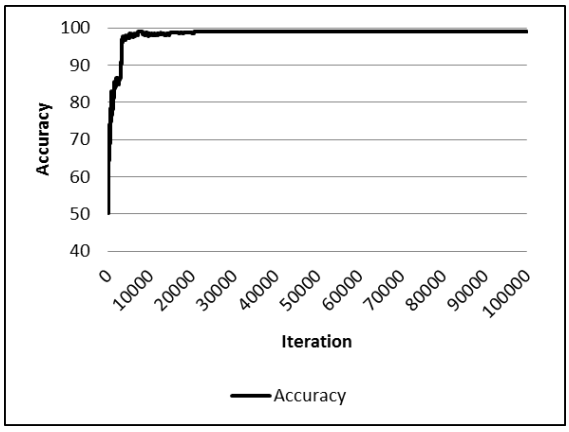


Figure 7-5 : Convergence Graph Synthetic Data in RGFGSA Method (10FCV)

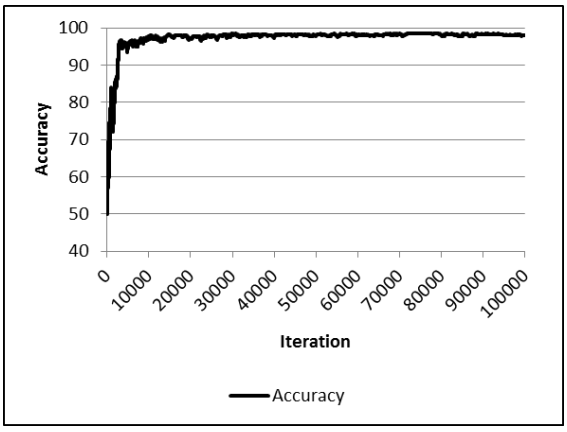


Figure 7-6 : Convergence Graph Synthetic Data in GSA Method (10FCV)

Besides the analyses on the algorithm performance, resultant clusters visualisation on the visual field grid is observed and comprehended according to the clinical practices. One of the resultant clusters from the experiments was chosen for visualisation on the visual field grid map. The best accuracy of 10FCV (88.83%) was chosen and illustrated on the visual field grid map as shown in Figure 7-7. The resultant clusters have four clusters which most of the visual field locations are in the same group in cluster one. The biggest cluster of visual field locations is mainly positioned at the peripheral of the grid. It is the same with cluster four which three of the locations are situated at the peripheral. This pattern of visual field location is corresponding to the clinical evidence that the onset of glaucoma are in the periphery of vision and near to the blind spot (Broadway, 2012; Heijl and Patella, 2002).

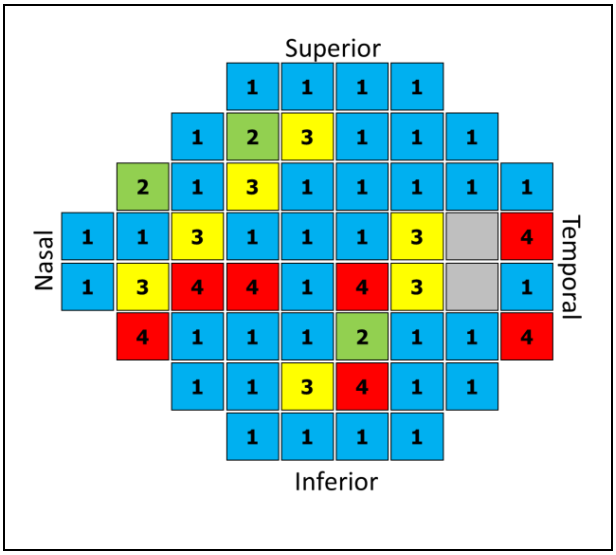


Figure 7-7 : Visualisation of Resultant Clusters on VF Grid Map

7.4 *kN* Experiments

The *kN* experiments present a new strategy in the RGF procedure by manipulating the small change value, where *k* is a constant and *N* is the length of dataset variables. Small

change in the RGFGSA algorithm is obtained as a result of multiplying the new visit value with the length of variables in the data (synthetic: 45 variables). However, this piece of work manipulates the length of the dataset variables by applying a range of constants starting from 0.5 to 5.0 to be multiplied with the length of variables. Small value of the constant (*k*) would result less number of moves in small change and vice versa. Implementing *kN* in experiments is to observe the effect of RGFGSA algorithm which lesser or larger number of moves might be influencing the efficacy of the algorithm in exploring the search for clustering. The motivation of having this strategy in the RGF procedure is that, to see whether a larger or lesser number of moves potentially improves the search performance with high WK and predictive accuracy as far as reverse-engineer is concerned. This experiment is tested on the synthetic data owing to the known clusters that enables to measure the efficacy of the experiment strategy with higher WK as well as improving predictive accuracy. The hypothesis behind this experiment is larger exploration (high number of moves) in the search at early iteration would potentially more efficient for the algorithm to find the best clusters in data.

As explained in Algorithm 7-2 (line 3), the small change (a number of moves) is derived from the multiplication of *newvisit* (the GSA new visit distribution from Equation 2-13) and the size of variables (*N*). For the synthetic data, the small change is to be factored with 45 (45 variables). These experiments manipulate a constant (*k*) to produce either a larger length of the variables or lesser. The range of the constant *k* is shown in Figure 7-8.

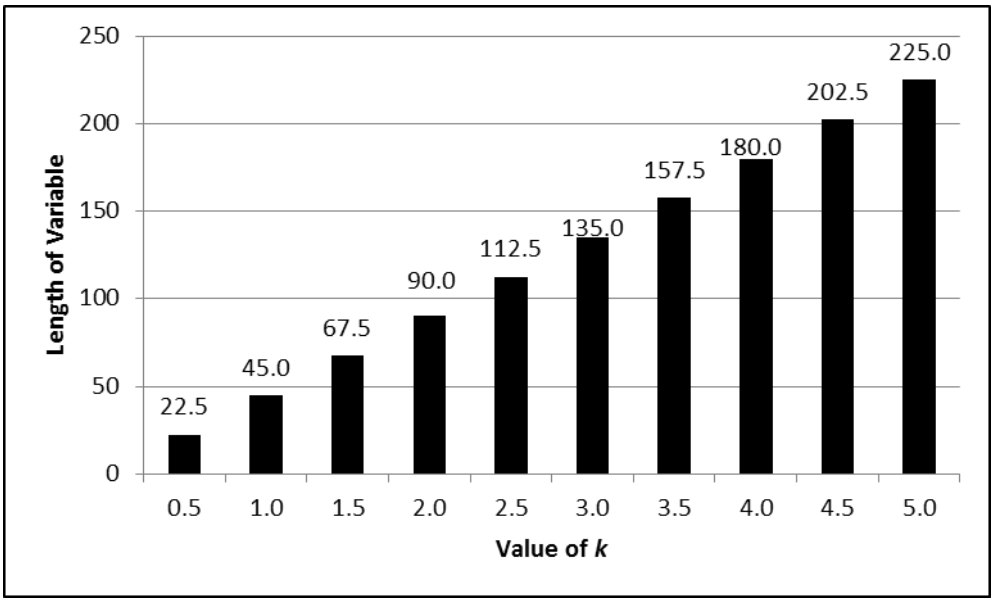


Figure 7-8: The Length of Variables for Small Change in *kN* Experiments

Each of the *kN* value, as shown in Figure 7-8 above, are experimented using the NBU classifier with 10FCV. As learned from the previous experiment results (10FCV strategy) that the average convergence point is less than 10,000 iterations, therefore these experiments are run for 10,000 iterations with 10 repeats. However, in order to get the same efficiency of the RGFGSA method, the same parameter values from the previous experiment (q_a and q_v) are used.

7.4.1 Results

Each of the *kN* experiments final results is obtained and tabulated for comparison as shown in Table 7-15. The full complete 10 experiments results are presented in **Appendix 7-B**.

k	Average Accuracy	Average WK
0.5	97.749	0.851
1.0	97.724	0.872
1.5	97.686	0.817
2.0	98.063	0.896
2.5	97.389	0.806
3.0	97.758	0.864
3.5	97.629	0.807
4.0	97.828	0.898
4.5	97.771	0.841
5.0	97.793	0.857

Table 7-15 : The *kN* Experiments Results for Synthetic Data

Figure 7-9 shows the results ranked by average accuracy which the highest average accuracy was obtained from experiment with k=2.0.

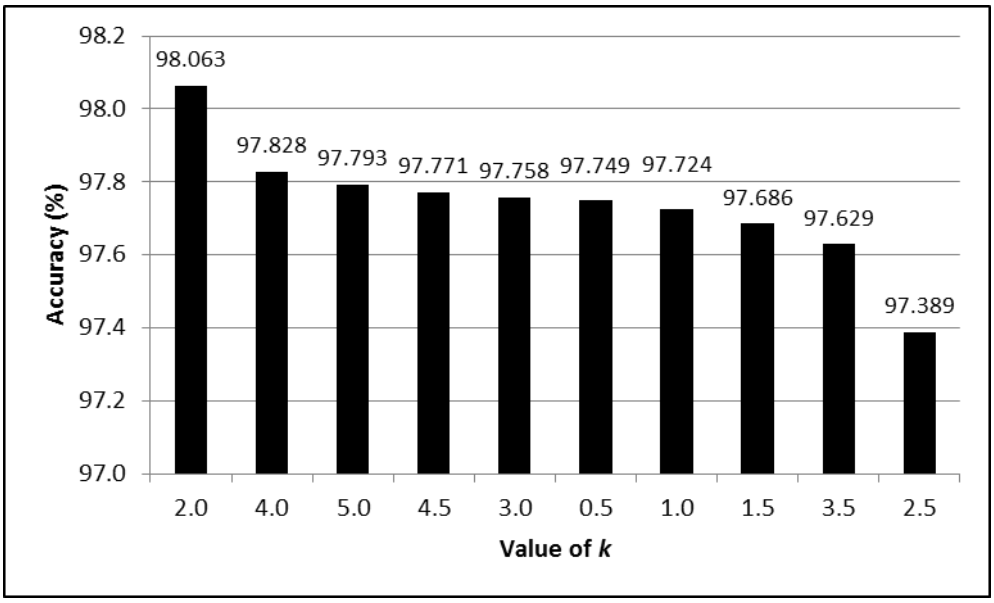


Figure 7-9: The *kN* Experiment Results Ranked by Accuracy for Synthetic Data

Meanwhile, the experiment with k=4.0 has the highest WK (0.898) as shown in Figure 7-10. It clearly can be seen from the figure that the top 3 in the rank are from the experiments with k positive integer value. As far as *reverse engineering* is concerned, this result concludes that the positive integer values for variable k in RGFGSA algorithms could find a better solution for data clustering arrangements (measure in WK).

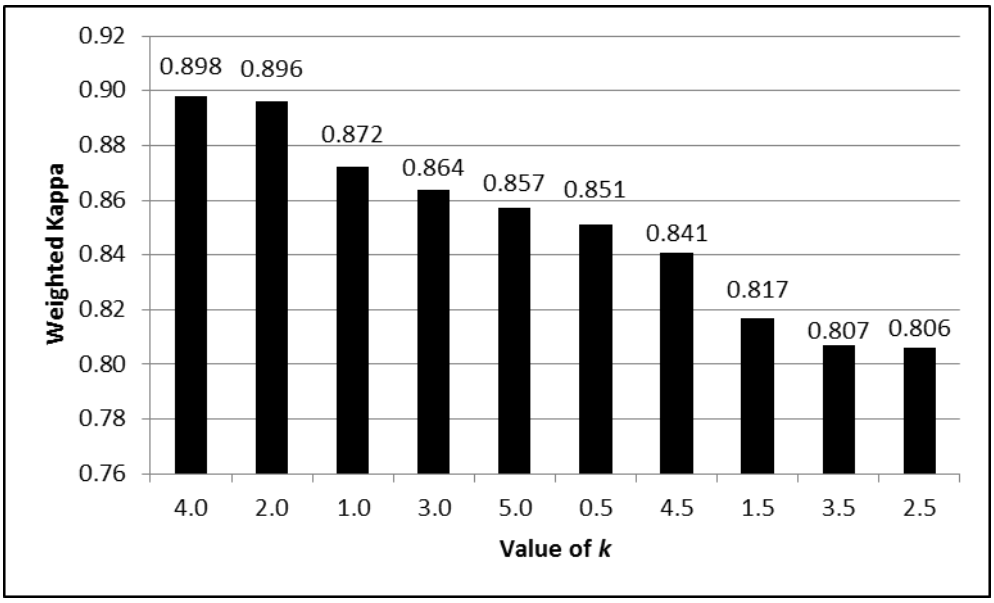


Figure 7-10 : The *kN* Experiments Results Ranked by WK for Synthetic Data

Figure 7-11 and 7-12 exhibit the convergence graph for the experiments with k=2.0 and k=1.0 respectively. From the graphs, it can be seen that the search with RFGFSA experienced too much noise in the search. This might be due to the inappropriate q_a and q_v (obtained with 100,000 iterations) values for 10,000 iterations.

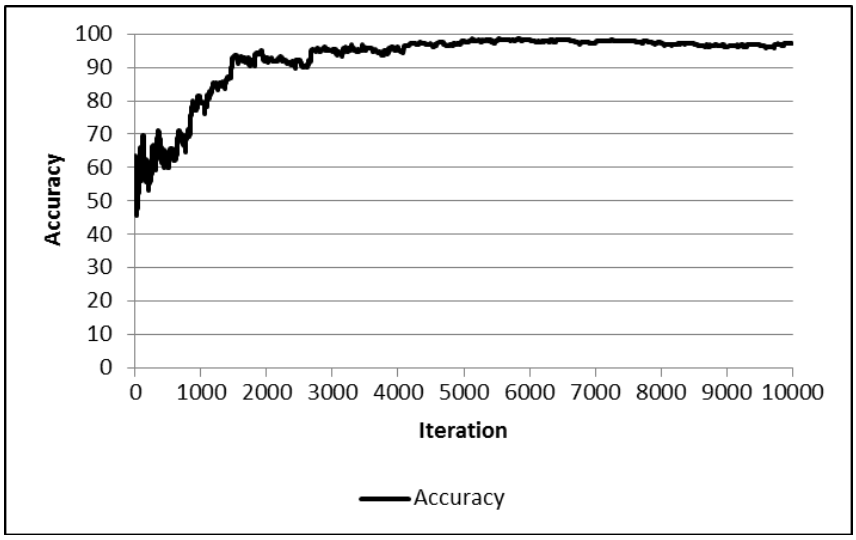


Figure 7-11 : Convergence Graph for *kN* Experiment (*k*=2.0)

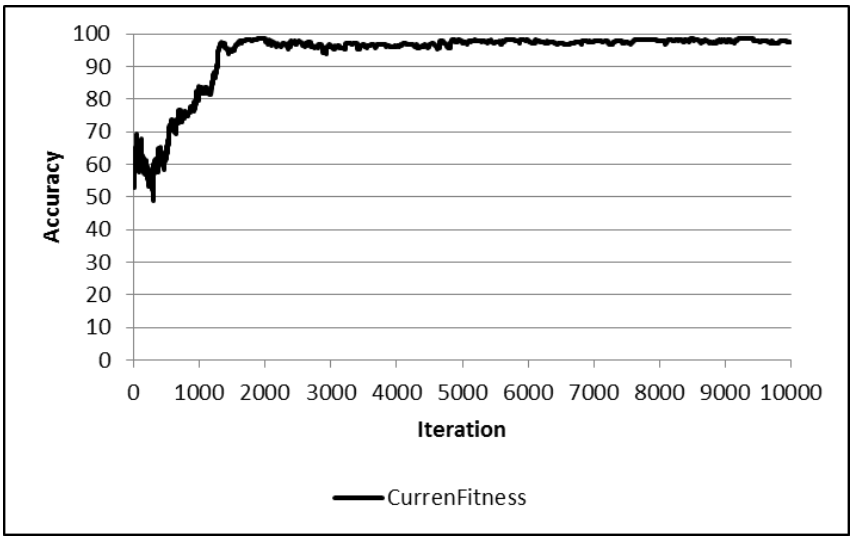


Figure 7-12 : Convergence Graph for *kN* Experiment (*k*=1.0)

7.4.2 Discussion

The results from the *kN* experiments are indecisive to infer meaningful information. There are inconsistent results between the accuracy and WK by ranking. However WKs have

shown better results with the positive interger *k* value (*k*:1,2,3,4,5). The accuracies within the *kN* experiments have not much different between each other. The different between the highest accuracy and the lowest accuracy in these experiments is 0.674% (Accuracy for *k*=2.0 and Accuracy for *k*=2.5). There are also very little different between WKs within the *k* values experiments. The different between the highest and the lowest WK is 0.092 (WK for *k*=4.0 and WK for *k*=2.5).

7.5 Subsets of Visual Field Data

Another supplementary experiment was carried out on visual field data based on three subsets of visual field data. The goal of this experiment is to see the significance of visual field records in predicting visual field loss. Furthermore, with these datasets, the performance of the SMC can be further measured on the real data based on predictive accuracy and WK.

As discussed earlier in Chapter 3, the data exploration has demonstrated the three subsets of visual field data: early record datasets, middle record dataset, and latest record dataset. These datasets consist of different number of records due to some patients have limited visual field test records (explained in Chapter 3, section 3.1.1). Early record dataset has 1,580 records, middle record dataset has 1,083 records and latest record dataset has 1,284 records.

The three optimisation methods (SA, GSA, and RGFGSA) are tested on the three datasets using the three modelling strategies (10FCV, 2FCV, and NoFCV). The classifier used in the experiments is NBU and is run for 100,000 iterations (25 repeats). The experimental setups are briefly shown in Table 7-16.

Method	Dataset Record	Repeat	Iteration	Classifier	Modelling Strategy
SA	Early	25	100,000	NBU	10FCV 2FCV NoFCV
	Middle				
	Latest				
GSA	Early				
	Middle				
	Latest				
RGFGSA	Early				
	Middle				
	Latest				

Table 7-16 : Experimental Setup for the 3 Dataset Categories

7.5.1 Results

Predictive Accuracy

Table 7-17 presents the predictive accuracy results for the early dataset's experiments. All three methods predict better with 10FCV modelling strategy experiment with the average accuracy 87.649%, 87.729% and 87.698%, for SA, GSA and RGFGSA respectively.

Modelling Strategy/Method		Predictive Accuracy (%) Early Dataset		
		10FCV	**2FCV**	**NoFCV**
SA	Min	86.995	86.108	86.646
	Max	88.431	87.225	87.468
	Average	87.649	86.651	87.066
GSA	Min	87.229	86.272	86.519
	Max	88.568	87.070	87.405
	Average	87.729	86.786	87.019
RGFGSA	Min	87.105	86.354	86.329
	Max	88.435	87.253	87.405
	Average	87.698	86.745	86.954

Table 7-17 : Predictive Accuracy for Early Dataset Experiments

Likewise in the middle dataset's experiments (Table 7-18), the best average accuracy is recorded by the experiments with the 10FCV modelling strategy: SA 88.938%, GSA 88.752%, and RGFGSA 89.034%. Interesting result was found in the middle dataset's experiment (10FCV) with highest accuracy 90.225% by the SA method. Meanwhile, the RGFGSA average accuracy is the highest amongst others with 10FCV and 2FCV in the middle dataset's experiment.

Modelling Strategy/Method		Predictive Accuracy (%) Middle Dataset		
		10FCV	**2FCV**	**NoFCV**
SA	Min	87.976	86.714	87.073
	Max	90.225	87.291	87.627
	Average	88.938	86.979	87.309
GSA	Min	86.848	86.843	86.704
	Max	89.788	87.323	87.535
	Average	88.752	87.042	87.206
RGFGSA	Min	88.578	86.631	86.888
	Max	89.606	87.309	87.535
	Average	89.034	87.004	87.228

Table 7-18 : Predictive Accuracy for Middle Dataset Experiments

The latest dataset's experiments results shown in Table 7-19 reveal that 10FCV modelling strategy is better to improve prediction accuracy than the other methods. The average accuracy results with 10FCV are 87.659%, 87.672%, and 87.709% for SA, GSA and RGFGSA respectively.

Modelling Strategy/Method		Predictive Accuracy (%) Latest Dataset		
		10FCV	2FCV	NoFCV
SA	Min	86.997	85.728	86.371
	Max	88.439	86.386	87.305
	Average	87.659	86.029	86.816
GSA	Min	86.838	85.798	86.293
	Max	88.248	86.636	87.350
	Average	87.672	86.178	86.800
RGFGSA	Min	87.080	85.623	86.137
	Max	88.652	86.620	87.150
	Average	87.709	86.093	86.701

Table 7-19 : Predictive Accuracy for Latest Dataset Experiments

Based on the observation of the predictive accuracy results in all datasets experiments, the middle dataset is better (with higher accuracy) to predict glaucoma deterioration using visual field data than the early dataset and latest dataset experiments. Obviously from the results, the 10FCV modelling strategy is the best approach to model the data with high average accuracy. RGFGSA has higher average accuracy with 10FCV modelling strategy in the middle and latest dataset than the other methods.

Weighted Kappa Statistic

WKs of the resultant clusters against the 6NFB are inconclusive. Table 7-20 shows the WK results (in average) for the experiments by dataset categories, methods and modelling strategies. The highest WK in these experiments is 0.0063 (very poor agreement with the 6NFB).

Dataset/Modelling Strategy/Method		10FCV	2FCV	NoFCV
Early	SA	-0.0073	-0.0038	-0.0119
	GSA	0.0012	-0.0100	-0.0106
	RGFGSA	-0.0016	-0.0069	-0.0109
Middle	SA	-0.0016	-0.0009	-0.0027
	GSA	0.0022	-0.0064	0.0039
	RGFGSA	0.0040	0.0052	0.0020
Latest	SA	0.0048	-0.0048	-0.0098
	GSA	0.0063	-0.0050	-0.0047
	RGFGSA	0.0024	-0.0059	-0.0019

Table 7-20 : Weighted Kappa of Resultant Clusters against 6NFB

As the WK results are found very poor in these experiments, consistency of the resultant clusters is measured by computing WK among the resultant clusters. From Table 7-21, the highest average WK among the resultant clusters is obtained from the early dataset's experiment using the GSA method with NoFCV (0.354 – fair agreement). With the 2FCV modelling strategy, the best average WK (0.196) among the resultant clusters is from the early dataset using the RGFGSA method. With this evidence of cross comparing the resultant clusters, it shows that the algorithm produces more consistent clustering arrangement in the early dataset than the others. Meanwhile WKs are found very poor in experiments with the 10FCV modelling (all datasets and methods).

Method	Dataset	Weighted Kappa		
		10FCV	2FCV	NoCV
SA	Early Dataset	0.003	0.100	0.231
	Middle Dataset	0.005	0.041	0.098
	Latest Dataset	0.007	0.044	0.100
GSA	Early Dataset	0.009	0.162	0.354
	Middle Dataset	0.002	0.050	0.097
	Latest Dataset	0.004	0.072	0.107
RGFGSA	Early Dataset	0.004	0.196	0.264
	Middle Dataset	0.011	0.040	0.105
	Latest Dataset	0.002	0.092	0.097

Table 7-21 : Average Weighted Kappa among the Resultant Clusters

Convergence Point

Table 7-22 to 7-24 show the average convergence point (fitness tolerance) for the experiments. The GSA method was found more efficient than the other methods in the early and latest dataset with the 10FCV modelling strategy (1,350.20 and 1,065.12 respectively). Meanwhile in the middle dataset's experiment, RGFGSA outperformed GSA with the best average convergence point 1,680.80 (with the 10FCV). As for the SA method, the average convergence points in all datasets experiments are too high. The best average convergence point recorded by SA is 18,751.60 in the middle dataset's experiment with 10FCV. On the other hand, RGFGSA was found consistent to be the most efficient method in all dataset experiments with 2FCV modelling strategy.

Method/Modelling Strategy	Early Dataset	
	10FCV	2FCV
SA	21,313.24	37,182.44
GSA	1,350.20	38,182.08
RGFGSA	2,385.56	35,215.76

Table 7-22 : Average Fitness Tolerance Convergence Point for Early Dataset

Method/Modelling Strategy	Middle Dataset	
	10FCV	2FCV
SA	18,751.60	33,986.96
GSA	7,577.32	25,439.12
RGFGSA	1,680.80	24,510.28

Table 7-23 : Average Fitness Tolerance Convergence Point for Middle Dataset

Method/Modelling Strategy	Latest Dataset	
	10FCV	2FCV
SA	19,029.72	36,996.12
GSA	1,065.12	35,351.68
RGFGSA	1,623.92	32,788.96

Table 7-24 : Average Fitness Tolerance Convergence Point for Latest Dataset

7.5.2 Discussion

The experiments' results from the three subsets of visual field data reveal that SMC with the RGFGSA method produces more quality results in terms of predictive accuracy, WK and convergence point than the other methods. Prediction of AGIS using the SMC clustering technique is better in the middle dataset than the early and latest datasets. It also can be concluded that 10FCV is the best modelling strategy as this strategy highly improved the accuracy in all datasets. In average accuracy, RGFGSA was found the best method to predict AGIS in the middle dataset with the 10FCV and 2FCV modelling strategy and in the latest dataset with 10FCV. However the best result (highest accuracy) was from the SA method in the middle dataset experiment with the 10FCV modelling strategy.

Despite the fact that 10FCV was the best (high accuracy) strategy for modelling the three datasets of visual field, the WK results have shown that the resultant clusters are more consistent with NoFCV modelling strategy (all datasets and methods) than the 10FCV and 2FCV with high WK. The experiments with the 10FCV and 2FCV modelling strategies have very poor WK results among the resultant clusters even though the predictive accuracy values are very close (among the three modelling strategies and datasets). This indicates that the 10FCV and 2FCV modelling strategies have produced different clustering arrangements in the 25 set experiments compared to NoFCV. With respect to the phenomenon of close predictive accuracy values and different clustering arrangements in the experiments, this might be due to cross validation. To overcome and comprehend this situation, further experiments can be done using other strategies of cross validation and other methods of sampling the visual field datasets.

The convergence point analyses have found that there is no significant pattern can be seen to identify which method is the most efficient in the three dataset experiments with the

10FCV modelling strategy. However, GSA and RGFGSA have shown promising results in the 10FCV in the three datasets experiments compared to SA. Additionally, in these experiments, it is noticed that RGFGSA consistently has become the fastest method to converge in the search with the 2FCV modelling strategy. Nevertheless, the average convergence points in the 2FCV experiments are rather higher than 10FCV (best: 24,510.28). The SA method has been clearly shown to be the most inefficient method in these experiments.

7.6 Summary

Chapter 7 demonstrated the use of RGFs representation in the GSA algorithm's search procedure in SMC to garner quality solutions. The aim of this work was to improve the efficiency of SMC by removing the inherence of degeneracy in the combinatorial problems. To prove the method's efficiency, the novel algorithm, which is namely, RGFGSA, was examined on the real data (visual field) and synthetic data. The three measurements for the method performance: predictive accuracy, WK and convergence point, have shown positive results. RGFGSA appears to be an effective approach to reduce degeneracy problem in SMC as the method was found suffer less from the effects of a noisy fitness function compared to GSA. This was shown by the comparison of convergence graph between RGFGSA and GSA (Figure 7-5 and Figure 7-6).

Although the average accuracy results (of 25 experiments) of RGFGSA were found inconsistent to be the highest in the real data, the method has appeared to be consistent with highest average accuracy results in the synthetic data compared to the other methods. The RGFGSA method also has recorded the best result (highest accuracy) in the visual field data with the 10FCV modelling strategy. In comparison with the benchmark of this study, RGFGSA' predictive accuracies were still found higher than *K*-means and the 6NFB.

The WK results for the visual field data were very poor in all methods and modelling strategies. However, exceptional results obtained in the synthetic experiments have shown that the RGFGSA method has found the perfect clusters (the known clusters) in all 25 experiments (with 2FCV). This is another evidence that the method is capable to remove the inherence of degeneracy in SMC. As for method's efficiency, RGFGSA has appeared to be the most efficient algorithm in terms of runtime (visual field data), and had faster convergence with lower average convergence point than SA and GSA in the synthetic data (10FCV and 2FCV).

Two supplementary experiments were carried out to further explore the RGFGSA algorithm in this work. First experiment was testing the algorithm with a range of small change size, namely, *kN* experiments. Second experiment was testing the method on the three datasets of visual field data: early, middle and latest record dataset. The *kN* experiments have discovered that the WK value is higher in a positive integer value of *k*. The discovery of this rule in *k*N experiment (k = 2.0, 4.0 and 5.0) for RGFGSA's small change is worth to be applied in real data. However, the significant difference between the WK values in this experiment is extremely tiny.

The experiments' results from the three datasets of visual field have shown consistent clustering arrangements. In the experiments, 10FCV appears to be the best modelling strategy to improve predictive accuracy. The middle dataset of visual field was found the best dataset to predict the AGIS score using the three methods and modelling strategies as the average accuracies are higher than other datasets. There is no conclusion can be made from the WK results of the three dataset of visual field experiments in 10FCV. However, within these experiments, the resultant clusters are more consistent in the NoFCV modelling strategy than others with fair agreement WK (among the resultant clusters). Also, early dataset was found to be much more consistent in resultant clusters. An interesting finding on the convergence point results is that, the search has (with GSA

and RGFGSA) converged earlier in the three visual field dataset categories than the sampled dataset.

To summarise, with the RGFGSA method, SMC has shown a significant improvement on accuracy in the real data compared to the benchmark methods as well as the synthetic data. It can also be concluded that the RGFGSA method reduces degeneracy in SMC, more consistent (WK) clusters with the 6NFB than the other method with subset dataset of visual field, and works very efficient and effective with 2FCV. Based on this remark, it is viable to employ the RGFGSA algorithm in SMC.

Chapter 8

Conclusions

8.1 Overview

Chapter 8 summarises the entire works have been carried out in this research. The aim of this chapter is to highlight the research outcomes and findings. Building on the findings, research limitations and proposals for future work are presented. Chapter 8 comprises of five sections. Section 8.2 summarises all chapters of the thesis. This section also recaps the findings of each work presented in this research. Section 8.3 presents the contributions of the research, which will be of benefit to clinical practitioners (8.3.1) and the machine learning community (8.3.2). Limitations of the study are discussed in section 8.4. Finally, based on the limitation and findings, this thesis proposes future work in section 8.5.

8.2 Summary of the Dissertation

This dissertation started with introducing the subject, defining the problem statement and establishing the research questions in the domain problem: Glaucoma. Managing glaucoma in patients using visual field data is the common practice. Visual field data are multi-dimensional data in nature with 52 variables (visual field locations). The problem with analysing the data is that modelling the visual field data with a large number of variables will cause overfitting and inefficient modelling (Clements and Hendry, 2002). This problem arises from the need to find visual field locations that are highly related. The current practice of analysing the data uses the six nerve fiber bundles (6NFB) to predict the deterioration of glaucoma. However, as high predictive accuracy of visual field data is concerned, obtaining other groups of visual field locations with higher prediction accuracy than the 6NFB is a contribution. Since there is no gold standard

method in analysing the data, a novel clustering technique namely, Simultaneous Modelling and Clustering (SMC) was proposed in this research to model the data with high predictive accuracy. Within this research, a few of the SMC's algorithms have been developed using notable machine learning techniques in order to obtain the best clusters of visual field locations as well as improving prediction accuracy. As stated in Chapter 1, the research questions for this study are defined as follows:-

Research Question (RQ)	**Question**
RQ1	What is the baseline accuracy for predicting visual field loss across all patients when the visual field points are aggregated to the six nerve fiber bundles (6NFB)?
RQ2	Can visual field data being clustered using model-based clustering to improve the baseline accuracy as in question RQ1? a. Can model based clustering techniques improve on the base line accuracy (c.f. RQ1)? b. Does visual field points arrangement from the model-based agrees with the clinical evidence of glaucoma deterioration?

Research Question (RQ)	Question
RQ3	Does the choice of heuristic search techniques in the model-based clustering technique (SMC) effectively improve prediction accuracy of visual field loss?
RQ4	Can the SMC approach improves visual field loss prediction using a subset of the patients' records?

Table 8-1 : Research Questions

Chapter 2 has provided the literature review on the research areas. The domain problem, the concept of machine learning, the techniques and methods have been discussed therein. The methods used in this research were identified beforehand based on the review from the literature.

Data exploration has been carried out on the visual field data in Chapter 3 and discovered that the data suffer from the class imbalance problem. Due to imbalanced data, a simple analysis has been carried out to validate the evaluation metric (classification accuracy) of the imbalanced data. From the overall classification accuracy the analysis results indicated that the classifier performance was good even on imbalanced data. Chapter 3 has also highlighted this research has applied re-classified data class for efficient classification as adopted in the literature.

Chapter 4 has introduced a novel clustering technique, SMC which has been applied to the visual field data. This chapter comprised of a set of initial experiments and then some further experiments using SMC. The initial experiments were carried out to obtain the

baseline predictive accuracy for the data using the 6NFB. The initial experiments have discovered the range of accuracy for predicting visual field loss (AGIS): between 76.29% and 86.11%. The accuracy range was obtained by performing classification experiments on all record of visual field data (totalling 12,159) using the J48, NB and MNB classifiers. This range of accuracy is used as the baseline in this research. Revelation of the accuracy baseline has answered RQ1. Meanwhile, the SMC experiments carried out in Chapter 4 tested SMC on the real data (visual field) to improve predictive accuracy of visual field loss from the baseline accuracy. These experiments have found that the model-based clustering technique (SMC) improves prediction accuracy in the real data. The SMC results have proven that the technique can improve visual field loss prediction from the baseline accuracy. The best accuracy (88.49%) was recorded using the NB classifier. With this record, SMC improved prediction accuracy by 2.66% from the upper limit (86.11%) of baseline accuracy. Moreover, there was a significant improvement in accuracy (by 9.23%) using the MNB. This finding answered in part RQ2.

Another conclusion can be made from the work in Chapter 4 is the agreement level of the resultant clusters of SMC with the 6NFB. The resultant clusters of the SMC experiments were retrieved for WK analysis to measure the agreement level between the resultant clusters and the 6NFB. The WK results were found to be very poor (near to 0.0) and these results revealed that the SMC clusters of visual field locations have very different clustering arrangement with the 6NFB even though predictive accuracy is improved. This finding answered RQ2-a that the resultant clusters do not resemble the 6NFB.

Additionally, the resultant clusters were analysed with visualisation of the SMC's clusters on the visual field grid. The visualisation was the mapping of the 52 visual field locations (the SMC's clusters) on the visual field grid. From the visualisation of the clusters arrangement discovered by SMC, the result was in accordance with clinical evidence. The clinical evidence indicates that glaucoma first starts at the periphery of the visual field

grid and near to the blind spot. The larger clusters of the SMC's result appeared on the periphery of the visual field grid. Also, some of the visual field locations from are located near to the blind spot and the periphery of the grid. This finding on the visualisation of clustering arrangement corresponds to RQ2-b.

Chapter 5 has validated SMC on a synthetic dataset. As the WK results in Chapter 4 were poor, therefore, devising a synthetic data to validate SMC was necessary. Even though the clustering arrangement in the synthetic data is very simplistic: the first 15 variables are in cluster one, the next 15 variables are in cluster two, and so on, however, the data have been randomised in the SMC process. This requires efficient heuristic effort from the optimization methods used in this study to search the known clustering arrangement. Additionally, the 6NFB clustering arrangement also has fairly simple arrangement: visual field point 1 to 5 are grouped in the same cluster. Furthermore, the data has the same nature and properties such as data distribution, number of variables and records.

With regards to modelling strategies, the work in Chapter 5 has been introduced with the three modelling strategies in the experiments. The experiments from the synthetic data have shown promising results in classification accuracy and WK. The resultant clusters of SMC have high agreement (measured using WK) with the known clusters. With the high WK results in the synthetic data, SMC was proven workable clustering technique. The results also supported the hypothesis underpinning this work that the higher prediction accuracy, the better quality clustering arrangement. This is because SMC is designed in such a way to only retain solutions (clustering arrangements) with high classification/prediction accuracy (the fitness value) during the heuristics search. Moreover, the perturbation (small change), which is carried out to produce a new solution during the search, is based on a clustering arrangement with high classification/prediction accuracy. Therefore, the assumption in this study is made that higher accuracy represents a good solution.

Moreover, the 2FCV modelling strategy has shown higher results in accuracy and WK than other strategies. Thus, the three modelling strategies have been used in the subsequent works in Chapter 6 and 7.

Additionally, clustering using *K*-means technique was carried out on both datasets (visual field and synthetic) in Chapter 5. These experiments were conducted to see the performance of the common clustering technique as a comparison with the SMC technique. The *K*-means results were found lower than SMC in prediction accuracy of visual field data. Meanwhile for the synthetic data, the *K*-means results were poor with average only 60.91% accurate (with 10FCV) as compared to SMC. As a conclusion, *K*-means has appeared to be less effective to improve accuracy in both datasets.

Chapter 6 introduced the advanced optimisation method in SMC in order to improve the work carried out in Chapter 4 and 5. The GSA method used in this chapter is the advanced annealing version of SA. Chapter 6 revealed a finding on the performance of SMC with the application of GSA. The GSA method has appeared to be more efficient (faster convergence point) in searching the best clusters in the real data than the existing methods with 10FCV. The GSA optimisation method has been proven to better improve predictive accuracy using the clusters obtained by the method in the real and synthetic data than SA. The method is also outperformed the benchmark accuracies (*K*-means and the 6NFB) for this study. SMC with the GSA optimisation method has shown huge improvements in WK results in the synthetic data's experiments. The application of GSA in SMC has appeared to be more efficient (faster obtaining solution in the search) in the experiments with lower average convergence points. Other key findings of Chapter 6 are modelling strategy and number of iterations in the experiments. The 2FCV modelling strategy was the best modelling strategy in the work of Chapter 6. Also, as proven in the results, a larger number of iterations in a search advantage the GSA method to obtain a quality solution (high accuracy in the real data and high WK in the synthetic data). This finding

contributes towards answering RQ3. An additional finding in Chapter 6 was the visualisation of the visual field locations clustered by SMC. Larger clusters are located in the peripheral and the blind spot on the visual field grid. This result is consistent with Chapter 4, which corresponds to the clinical evidence and provides solution to RQ2.

Chapter 7 comprised of an additional set of SMC experiments. The main focus of the chapter was on the degeneracy issue where the GSA method was employed with Restricted Growth Functions in the search. The RGFGSA method has shown effective to remove the degeneracy problem in the search. RGFGSA was found to be the most efficient method than SA and GSA with the lowest average convergence point within the 25 set experiments using 10FCV and 2FCV in synthetic data. An anticipated side effect of removing degeneracy was that the fitness function became less noisy. The method also has been proven very effective with the 2FCV modelling strategy in obtaining the known clusters in the synthetic data in all 25 set experiments. These results contributed to WK = 1.0. Again, these findings provide the answer for RQ3, where heuristics search methods are the key contribution in improving the performance of SMC. In terms of predictive accuracy, SMC with the RGFGSA method is also found predicts better than the benchmark accuracies (*K*-means and the 6NFB).

The other supporting experiments carried out in Chapter 7 were the *kN* and the three subsets of visual field data experiments. These experiments were carried out to identify some useful properties regarding the SMC method as well as the data (visual field). The results in the *kN* experiments have shown that the classification accuracy on the synthetic with the RGFGSA method was higher with a positive integer value of *k*. The experiment results with *k* values 2, 4 and 5 were ranked the top three in classification accuracy. This finding indirectly contributes an answer to RQ3.

Meanwhile the three subsets of visual field data experiments have found that middle record dataset was the best dataset to be used for predicting visual field loss with high predictive accuracy records. The highest accuracy was from SA with 90.23% accurate and in an average accuracy, RGFGSA was the best with 89.03% accurate. These results were found within the 10FCV experiments. In terms of WK result, NoFCV was found to be the best modelling strategy to produce consistent resultant clusters within the 25 set experiments. The best WK (average) among resultant clusters was 0.354 from the early record dataset with the GSA method. This indicates that the experiments on the early record dataset of visual field resulted in more consistent resultant clusters. Findings from the three visual field data categories clearly have answered RQ4.

8.3 Contributions

The contribution of this research is two-fold. Firstly, this research contributes knowledge to medical practitioners in managing glaucoma using visual field data. Secondly, it offers practical recommendations in the field of machine learning.

8.3.1 Clinical Contributions

Discovery of visual field location arrangements with high predication accuracy (AGIS score) that different than the 6NFB could draw an interest from medical perspective in diagnosis glaucoma. From the clustering arrangement of visual field locations suggested by the SMC method, the pattern of the deteriorated visual field locations can be studied by medical practitioners to predict glaucoma in patients. The main objective of this research was to search for compound variables which are highly related subsets in predicting visual field loss instead of the 52 variables and the 6NFB. The compound variables of the visual field data were searched using the SMC technique in order to have

high predictive accuracy. The end result of SMC was in the form of clusters of the visual field locations. The clusters were studied for the significance of the visual field locations to see the association among them, which indicates visual deterioration. The resultant clusters found by SMC have the same agreement with the clinical evidence as shown in the visualisation on the visual field grid analysis. Larger clusters are located in the periphery and near to the blind spot of the visual field grid. Thus, these patterns of visual deterioration mapped on the visual field grid could equip medical practitioners with a prediction of glaucoma progression. With the information of visual deterioration, an advance appropriate treatment could be delivered in patients in order to prevent the disease from further damage.

Although the resultant clusters have poor WK results with the 6NFB, the visualisation of resultant clusters supports the resultant clusters that the size of clusters found by SMC can be considered by medical practitioners in predicting glaucoma loss using visual field data. Chapter 4 and Chapter 7 presented the range of cluster sizes can be used on visual field data to predict visual field loss. In Chapter 4, it was discovered that the minimum clusters' size is 4 and the maximum is 17 in all data experiments. The sampled experiments however discovered 5 and 19 for the minimum and maximum clusters' size respectively.

The best accuracy predicting AGIS using visual field data was with 12 clusters. This was doubled size from the 6NFB. However in Chapter 7, with the advanced optimisation method used in SMC, smaller clusters sizes were suggested. The advanced methods suggested clusters sizes between 2 to 11 and 3 to 14 by RGFGSA and GSA respectively.

Another contribution of this study to medical practitioners is prediction of AGIS using the three datasets of visual field data. This work was presented in Chapter 7 where experiments were conducted on the subsets of visual field data: early records, middle

records and latest records. Middle records dataset was found the best dataset to have high prediction accuracy of visual field loss in patient with 90% accurate.

8.3.2 Contribution in Machine Learning

The outcomes of this research have contributed some practical knowledge of machine learning in terms of the methods and strategies used in this research. This study proposed a novel clustering and classification technique, namely, SMC, which was applied to modelling visual field and synthetic data. The first notable contribution in this field is that, SMC has been proven to be effective and efficient with data clustering and classification in this study, which presented in visual field data and synthetic data experiments using certain modelling strategies and optimisation techniques. With SMC, both datasets have appeared to have improvement in prediction accuracy. The results demonstrated in Chapter 4, which is the introduction of SMC on the visual field data, have shown improvements from the baseline accuracy. The advanced application of the optimisation methods in SMC presented in chapter 6 and 7 are also improved predictive accuracy from the benchmark accuracies (*K*-means and the 6NFB). Additionally in Chapter 5, SMC has improved classification accuracy of the synthetic data with high WK. These results indicated a substance contribution as the application of SMC on other datasets and problems has produced promising results. Moreover, the novel technique proposed in this study is a generic modelling and clustering technique, which can be applied to other domain problems and modelling techniques such as forecasting and regression.

The optimisation methods (SA, GSA, and RGFGSA) that applied in SMC presented in experiments in Chapter 4, 5, 6 and 7 appear to have promising results. It can also be concluded that the use of the annealing methods in SMC were effective and efficient. In

Chapter 4, the SA optimisation method was found the best method amongst other methods (RMHC and RRHC) to improve predictive accuracy of visual field data.

GSA is another annealing optimisation method, which was proven to produce good results in continuous problems in the literature (dos R Correia *et al.*, 2005; Tsallis and Stariolo, 1996), has been employed in SMC. Inspired from positive results in the literature, the GSA has been adopted in the discrete optimisation problem (SMC) as presented in Chapter 6. The method has improved predictive accuracy in the real data and synthetic data. These results confirmed that GSA is compatible with discrete problems.

The advanced method of annealing namely RGFGSA was designed to demonstrate that SMC could be more efficient than the preceding optimisation methods in terms of convergence point with RGFGSA. The RGFGSA method has obtained its name as the RGF procedure was adopted in the GSA method with the goal to remove degeneracy in the search. This novel algorithm was found effective to remove degeneracy in the search as reported in Chapter 7. With RGFGSA, the search has converged relatively faster than the other methods. As a result of removing degeneracy, RGFGSA is able to reduce noisy fitness in the search.

Furthermore, this research has explored the three modelling strategies: 10FCV, 2FCV and NoFCV in the experiments. The introduction of these modelling strategies in this study with the objective to observe the cause and effect of number of partitions in cross validation towards the performance of SMC which is measure in classification accuracy. This study has introduced 2FCV with 10 repeats in the experiments and this modelling strategy has appeared effective in the experiments. This was proven by the results from Chapter 6 and Chapter 7, where the known clusters of the synthetic data were found by the modelling strategy with GSA and RGFGSA. Also, with 2FCV, the RGFGSA method

has shown a superior performance search compared to other methods in the synthetic data.

Finally, this work has provided empirical evidence that a larger number of iterations have allowed an extensive search. There were studies in the literature on algorithms that effectiveness and efficacy performance of the algorithms are affected by the number of iterations (Mirjalili and Gandomi, 2017; Askarzadeh, 2016). Based on this knowledge from the literature, number of iterations was increased to 50,000 and 100,000 in the experiments carried out in Chapter 5 and 6 respectively. Annealing methods with a larger number of iterations has appeared to be very effective in the search as values of the variables used in the methods such as artificial temperature and visiting distribution were affected by the number of iterations.

8.4 Limitations

Limitations of research are constraints and conditions, which cannot be controlled by researchers, influence research findings and its interpretations. This research has experienced some limitations. However, the scope and objectives of the research are within acceptable tolerance limits. Assumptions within the limitations are helpful to support the achievement of the research goals and objectives. There are six limitations have been identified in this research.

8.4.1 Data Quality

This research used secondary data which were retrieved with permission from Moorefield Eye Hospital London. In secondary data, researchers do not have control on data

collection from subjects. Thus, imbalance class issue, which was discussed in Chapter 2, was found in this research and the limitation needs to be dealt with. The summary statistics has shown that the visual field data have high number of patients in two visual field test records. There is only one patient with 42 visual field test records. Also, there are more than 40% from the data were classified with no severity of visual field loss (AGIS = 0). Having class imbalance in data would potentially mislead the modelling.

The implication of imbalance data and class affects the performance of classification as discussed in Chapter 2. Therefore, an analysis was carried out on the imbalance data to validate classifier performance on the data. This imbalance data and class issues were looked into another perspective by analysing the data using *G-mean* to measure the classifiers performance on the data. The *G-mean* analysis results indicate that the classifier performance on the data appears to be good. Therefore, this research takes the data as good and over-sampling and under-sampling were not adopted in this research.

8.4.2 The Subsets Data

The subsets of visual field data were presented in Chapter 2. This datasets were also presented in Chapter 6 where the datasets were experimented for the significant of visual field subsets to predict the AGIS score. As the visual field data are imbalanced, this issue also affect the subsets data: early records, middle records, and latest records. These subsets data were sampled with a few assumptions based on the number of visual field records in patients. The assumptions include patients with at least two records are eligible for the early records dataset. For the middle records dataset, patients with at least four visual field records are eligible. The latest dataset are sampled from all patients except the patients with two visual field records. One visual field record was sampled from each eligible patient for the datasets.

The subsets data however can contain bias. Patients' records which were sampled in the latest dataset can be in the middle dataset. This is because patients in the latest dataset with three records can also be considered in the middle dataset. These patients may have several visual field tests which are not included in this research data or the patients may need to have more visual field test to be eligible in the latest dataset.

This research also took the assumption that all patients' records are tested during the onset of glaucoma. However, the patient with only little visual field records can be in the severe stage of glaucoma and these patients were sampled in the early dataset.

8.4.3 Classifiers

The classifiers used in this research were J48, NB, MNB and NBU. The selection of these classifiers for the experiments was made due to the efficiency of the classifiers in terms of runtime as discussed in Chapter 3. However there are many other good classifiers with higher accuracy than these classifiers can be used. This limitation of research was considered in the experiments due to the aim of this research was to validate and verify the SMC technique. Besides finding the best clusters for modelling visual field data, the main focus of this research is to validate the SMC technique instead of finding the best classifiers. For this reason, other classifiers with high accuracy such as Neural Network and Multi-Layer Perceptron may not be appropriate in the experiment due to being highly computationally expensive.

8.4.4 Optimisation Technique

Optimisation methods are the main component of SMC. Simple optimisers were employed in SMC such as hill climbing methods. This was justified by easy to implement and the effectiveness of the methods from the literature. The annealing optimisers were used in SMC owing to the methods are able to avoid local optima in a search. Even though the SA method has recorded the highest accuracy with 90%, the fine-tuned algorithm (RGFGSA), which was presented in Chapter 7, has not reach 90% accurate of prediction. Nevertheless, RGFGSA successfully solved degeneracy in the search and converged much quicker (on the synthetic data with 10FCV).

8.4.5 Parameters for the Methods

The annealing optimisation methods used in this research are well known as highly assumption-based methods. Variables and parameters such as artificial temperature, cooling rate, visiting distribution, acceptance index and visiting index require user calibration and assumptions. The assumptions to these values may affect the performance of the methods used in the experiment. Furthermore, a nature of problem also affects the method performance.

Simulated Annealing

Setting the value for initial temperature in SA plays a critical role in acceptance probability. The initial value should not be high enough that the algorithm simply conducts a random search causing excessive computation time. Therefore, a careful procedure for setting the value is essential. This research runs a separate simulation to

estimate the initial temperature in SA, and the approach of estimating the value was proven effective by Swift *et al.* (Swift *et al.*, 2004). However, there are many other ways in obtaining initial temperature value for SA. Kirkpatrick et al (Kirkpatrick *et al.*, 1983) suggested initial temperature value by obtaining the maximum different of two neighbouring solutions. One study (Jung, Jayakrishnan and Park, 2016) has a separate procedure to compute the initial temperature by obtaining the different between the best and the average solution values.

Generalised Simulated Annealing

The GSA method has more parameters which require user setup especially the acceptance index (q_a) and visiting index (q_v). The right acceptance index of GSA was searched by running a simulation using a range of acceptance index values. The best result for the simulation obtained the value for the acceptance index. However, exploration of the value in this research has limited the value to 3 decimal places.

As for the visiting index, the value in this research was derived using mathematic tool with a few of assumptions. The visiting index value is a corresponding value to the acceptance index. Similarly to the initial temperature for GSA, assumptions values for different fitness and acceptance probability were used in order to obtain the initial temperature value.

8.4.6 Medical Endorsement

The results presented in this thesis were analysed in view of algorithms effectiveness and efficiency utilised in the proposed method (SMC) on visual field data. Besides inventing

the best (compatible) algorithms using this clustering and classification approach, the study has set the objective to analyse and describe the significant clustering arrangements of visual field that highly predict AGIS. However, the discussions on the results were limited to non-medical opinions and only based on the literature which this study made. Medical endorsement on results can be meaningful to the study for future work.

8.5 Future Work

This future work section looks at improvement to the experiments can be carried out to further explore SMC. The improvement areas that can be made are methods and generalisation of SMC on other datasets.

8.5.1 Methods

Optimisation Technique

There are plenty of methods available in literature with promising results can be used in solving optimisation problems. According to Fister Jr *et al.* (Fister Jr *et al.*, 2013), the choice of a method is depending on a nature of problems. In this thesis, the main focus is on single population (Local Search) techniques. Population based methods such as Genetic Algorithms have operators which can result in offspring (in the case of GAs) that are invalid: result in a representation that is not a valid clustering arrangement. It is noted that there are complex representations and operators that can overcome this problem to some extent, however the focus on this thesis is the SMC process as a proof of concept, not the adaptation of the method towards population-based methods. This will be explored as future work.

As far as higher accuracy is concerned, other optimisation techniques as part of SMC would be worth considering. A nature-inspired optimiser namely, the Firefly Algorithm was proven outperformed GA and Particle Swarm Optimisationin a study (X. Yang, 2009) could also be used to test SMC.

Classifiers

With regard to high classification in modelling, SMC is worth to employ other classifiers with high accuracy classification in the experiments although some of the other available classifiers have a high computational overhead. In preliminary experiments of this study, classifiers such as logistic regression, SMO (the Support Vector Machine classifier in WEKA) and neural network have been tested. Even though the performance of the classifiers in terms of runtime was fairly poor compared to others, the accuracy results from these methods were above the average baseline accuracy.

Parameters Values

As discussed in the limitation section, assumptions were made in the GSA method in order to obtain the values for the parameters. Also, the range of acceptance index was limited to a certain range. Therefore, further exploration on these parameters values is worth in order to improve the performance of SMC with the optimisation methods.

Small Change

Chapter 6 has presented the *kN* experiments which is a size of the small change was manipulated to have smaller or larger size of the small change. These experiments were

conducted on the synthetic data using the RGFGSA method. The results of the experiments have provided some good information that can be used to improve classification accuracy. Extending the same experiment to visual field data in the future work is worth. Furthermore, applying the same approach (the *kN*) with GSA and other methods is another area worth exploring.

Noisy Fitness

Noisy fitness is a common issue in optimisation problems especially in real datasets. SMC has been proven less susceptible to noisy fitness in the synthetic data with RGFGSA. However, the noisy fitness problem in the visual field data is still an area that warrants further investigation. Therefore there is a need for further investigation on the real data noisy fitness. From the literature, there are a number of studies on glaucoma (Henson *et al.*, 1997; Liu *et al.*, 1994) to fix this problem. Another way that potentially solves noisy fitness is elementary landscape analysis. Elementary landscape analysis, which is a mathematical formalism of the search space of a combinatorial optimisation problem, studies neighbourhood move operator to define adjacency between points in the search space (Whitley *et al.*, 2014). An investigation of noisy fitness on the visual field using this technique is an open research opportunity.

Seeded Search

Seeded search (Gravina, Liapis and Yannakakis, 2017) in optimisation problems is commonly used as a guide to initialise the point of search in the search space. Seeded search can be used if one is provided with information to help the search becomes more efficient than just having a random point of search. Since the range of cluster sizes found by SMC is between 4 and 19, therefore this information can be used in a future work

experiment to seed the search. Also, the 6NFB can be seeded as the initial clusters in the search owing to the high baseline accuracy with the 6NFB (76% - 86%). With this strategy, SMC is expected to perform more efficient in terms of convergence than using random clusters as the initial clusters.

Algorithm Runtime

The three modelling strategies used in this research have shown various performance results in terms of runtime. This is due to the fact that a higher number of cross validation folds and number of repeats in modelling the data requires more computational efforts. Therefore, it is worth extending this research by conducting a formal analysis on the algorithms' runtime.

8.5.2 Dataset

Application of SMC on other datasets with high accuracy of classification motivates further study on the method. This includes visual field datasets and other domain problems.

Visual field data

The experiments of the three subsets of the visual field data, which are early records, middle records and latest records, have shown promising results (predictive accuracy) on middle records dataset. However, the experiment results of the early records dataset are found to have consistent clustering arrangement results (with NoFCV) compared to the other datasets even though the predictive accuracy values are a little difference among

them (early, middle and latest datasets). For this reason, further experiments can be done by using other modelling strategies and other methods of sampling the three subsets of the data in order to further understand this situation.

As discussed in the limitation section, the datasets were derived with a few assumptions. The assumption of obtaining the datasets may lead to bias due to some patients with few visual field records may be wrongly classified in the datasets. Therefore, standard criteria represent each dataset (early, middle and latest) should be established in order to avoid bias in classification. Refining the datasets would probably give a better exploration of the data using the SMC technique.

Other future work opportunities in the domain of visual field data analysis are to extend the experiments to using data from both eyes. Visual field data with a combination of both eyes would result in a 104 variable dataset. Investigation on both eyes would be investigated the relationship and associated deterioration of vision between the right and left eyes, which is known to be related.

Other datasets

Besides the visual field data experiments, the good results presented in the synthetic data experiments inspire SMC to be generalised to other domain problems. SMC is a model-based clustering technique for discrete problems, which the technique works for dimensionality reduction in data. Datasets of other domain problems with high dimensional data such as gene expression (Heimberg *et al.*, 2016) and deep sequencing (Veneziano *et al.*, 2016) can be used to test SMC in identifying relevant features in data for classification.

Appendix 2A Decision Tree

Entropy of target variable:-

$$Entropy(C) = -\sum_{i=1}^{c} P_i \log_2 p_i$$

Entropy of two attributes:-

$$E(C, X) = \sum_{c \in X} P(c)E(c)$$

Information gain:-

$$Gain(T, X) = E(T) - E(T, X)$$

Example from the following dataset:-

No	Variable-a	Variable-b	Variable-c	Target Variable
1	a1	b2	c2	Yes
2	a2	b2	c1	No
3	a1	b1	c2	No
4	a1	b2	c3	Yes
5	a2	b1	c2	No
6	a1	b2	c2	Yes
7	a3	b2	c1	Yes
8	a1	b1	c2	Yes
9	a3	b1	c3	Yes
10	a3	b2	c2	No
11	a1	b1	c1	Yes
12	a2	b2	c3	No
13	a1	b1	c3	Yes
14	a2	b2	c1	Yes

Target Variable	Frequency	Probability
Yes	9	0.64
No	5	0.36
Total	14	1.00

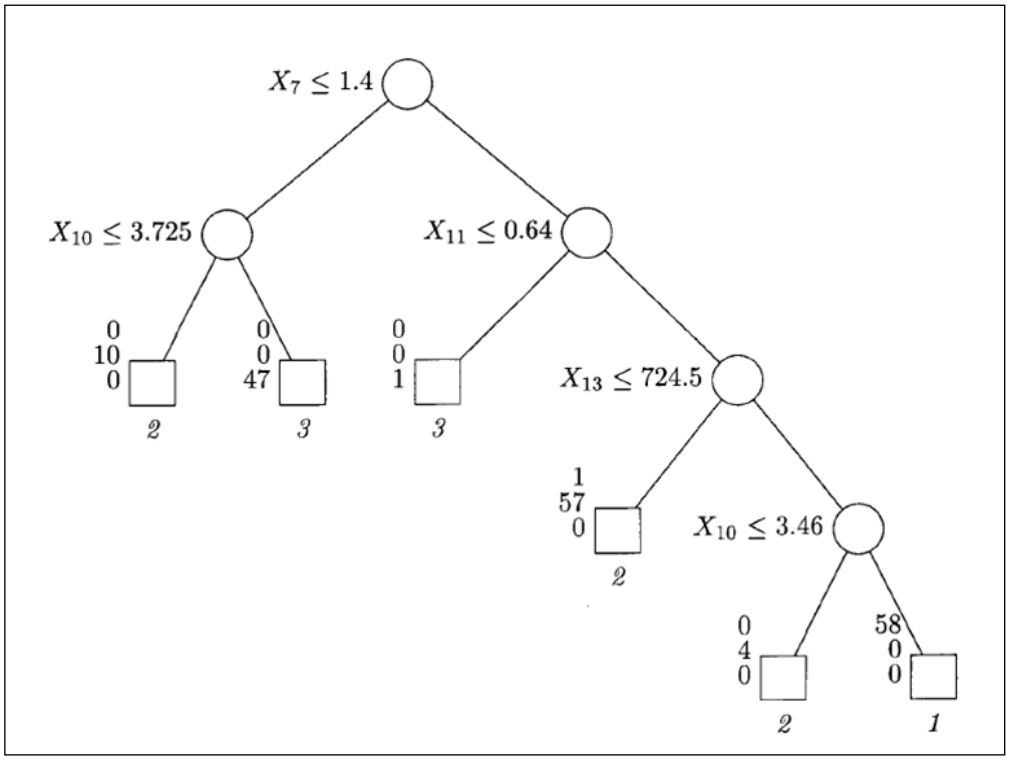
Variable-a	Yes	No	Total	Prob-Class-a	Probability-Yes	Probability-No	Entropy
a1	6	1	7	0.50	0.86	0.14	0.59
a2	1	3	4	0.29	0.25	0.75	0.81
a3	2	1	3	0.21	0.67	0.33	0.92
Total	9	5	14	1.00	0.64	0.36	0.94
Information Gain = -0.059							

Variable-b	Yes	No	Total	Prob-Class-b	Probability-Yes	Probability-No	Entropy
b1	4	3	7	0.50	0.57	0.43	0.99
b2	5	2	7	0.50	0.71	0.29	0.86
Total	9	5	14	1.00	0.64	0.36	0.94
Information Gain = 0.016							

Variable-c	Yes	No	Total	Prob-Class-c	Probability-Yes	Probability-No	Entropy
c1	3	1	4	0.29	0.75	0.25	0.81
c2	3	3	6	0.43	0.5	0.5	1.00
c3	3	1	4	0.29	0.75	0.25	0.81
Total	9	5	14	1.00	0.64	0.36	0.94
Information Gain = **0.048**							

The decision node is chosen based on the highest information gain from the variables. For this example, the variable *c* is the decision note.

Then, entropy is compute again for every split branch. The branch with ‘0’ entropy will be a leaf node. Meanwhile, branches with entropy with more than ‘0’ need further splitting. Example of a decision tree diagram (Shih, 1999) is as follow:-



Appendix 2B Naïve Bayes Classifier

Bayes Theorem is defined as:-

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Example from the following dataset:-

No	Variable-A	Variable-B	Variable-C	Target Variable (*c*)
1	a1	b2	c2	Yes
2	a2	b2	c1	No
3	a1	b1	c2	No
4	a1	b2	c3	Yes
5	a2	b1	c2	No
6	a1	b2	c2	Yes
7	a3	b2	c1	Yes
8	a1	b1	c2	Yes
9	a3	b1	c3	Yes
10	a3	b2	c2	No
11	a1	b1	c1	Yes
12	a2	b2	c3	No
13	a1	b1	c3	Yes
14	a2	b2	c1	Yes

Target Variable	Frequency	Probability
Yes	9	0.64
No	5	0.36
Total	14	1.00

If one record is given such as a1, b2 and c3, the followings are computed.

The record (*a*1, *b*2 and *c*3) probability to be classified as 'Yes' is computed using Bayes Theorem as below:-

Variable-A	**Yes**	**No**	**Total**	**P(A|Yes)**	**P(A)**
a1	6	1	7	0.67	0.50
a2	1	3	4	0.11	0.29
a3	2	1	3	0.22	0.21
Total	9	5	14	1.00	1.00

Variable-B	**Yes**	**No**	**Total**	**P(B|Yes)**	**P(B)**
b1	4	3	7	0.44	0.50
b2	5	2	7	0.56	0.50
Total	9	5	14	1.00	1.00

Variable-c	**Yes**	**No**	**Total**	**P(C|Yes)**	**P(C)**
c1	3	1	4	0.33	0.29
c2	3	3	6	0.33	0.43
c3	3	1	4	0.33	0.29
Total	9	5	14	1.00	1.00

Variable	**Class Value**	**P(X_i)**	**P(A_i|Yes)**	**P(Yes|A_i)× P(Yes)**
A	a1	0.50	0.67	0.43
B	b2	0.50	0.56	0.36
C	c3	0.29	0.33	0.21

***P*(*c* = 'Yes')** is 0.64

$$P(c = Yes \mid a1, b2, c3) = \frac{(P(a1 \mid Yes) \times P(Yes)) \times (P(b2 \mid Yes) \times P(Yes)) \times (P(c3 \mid Yes) \times P(Yes))}{P(a1) \times P(b2) \times P(c3)}$$

$$P(c = 'Yes' \mid a1, b2, c3) = \frac{0.43 \times 0.36 \times 0.21}{0.50 \times 0.50 \times 0.29}$$

$$P(c = 'Yes' \mid a1, b2, c3) = \frac{0.032508}{0.0725}$$

$$P(c = 'Yes' \mid a1, b2, c3) = 0.448386 \approx 0.45$$

The record (*a*1, *b*2 and *c*3) probability to be classified as 'No is computed using Bayes Theorem as below:-

Variable-A	**Yes**	**No**	**Total**	**P(A|No)**	**P(A)**
a1	6	1	7	0.20	0.50
a2	1	3	4	0.60	0.29
a3	2	1	3	0.20	0.21
Total	9	5	14	1.00	1.00

Variable-B	Yes	No	Total	P(B|No)	P(B)
b1	4	3	7	0.60	0.50
b2	5	2	7	0.40	0.50
Total	9	5	14	1.00	1.00

Variable-c	Yes	No	Total	P(C|No)	P(C)
c1	3	1	4	0.20	0.29
c2	3	3	6	0.60	0.43
c3	3	1	4	0.20	0.29
Total	9	5	14	1.00	1.00

***P*(*c* = 'No')** is 0.36

Variable	Class Value	$P(X_i)$	$P(A_i\|No)$	$P(No\|A_i)\times P(No)$
A	a1	0.50	0.20	0.072
B	b2	0.50	0.40	0.144
C	c3	0.29	0.20	0.072

$$P(c='No'|\ a1,b2,c3) = \frac{(P(a1|\ No)\times P(No))\times(P(b2|\ No)\times P(No))\times(P(c3|\ No)\times P(No))}{P(a1)\times P(b2)\times P(c3)}$$

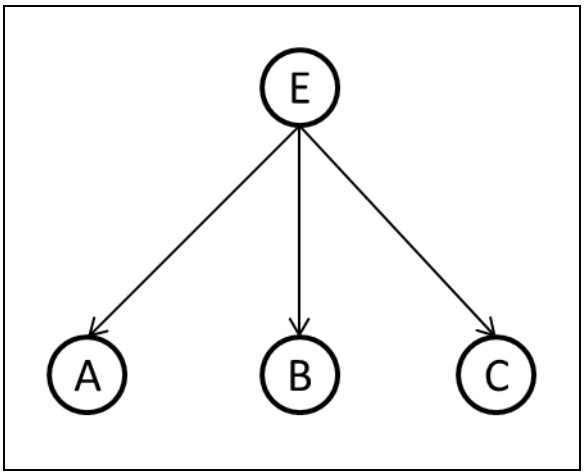
$$P(c='No'|\ a1,b2,c3) = \frac{(0.072)\times(0.144)\times(0.072)}{0.50\times 0.50\times 0.29}$$

$$P(c='No'|\ a1,b2,c3) = \frac{0.000746}{0.0725}$$

$$P(c = 'No' | a1, b2, c3) = 0.010296 \approx 0.01$$

From the computation of probability ***Yes*** and ***No*** using Naïve Bayes Theorem for the record $a1$, $b2$ and $c3$, the probability value of $P(c = 'Yes' | a1, b2, c3)$ (0.45) appear to be greater than $P(c = 'No' | a1, b2, c3)$ (0.01). Therefore, the new record is classified as **Yes**.

Figure bellows shows the graphical model of the dependencies between the variables within this example.



Appendix 2C Multinomial Naïve Bayes

The example of Multinomial Naïve Bayes is shown in the following example:-

Training Documents		
Document	**Words**	**Class (*x*)**
D1	China UK Japan	East
D2	Malaysia UK Germany	West
D3	Germany UK Ireland China	West
D4	Japan Germany Malaysia	East
D5	Germany UK Malaysia	West
D6	Malaysia China Japan	East
D7	Germany UK	West

Vocabulary
China
UK
Japan
Germany
Malaysia
Ireland

No. of vocabulary is 6

The followings are test documents to be classified:-

Test Document (TD)	
TD	**Words**
TD1	Germany Japan Malaysia
TD2	Germany Germany UK

From the training documents, the prior probability of the classes is computed:-

$P(x = East) = 3/7 \approx 0.42857$ $P(x = West) = 3/7 \approx 0.57143$

The probability of the class for Test document 1 (TD1) is computed as follows:-

In TD1, words ‘Germany’, ‘Japan and ‘Malaysia’ are presented. Thus, the probability of these words for the class ‘East’ and ‘West’ is computed. Firstly, all documents which are classified as ‘East’ are retrieved as the table below:-

Training Documents		
Document	**Words**	**Class**
D1	China UK Japan	East
D4	Japan Germany Malaysia	East
D6	Malaysia China Japan	East

The equations for Multinomial Naïve Bayes as explained in Chapter 2 are as follows:-

$$P(c \mid t_i) = \frac{P(c)P(t_i \mid c)}{P(t_i)} \text{ (Equation 2c-1)}$$

$$P(t_i \mid c) = \alpha \prod_n P(w_n \mid c)^{f_{ni}} \quad \text{(Equation 2c-2)}$$

$$\widehat{P}(w_n \mid c) = \frac{1 + F_{nc}}{N + \sum_{x=1}^{N} F_{xc}} \quad \text{(Equation 2c-3)}$$

$$P(t_i) = \sum_{k=1}^{|C|} P(k)P(t_i \mid k) \quad \text{(Equation 2c-4)}$$

Equation 2c-3 is used to compute $P(t_i \mid c) \approx P(Germany, Japan, Malaysia \mid East)$

P(Germany|East)	1+1/(9+6)	0.1333
P(Japan|East)	3+1/(9+6)	0.2666
P(Malaysia|East)	2+1/(9+6)	0.2000

Then, documents with class ‘West’ are retrieved.

Training Documents		
Document	**Words**	**Class**
D2	Malaysia UK Germany	West
D3	Germany UK Ireland China	West
D5	Germany UK Malaysia	West
D7	Germany UK	West

P(Germany|West)	4+1/(12+6)	0.2777
P(Japan|West)	0+1/(12+6)	0.0555
P(Malaysia|West)	2+3/(12+6)	0.2777

$P(TD1)$ is computed as in Equation 2c-4 as follows:-

$$P(TD1) = [P(East) \times P(Germany \mid East) \times P(Japan \mid East) \times P(Malaysia \mid East)] + [P(West) \times P(Germany \mid West) \times P(Japan \mid West) \times P(Malaysia \mid West)]$$

$$P(TD1) = 0.00305 + 0.00245 \approx 0.0055$$

$$P(East \mid TD1) = P(East) \times P(Germany \mid East) \times P(Japan \mid East) \times P(Malaysia \mid East)$$

$$P(East \mid TD1) = 0.42857 \times 0.1333 \times 0.2666 \times 0.2000 \approx 0.00305$$

$$P(East \mid TD1) = \frac{0.00305}{0.0055} \approx 0.5545$$

$$P(West \mid TD1) = P(West) \times P(Germany \mid West) \times P(Japan \mid West) \times P(Malaysia \mid West)$$

$$P(West \mid TD1) = 0.57143 \times 0.2777 \times 0.0555 \times 0.2777 \approx 0.00245$$

$$P(East \mid TD1) = \frac{0.00245}{0.0055} \approx 0.4456$$

As the probability of the document for class ‘East’ (0.5545) is higher than class ‘West’ (0.4456), TD1 is classified as class ‘East’.

The other test documents are computed and the following is the result of classification:-

Document	Words	Probability East	Probability West	Class
TD1	Germany Japan Malaysia	0.5545	0.4456	East
TD2	Germany Germany UK	0.0766	0.9234	West

Appendix 2D *K*-Means Algorithm

Algorithm 2D : KMeans (X, k)	
Input:	X = Dataset $k > 1$ $C_i : i \to \{1,\ldots,k\}$
1	Assign the objects (rows) randomly to C_i ensuring no cluster is empty
2	Calculate the centres of each cluster
3	Allocate each object to the new centres by minimising the sum of squares error, SS(X)
4	Repeat steps 2 and 3 until the terminating condition is met
Output:	Set of clusters

$k > 1$

$c_{i=1..k}$

c_i = centroid of cluster *i*.

$D = \sum_{i}^{k} \sum_{j}^{n} | x_{ij} - c_i |$, distance between object and cluster's centroid.

Appendix 4A Correlation of AGIS and Number of VF Test

Correlation between the variables is computed as follows:-

Firstly the number of tests for every patient is counted. There are 1,580 patients in the data with 2 and 42 for minimum and maximum number of tests respectively. An example is as follow:-

PID	Count
4	6
5	6
6	3
7	9
8	6
9	6
10	4
11	3

PID denotes patient ID, which is an internal non-medical record based identifier.

Next, number of patients for each number of tests is counted. Majority patients have 2 number of tests (296 patients). There are 38 groups for number of tests. The table below shows the sample of the data for number of tests and number of patients.

Number of Tests	No. of Patients
2	296
3	201
4	162
5	128
6	91
7	74
10	52

Then, the AGIS score (unclassified score) from each patient is sampled by taking the lowest AGIS score from the patients records.

PID	AGIS	Number of Tests
4	17	6
5	8	6
6	13	3
7	11	9
8	3	6
9	12	6
10	0	4
11	14	3
57	5	2

Subsequently, the AGIS score for patients within the same group of number of test is summed. Meaning that, all patients with 2 number of tests are accumulated for the AGIS score.

Number of Tests	No. of Patient	Sum of AGIS
2	296	1279
3	201	753
4	162	618
5	128	351
6	91	290
7	74	169
10	52	96
8	47	121
13	45	96

Based on the observation of numbers of tests, there are 14 groups of number of tests that have less than 9 patients in the groups. Therefore, the correlation for the number of tests and the AGIS score is computed from the sampled for the group of number of test with number of patients more than 9 as follow (the table is sorted in No. of Patient order):-

Number of Tests	No. of Patient	Average Sum of AGIS
2	296	4.3209
3	201	3.7463
4	162	3.8148
5	128	2.7422
6	91	3.1868
7	74	2.2838
10	52	1.8462
8	47	2.5745
13	45	2.1333
12	42	2.0000
15	40	2.6750
9	39	1.9231
11	38	3.1579
16	35	2.0857
14	30	2.5333
21	29	1.1724
17	28	2.0357
20	28	0.5357
18	26	1.3462
19	24	1.0417
22	23	1.3913
23	20	0.1500
24	20	0.7500
27	10	3.1000

The correlation for the two variables is -0.7247. This shows that the higher the number of test, the lower the AGIS score recorded by patients.

Appendix 6A Newton Raphson Method

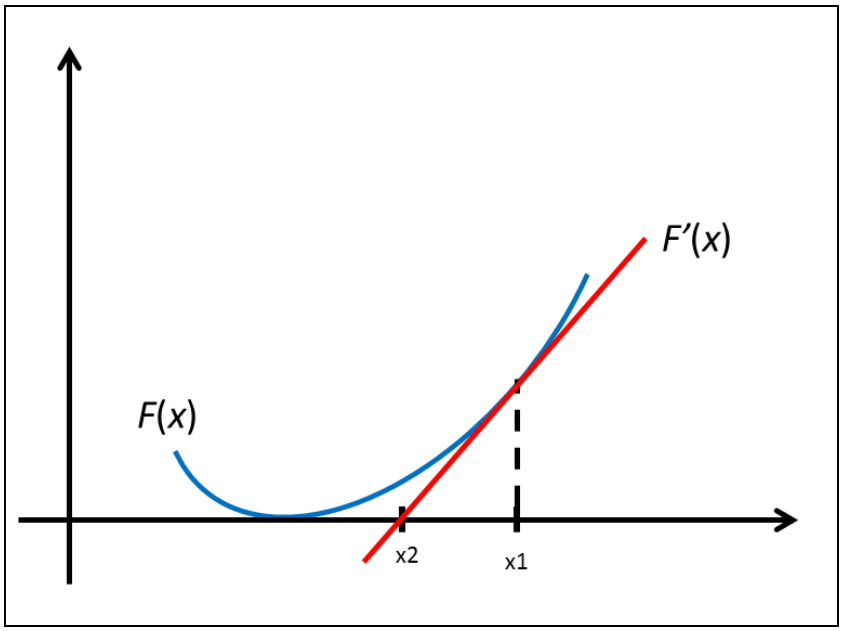
The Newton Raphson equation is as follow:-

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad \text{(Equation 5A-1)}$$

Pseudo code for the Newton Raphson method is as follow:-

Algorithm 5A : Newton Raphson	
Input:	X = 1 Y = f(X)
1	While |Y| > 0.001
2	X = Equation 5A-1
3	Y = f(X)
4	End while
Output:	Get Y and X

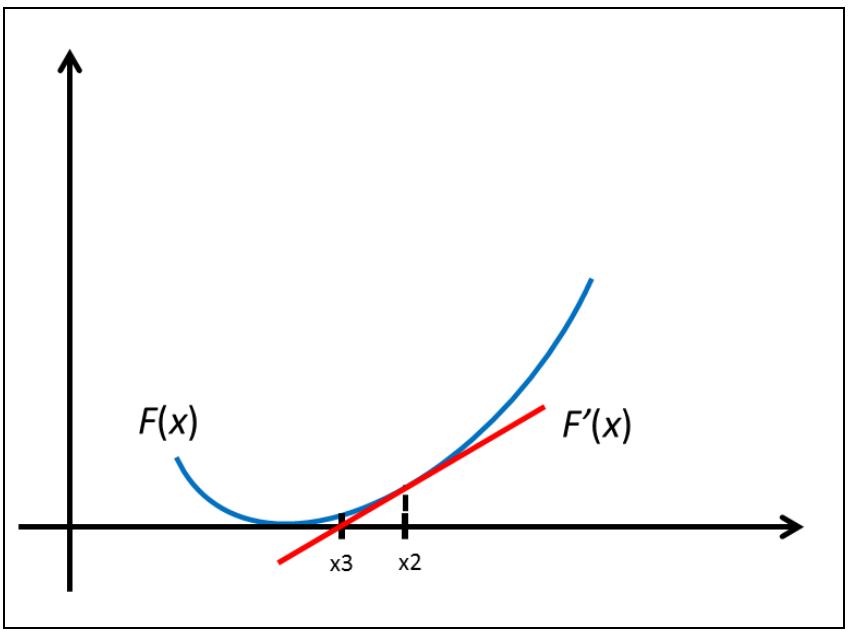
The graphical example of Newton Raphson is as follow:-



Iteration of obtaining an *x* value which equivalent to $f(x_4) \approx 0$ starts best with $x = 1.0$

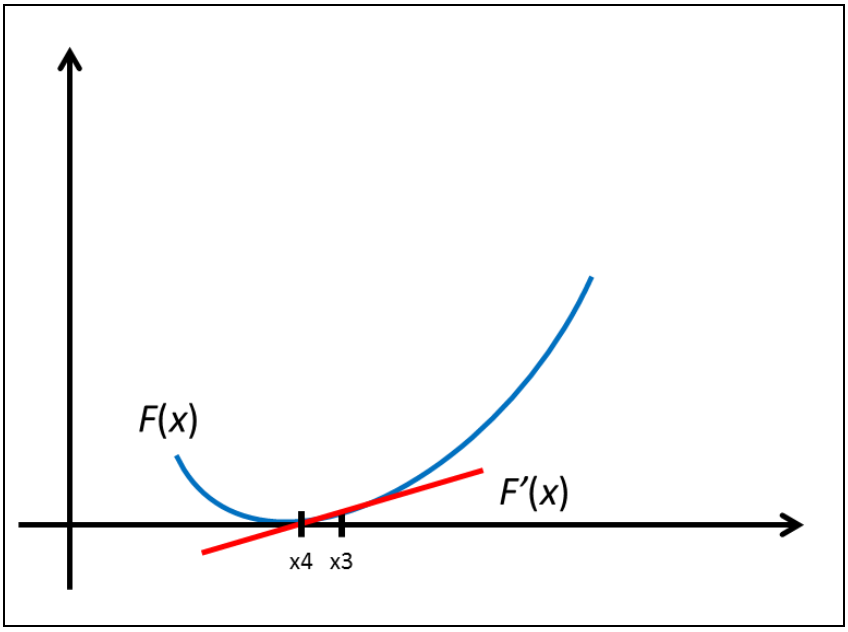
$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

The value of x_2 is obtained from the tangent function (the red line).



$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

The value of x_3 is obtained from the tangent function (the red line).



$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)}$$

Iteration stops when the value of x_4 is equivalent to $f(x_4) \approx 0$

Appendix 6B Simulation Experiment Results

Simulation Experiment for GSA Parameter (Visual Field)

Visual Field Experiment for GSA Parameter Comparison
Re-sampling VF Data Every Iteration
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.01				0.02				0.03			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	87.069	2,328	87.136	-0.0110	87.529	9,068	87.529	1.47E-02	87.633	1,141	87.6329	-0.020
2	86.949	5,400	86.949	-0.0026	87.096	9,593	87.096	0.015	87.491	9,397	87.5063	-0.030
3	87.591	1,468	87.591	0.0395	87.561	3,063	87.561	0.001	87.234	4,060	87.2342	-0.036
4	87.011	1,833	87.011	0.0301	87.955	339	87.955	0.014	87.715	5,126	87.7146	0.006
5	87.540	9,441	87.540	-0.0134	87.610	1,209	87.610	-0.023	87.337	7,836	87.3368	0.040
6	87.228	7,674	87.228	-0.0387	87.528	5,215	87.528	-0.034	87.781	9,346	87.7813	0.034
7	86.890	3,323	86.890	0.0104	87.116	7,552	87.116	-0.002	87.235	3,091	87.2349	0.047
8	87.463	3,061	87.463	-0.0644	86.888	9,875	86.888	-0.017	87.219	7,205	87.2191	0.031
9	87.512	3,459	87.512	-0.0103	86.935	1,520	86.935	-0.007	87.285	7,423	87.2845	-0.039
10	87.246	1,353	87.246	-0.0325	86.959	8,785	86.959	-0.006	87.471	3,457	87.4712	0.026
	87.250	**3,934**	**87.257**	**-0.0093**	**87.318**	**5,622**	**87.318**	**-0.004**	**87.440**	**5,808**	**87.4416**	**0.006**

	0.04				0.05				0.06			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	87.284	4,135	87.284	0.008	87.499	7,436	87.499	-0.008	87.173	4,571	87.1725	0.006
2	87.058	2,110	87.058	0.022	87.758	5,242	87.758	-0.015	87.590	4,766	87.5901	0.015
3	86.843	5,124	86.843	0.051	88.008	4,229	88.008	-0.036	87.416	8,818	87.4163	-0.021
4	87.757	9,393	87.760	0.038	87.244	1,779	87.244	-0.030	87.645	9,481	87.6452	-0.040
5	87.752	9,703	87.752	-0.025	88.088	3,254	88.088	-0.022	87.154	8,580	87.2384	-0.037
6	87.659	692	87.659	0.006	87.196	5,026	87.196	0.035	87.244	5,323	87.2856	-0.041
7	86.883	3,429	86.923	-0.001	87.377	5,359	87.377	-0.015	87.338	3,676	87.3380	-0.019
8	87.639	6,439	87.639	0.018	87.579	4,442	87.579	-0.003	87.265	4,721	87.2652	0.032
9	88.761	4,166	88.761	0.005	86.907	7,907	86.950	0.044	87.085	8,993	87.0848	0.086
10	87.177	2,285	87.177	-0.013	87.576	2,495	87.576	-0.028	87.727	9,887	87.7273	-0.037
	87.481	**4,748**	**87.486**	**0.011**	**87.523**	**4,717**	**87.528**	**-0.008**	**87.364**	**6,882**	**87.3763**	**-0.006**

	0.07				0.08				0.09			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	87.394	5,688	87.394	2.92E-03	87.542	6,855	87.542	-0.007	87.481	2,086	87.4809	0.009
2	87.333	8,877	87.333	-0.041	87.788	1,982	87.788	1.01E-02	86.970	3,265	86.9704	0.018
3	87.227	4,410	87.227	0.037	86.844	2,104	86.844	-0.008	87.151	9,932	87.1515	-0.005
4	87.164	9,882	87.164	0.033	87.434	2,128	87.434	0.006	87.776	7,211	87.7764	0.024
5	87.154	9,211	87.154	-0.004	87.267	7,223	87.267	0.035	87.474	3,049	87.5482	-0.006
6	87.662	863	87.662	0.028	87.650	6,984	87.650	-0.011	87.217	1,832	87.2166	0.046
7	87.267	7,103	87.267	-0.004	87.618	7,980	87.618	-0.017	87.425	6,394	87.5092	0.017
8	87.194	8,071	87.194	-0.030	87.605	4,959	87.605	-0.034	87.729	9,590	87.7287	0.031
9	87.229	8,845	87.229	-0.002	87.665	8,280	87.665	-0.001	87.760	9,214	87.7602	-0.033
10	87.471	8,417	87.511	-0.045	87.204	3,586	87.204	0.012	87.659	1,537	87.6589	-0.018
	87.309	**7,137**	**87.313**	**-0.003**	**87.462**	**5,208**	**87.462**	**-0.002**	**87.464**	**5,411**	**87.4801**	**0.008**

Visual Field Experiment for GSA Parameter Comparison
Re-sampling VF Data Every Iteration
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.1				0.2				0.3			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	87.141	2,528	87.141	-0.012	87.856	3,138	87.856	1.68E-04	87.381	9,885	87.381	-0.036
2	86.944	5,667	86.944	-0.009	86.765	5,872	86.765	-0.033	87.593	4,694	87.593	-3.5E-05
3	87.392	5,726	87.392	-0.005	87.355	9,892	87.355	0.016	87.296	7,321	87.296	-0.009
4	87.230	7,106	87.230	0.004	87.348	6,339	87.348	-0.013	86.605	1,041	86.605	0.054
5	87.665	4,837	87.665	-0.028	87.404	688	87.404	-0.038	87.175	1,624	87.175	-0.016
6	87.663	4,223	87.663	0.009	87.333	1,030	87.333	-0.004	87.105	2,079	87.105	-0.020
7	87.523	8,346	87.539	-0.035	87.230	2,068	87.230	-0.043	87.808	1,651	87.808	-0.009
8	87.546	8,015	87.546	0.036	86.923	6,037	86.923	0.026	87.310	9,373	87.310	-0.024
9	87.745	1,165	87.745	0.036	87.405	5,640	87.405	0.003	87.059	1,814	87.090	0.064
10	87.528	4,721	87.528	-0.017	87.289	6,614	87.289	-0.044	87.329	6,612	87.329	0.037
	87.438	**5,233**	**87.439**	**-0.002**	**87.291**	**4,732**	**87.439**	**-0.002**	**87.266**	**4,609**	**87.439**	**-0.002**

0.4				0.5			
Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
87.332	3,519	87.332	-0.007	83.209	10,001	87.433	-0.020
87.001	7,270	87.099	-0.026	84.591	10,001	87.059	0.021
87.158	5,668	87.244	-0.021	84.350	10,001	87.461	-0.015
87.646	6,757	87.646	-0.052	82.697	10,001	87.285	0.009
87.576	2,604	87.576	0.005	82.776	10,001	87.343	-0.037
87.647	5,994	87.647	-0.059	84.056	10,000	87.217	-0.027
87.567	7,666	87.567	-0.027	82.326	10,001	87.557	0.040
87.290	7,300	87.290	-0.003	83.335	10,001	87.987	-0.003
87.418	4,886	87.418	-0.023	83.486	10,001	87.325	0.016
87.134	5,777	87.278	-0.010	83.640	10,001	87.351	-0.019

Visual Field Experiment for GSA Parameter Comparison
Re-sampling VF Data Every Iteration
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.001				0.002				0.003			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Resu	WK
1	87.545	556	87.545	0.0336	87.381	8,774	87.409	-5.8E-03	87.367	781	87.367	-0.057
2	87.081	9,861	87.138	-0.0402	87.117	9,995	87.173	-0.005	87.315	4,400	87.315	-4.5E-02
3	87.098	8,977	87.098	-0.0012	87.227	9,203	87.227	0.026	88.241	2,355	88.241	0.020
4	87.457	7,652	87.457	-0.0412	87.752	9,639	87.752	0.047	87.770	4,268	87.770	-0.031
5	87.502	7,684	87.502	-0.0073	87.244	5,298	87.244	0.009	87.080	5,897	87.159	0.014
6	87.053	8,229	87.083	0.0492	87.166	4,193	87.347	-0.022	87.324	7,128	87.324	0.000
7	87.431	7,214	87.431	-0.0021	86.960	4,106	87.078	0.011	87.616	8,380	87.616	-0.031
8	87.132	9,397	87.132	0.0228	87.348	8,815	87.348	-0.001	87.336	7,606	87.336	0.014
9	87.256	2,009	87.256	-0.0231	87.432	6,735	87.432	-0.005	87.256	9,350	87.256	-0.010
10	87.676	3,494	87.676	-0.0185	86.837	536	86.837	0.052	88.168	483	88.168	-0.031
	87.323	**6,507**	**87.332**	**-0.0028**	**87.246**	**6,729**	**87.285**	**0.011**	**87.547**	**5,065**	**87.555**	**-0.016**

	0.004				0.005				0.006			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Resu	WK
1	87.297	641	87.297	-0.014	87.563	6,025	87.563	-0.032	87.791	8,989	87.791	0.0458
2	87.686	6,149	87.686	0.001	86.930	9,869	86.937	-0.001	87.282	8,011	87.282	0.0161
3	87.686	1,069	87.686	0.045	87.250	7,778	87.250	-0.009	87.461	5,839	87.482	0.0038
4	87.728	7,794	87.728	0.041	87.128	774	87.128	-0.029	87.222	7,970	87.222	0.0304
5	87.141	8,625	87.141	-0.025	87.210	7,204	87.210	0.016	87.197	4,133	87.197	-0.0121
6	87.409	8,993	87.409	-0.009	87.023	5,147	87.023	-0.002	87.346	6,184	87.399	0.0309
7	87.221	1,022	87.221	-0.004	87.331	4,630	87.331	-0.023	87.241	2,100	87.241	-0.0114
8	86.938	7,515	86.991	-0.014	87.838	9,896	87.838	0.046	87.536	5,376	87.536	-0.0366
9	87.468	7,534	87.529	0.000	87.354	3,471	87.354	-0.016	87.207	3,593	87.207	-0.0004
10	87.338	2,872	87.338	0.081	88.044	7,636	88.044	0.011	87.400	3,324	87.400	-0.0357
	87.391	**5,221**	**87.403**	**0.010**	**87.367**	**6,243**	**87.368**	**-0.004**	**87.368**	**5,552**	**87.376**	**0.0031**

	0.007				0.008				0.009			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Resu	WK
1	87.909	9,304	87.909	2.66E-02	86.959	9,431	86.959	-0.012	87.236	5,426	87.236	0.011
2	87.445	9,207	87.529	0.028	88.214	3,789	88.214	-9E-03	87.461	7,860	87.519	0.022
3	87.448	5,417	87.448	0.000	86.770	357	86.770	0.044	87.091	1,502	87.091	0.105
4	87.191	5,552	87.191	0.012	87.332	4,923	87.332	0.008	86.960	8,460	86.984	0.022
5	87.310	3,868	87.310	0.005	88.134	572	88.134	-0.017	87.401	6,346	87.401	0.071
6	87.688	7,538	87.688	-0.025	87.348	8,141	87.348	-0.042	87.054	2,630	87.054	-0.013
7	87.203	6,937	87.203	0.049	87.923	9,767	87.923	-0.008	87.597	3,440	87.597	-0.012
8	87.216	7,594	87.216	-0.006	87.193	8,848	87.224	0.061	87.903	4,331	87.903	-0.020
9	87.163	9,408	87.173	0.047	87.500	7,199	87.500	0.047	87.465	7,124	87.465	-0.013
10	87.184	7,479	87.184	-0.025	88.394	9,975	88.394	0.008	87.584	7,513	87.584	0.011
	87.376	**7,230**	**87.385**	**0.011**	**87.577**	**6,300**	**87.580**	**0.008**	**87.375**	**5,463**	**87.383**	**0.018**

Simulation Experiment for GSA Parameter (Synthetic Data)

Synthetic Data Experiment for GSA Parameter Comparison
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.001				0.002				0.003			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	97.922	8,753	97.922	0.554	97.845	9,513	97.845	6.3E-01	97.851	9,296	97.851	0.672
2	97.940	6,733	97.940	0.525	97.503	9,883	97.503	0.594	97.490	9,798	97.551	4.9E-01
3	90.279	9,564	90.279	0.369	97.862	9,843	97.902	0.712	97.478	5,728	97.478	0.515
4	92.030	9,733	92.030	0.401	98.066	9,865	98.066	0.515	97.858	9,645	97.858	0.606
5	97.481	4,426	97.481	0.483	97.380	9,188	97.380	0.433	98.030	8,467	98.030	0.827
6	98.049	9,200	98.099	0.692	96.903	9,806	96.974	0.456	97.675	9,881	97.675	0.509
7	97.925	9,591	97.966	0.621	98.343	8,871	98.343	0.683	97.494	5,027	97.494	0.512
8	97.610	9,899	97.610	0.486	97.628	9,073	97.628	0.412	97.655	9,664	97.655	0.525
9	97.237	9,303	97.309	0.392	97.255	9,547	97.370	0.385	97.176	6,521	97.182	0.439
10	97.250	6,693	97.250	0.506	97.708	7,860	97.708	0.476	97.874	7,760	97.905	0.824
	96.372	**8,390**	**96.389**	**0.503**	**97.649**	**9,345**	**97.672**	**0.530**	**97.658**	**8,179**	**97.668**	**0.592**

	0.004				0.005				0.006			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	98.027	9,635	98.029	0.795	97.920	9,947	98.052	0.618	97.922	9,678	97.922	0.723
2	97.317	7,349	97.317	0.426	98.075	9,246	98.075	0.718	97.802	9,943	97.802	0.630
3	97.558	9,510	97.611	0.557	97.898	8,604	97.954	0.575	97.833	9,880	97.934	0.712
4	97.658	9,685	97.658	0.550	97.768	9,752	97.768	0.569	97.133	9,984	97.138	0.603
5	97.539	9,857	97.663	0.541	97.961	7,893	97.961	0.473	97.388	9,053	97.417	0.525
6	97.960	9,856	97.960	0.712	97.936	9,954	98.023	0.582	97.561	9,944	97.568	0.493
7	97.824	7,261	97.824	0.648	96.802	4,475	96.802	0.399	97.677	9,363	97.677	0.563
8	91.419	1,779	91.419	0.313	97.493	5,335	97.493	0.560	97.770	5,194	97.866	0.575
9	97.657	9,574	97.657	0.808	97.843	9,944	97.843	0.683	89.354	5,904	89.354	0.412
10	97.769	9,979	97.807	0.436	98.146	5,037	98.146	0.660	97.415	9,802	97.484	0.603
	97.073	**8,449**	**97.095**	**0.579**	**97.784**	**8,019**	**97.812**	**0.584**	**96.786**	**8,875**	**96.816**	**0.584**

	0.007				0.008				0.009			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	98.125	6,774	98.125	7E-01	97.856	8,871	97.876	0.878	97.517	6,145	97.517	0.463
2	98.427	9,559	98.427	0.827	98.287	8,070	98.287	8.1E-01	97.508	7,650	97.508	0.550
3	97.248	9,561	97.324	0.416	97.843	8,441	97.843	0.531	98.187	2,358	98.187	0.827
4	98.204	9,771	98.204	0.606	97.985	9,402	97.985	0.751	97.683	9,386	97.702	0.695
5	97.583	9,014	97.583	0.443	97.207	7,459	97.207	0.582	97.184	3,560	97.184	0.550
6	97.314	7,908	97.354	0.371	91.521	9,821	91.645	0.541	97.412	8,757	97.533	0.412
7	97.920	4,510	97.920	0.496	97.806	3,978	97.806	0.795	97.962	9,895	97.974	0.648
8	98.106	9,864	98.106	0.675	97.987	8,863	97.987	0.701	97.748	9,468	97.748	0.557
9	97.656	5,737	97.656	0.591	97.664	9,080	97.701	0.563	98.027	7,112	98.027	0.531
10	97.362	9,993	97.362	0.489	97.568	6,771	97.568	0.535	97.627	9,830	97.627	0.550
	97.794	**8,269**	**97.806**	**0.562**	**97.172**	**8,076**	**97.190**	**0.669**	**97.686**	**7,416**	**97.701**	**0.578**

Synthetic Data Experiment for GSA Parameter Comparison
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.1				0.2				0.3			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	97.976	9,845	98.117	0.672	97.370	9,838	97.370	4.7E-01	97.928	8,301	97.957	0.544
2	97.355	4,481	97.355	0.460	98.155	7,849	98.155	0.751	97.909	9,963	97.968	7.0E-01
3	97.923	2,720	98.008	0.784	98.327	7,681	98.327	0.557	97.839	9,859	98.109	0.677
4	97.556	9,254	97.613	0.557	97.521	9,721	97.650	0.483	97.569	9,673	97.569	0.426
5	97.589	7,525	97.589	0.502	97.729	9,740	97.834	0.746	98.045	8,414	98.045	0.512
6	96.911	8,836	96.975	0.388	97.415	9,990	97.439	0.757	97.319	9,807	97.319	0.506
7	97.882	9,856	98.075	0.735	98.169	9,782	98.169	0.800	97.612	9,891	97.628	0.446
8	97.880	9,343	98.141	0.642	98.083	9,281	98.083	0.618	97.638	9,783	97.675	0.473
9	96.298	4,541	96.298	0.392	97.120	9,365	97.174	0.554	98.175	9,456	98.175	0.827
10	97.803	8,123	97.803	0.525	98.110	3,688	98.110	0.845	97.783	5,401	97.783	0.531
	97.517	**7,452**	**87.439**	**-0.002**	**97.800**	**8,694**	**87.439**	**-0.002**	**97.782**	**9,055**	**87.439**	**-0.002**

	0.4				0.5			
Run	Fitness	Fit_Conv	Best_Resu	WK	Fitness	Fit_Conv	Best_Resu	WK
1	97.776	9,809	98.098	0.894	62.695	10,001	67.367	-0.014
2	97.798	8,924	97.867	0.695	47.129	10,001	73.317	-0.006
3	97.752	7,557	97.752	0.499	47.983	10,000	68.587	-0.037
4	91.424	9,950	91.518	0.494	49.650	10,001	70.065	-0.020
5	97.408	9,993	97.497	0.502	48.030	10,000	69.905	-0.030
6	97.850	9,321	97.850	0.519	47.642	10,001	67.731	-0.011
7	93.220	9,099	93.220	0.512	50.196	10,001	66.539	-0.014
8	98.004	8,347	98.004	0.443	51.671	10,000	66.519	-0.006
9	97.687	9,931	98.105	0.878	41.807	10,001	69.362	0.005
10	98.124	9,744	98.124	0.566	51.321	10,001	67.118	-0.003

Synthetic Data Experiment for GSA Parameter Comparison
Modelling Strategy : 10 Folds
Iterations : 10,000
Modelling Method : Naïve Bayes Updateable

	0.01				0.02				0.03			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	97.536	9,080	97.555	0.483	98.131	7,121	98.199	6E-01	97.829	9,042	97.829	0.648
2	98.145	9,819	98.160	0.827	87.272	8,079	87.272	0.231	97.092	8,604	97.145	4.9E-01
3	97.344	9,526	97.397	0.423	97.474	9,700	97.518	0.686	98.082	9,808	98.108	0.642
4	98.065	7,681	98.065	0.512	98.503	5,497	98.503	0.582	97.632	6,573	97.632	0.512
5	97.905	8,773	97.905	0.712	97.574	8,147	97.574	0.460	92.039	9,524	92.268	0.450
6	97.012	9,529	97.012	0.499	93.283	9,060	93.292	0.659	91.456	8,633	91.456	0.334
7	97.317	9,153	97.401	0.443	97.930	9,463	98.003	0.550	98.201	5,602	98.201	0.784
8	97.615	9,884	97.615	0.642	98.123	9,991	98.123	0.660	97.342	9,728	97.345	0.423
9	93.249	9,731	93.249	0.498	91.371	8,981	91.371	0.298	91.624	9,962	91.641	0.383
10	97.873	9,752	97.922	0.499	97.790	7,616	97.790	0.557	98.104	9,500	98.104	0.811
	97.206	**9,293**	**97.228**	**0.554**	**95.745**	**8,366**	**95.764**	**0.527**	**95.940**	**8,698**	**95.973**	**0.548**

	0.04				0.05				0.06			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	97.456	7,897	97.482	0.531	96.984	8,946	97.032	0.506	97.81	8,312	97.810	0.703
2	97.702	8,571	97.702	0.506	98.138	8,887	98.148	0.588	97.88	9,855	97.898	0.563
3	97.620	2,954	97.620	0.519	97.094	9,998	97.246	0.466	98.11	9,842	98.216	0.677
4	97.938	9,997	97.938	0.703	98.347	5,145	98.347	0.642	95.93	9,781	95.933	0.453
5	97.974	3,421	97.974	0.550	97.957	9,472	97.957	0.486	98.01	8,708	98.009	0.569
6	97.928	9,727	97.928	0.591	89.809	9,912	89.809	0.361	98.48	4,635	98.479	0.759
7	97.402	8,748	97.402	0.476	96.634	9,644	97.041	0.402	91.83	9,938	91.833	0.294
8	98.106	9,802	98.130	0.563	93.096	8,598	93.096	0.593	97.83	9,584	97.831	0.630
9	97.987	8,354	97.990	0.718	97.526	9,653	97.526	0.566	98.29	8,251	98.294	0.588
10	97.733	9,589	97.758	0.695	98.186	6,948	98.186	0.600	98.07	9,171	98.074	0.751
	97.785	**7,906**	**97.792**	**0.585**	**96.377**	**8,720**	**96.439**	**0.521**	**97.23**	**8,808**	**97.238**	**0.599**

	0.07				0.08				0.09			
Run	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK	Fitness	Fit_Conv	Best_Result	WK
1	98.331	9,796	98.331	7.9E-01	96.822	9,721	96.822	0.496	97.068	9,423	97.068	0.439
2	97.216	5,900	97.238	0.489	98.034	8,004	98.034	4.9E-01	97.829	8,168	97.829	0.677
3	97.843	9,772	97.843	0.832	97.682	9,333	97.747	0.483	97.267	7,682	97.267	0.563
4	97.233	9,490	97.233	0.433	96.004	9,765	96.004	0.399	98.149	8,408	98.149	0.563
5	97.646	8,257	97.646	0.550	96.978	3,825	96.978	0.446	97.235	9,138	97.235	0.483
6	97.856	9,803	97.913	0.878	97.842	8,263	97.842	0.557	97.467	9,289	97.508	0.486
7	97.537	9,223	97.537	0.666	97.677	9,950	97.677	0.486	96.711	9,865	96.934	0.476
8	97.686	7,772	97.686	0.531	97.134	6,578	97.134	0.419	98.180	9,045	98.180	0.735
9	97.947	9,998	97.947	0.563	97.965	8,190	98.006	0.483	97.246	9,518	97.371	0.499
10	97.804	9,793	97.880	0.618	98.048	6,699	98.048	0.651	97.533	9,744	97.552	0.544
	97.710	**8,980**	**97.725**	**0.635**	**97.419**	**8,033**	**97.429**	**0.491**	**97.468**	**9,028**	**97.509**	**0.547**

Appendix 6C Computation of Confidence Interval

A Confidence Interval (*CI*) is a range of values computed using sample statistics to estimate an unknown population parameter with a given confidence level.

A confidence level is the proportion of all samples randomly drawn from the population whose confidence intervals contain the estimated population parameter. The common confidences level used are 90%, 95% and 99%.

The equation for computing *CI* used to compute the accuracy interval limit is as follows:-

$$CI = \bar{x} \pm Z \frac{s}{N}$$

Where $\bar{x}$ = mean of the sample

Z = a Z-score value

σ = standard deviation

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

N = sample size

This work uses 95% *CI* which means that the data (accuracy values) are likely lies between the *CI* with 95% confidence.

For 95% the *Z* value is **1.960**. Experiment results for the RGFGSA method (synthetic data) with 10FCV modelling strategy are as follows:-

Experiment No.	**Final Accuracy Value**
1	99.1556
2	98.9760
3	98.9003
4	98.3683
5	98.8190
6	98.5696
7	98.7594
8	98.5090
9	98.7694
10	99.0355
11	98.9131
12	93.2695
13	98.9278
14	98.9708
15	99.0162
16	98.5847
17	98.8684
18	98.9894
19	98.8505
20	99.0349
21	98.9845
22	98.7850
23	98.9327
24	98.7579
25	98.8538
Average	**98.6241**

$\bar{x}$ = 98.6241

Z = 1.960

s = standard deviation

N = sample size

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$\bar{x}$ = 98.6241

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = 30.6692$$

$$s = \sqrt{\frac{1}{24}(30.6692)}$$

$$s = 1.1304$$

Therefore the CI is computed as follows:-

$$CI = \bar{x} \pm Z\frac{s}{N}$$

Upper limit = $98.6241 + \left[1.960\,\frac{1.1304}{\sqrt{25}}\right]$

= 99.0672

Lower limit = $98.6241 - \left[1.960\,\frac{1.1304}{\sqrt{25}}\right]$

= 98.1809

Appendix 6D Prediction Accuracy Results

Prediction Accuracy Results – Visual Field

Data : Visual Field Data (Sampled Every Iteration)
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance 10 fold : 1.539244854
Fitness Tolerance 2 fold : 0.346508305

	10 Folds CV					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	88.2229	16	26,422	1,767	88.2229	0.0055
2	87.8989	15	61,321	2,594	87.8989	-0.0205
3	88.1751	15	23,160	23,159	88.1751	-0.0065
4	87.8255	14	66,504	3,092	87.8255	0.0327
5	87.8682	15	30,286	13,827	87.8682	-0.0201
6	87.5575	15	24,092	24,091	87.5575	-0.0359
7	87.3693	15	16,390	16,389	87.3693	0.0020
8	87.8388	12	93,359	4,733	87.8388	0.0081
9	88.1541	15	48,347	5,691	88.1541	0.0371
10	86.3165	14	11,430	72,190	86.3165	-0.0203
11	88.0525	15	55,008	2,870	88.0525	0.0228
12	88.0114	13	41,251	5,995	88.0114	-0.0250
13	88.4824	12	34,555	34,554	88.4824	0.0011
14	87.7130	18	17,336	552	87.7130	-0.0292
15	87.2492	12	63,688	8,898	87.7761	0.0176
16	87.8179	13	68,250	4,678	87.8179	0.0185
17	87.6310	13	41,751	4,670	87.6310	0.0149
18	87.6091	17	4,539	4,538	87.6091	-0.0061
19	87.2627	15	69,509	2,403	87.2627	-0.0098
20	87.5714	12	48,007	18,288	87.5714	-0.0271
21	87.7863	14	6,999	6,998	87.7863	0.0309
22	88.1325	14	17,742	17,741	88.1325	0.0165
23	87.7605	16	81,741	2,749	87.7605	0.0075
24	88.0132	16	74,968	6,290	88.0132	0.0165
25	87.6142	14	43,753	2,322	87.6142	0.0072
Average	**87.7574**		**42816.32**	**11643.16**	**87.7784**	**0.0015**

Data : Visual Field Data (Sampled Every Iteration)
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance 10 fold : 1.539244854
Fitness Tolerance 2 fold : 0.346508305

	2 Folds CV with 10 Repeats					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	86.2089	12	18,623	18,623	86.2089	0.0050
2	86.7437	11	51,703	51,703	86.7437	0.0025
3	86.5759	12	76,986	4,585	86.5759	-0.0005
4	86.1297	16	87,047	2,827	86.1297	0.0181
5	86.4557	9	89,419	7,750	86.4557	-0.0295
6	86.3639	13	68,228	19,411	86.3639	0.0161
7	86.3038	9	78,468	7,623	86.3070	-0.0242
8	86.1487	8	77,008	7,060	86.1487	0.0192
9	86.6139	11	61,834	4,627	86.6139	-0.0026
10	86.4399	11	94,370	17,533	86.4399	-0.0030
11	86.7215	7	87,916	87,915	86.7215	0.0040
12	87.0854	10	82,084	82,083	87.0854	0.0343
13	86.8797	9	84,686	30,338	86.8797	-0.0055
14	86.5854	7	48,243	10,697	86.5854	-0.0096
15	86.6709	13	59,014	13,094	86.6709	-0.0072
16	86.1709	14	5,085	5,085	86.1709	-0.0254
17	86.1108	12	92,073	5,374	86.1108	0.0065
18	86.6171	8	57,836	8,602	86.6171	-0.0046
19	86.5728	12	85,009	7,576	86.5728	0.0015
20	86.6297	12	79,325	79,325	86.6297	0.0293
21	86.1392	8	80,496	3,094	86.1392	-0.0086
22	86.4747	14	63,640	7,503	86.4747	0.0209
23	85.9430	17	707	707	85.9430	0.0469
24	86.2373	15	87,168	13,973	86.2373	-0.0033
25	86.3987	10	48,360	1,343	86.3987	-0.0194
Average	**86.4489**		**66,613**	**19,938**	**86.4490**	**0.0024**

Data : Visual Field Data (Sampled Every Iteration)
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance 10 fold : 1.539244854
Fitness Tolerance 2 fold : 0.346508305

	No Cross Validation					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	86.5823	10	23,130	23,130	86.5823	-0.0241
2	86.3924	13	52,521	52,521	86.3924	-0.0168
3	86.8354	7	75,818	75,818	86.8987	0.0192
4	86.6456	6	67,347	67,347	86.6456	0.0341
5	86.5823	9	49,703	49,703	86.5823	0.0049
6	86.5823	8	56,470	56,470	86.5823	-0.0422
7	86.2658	11	29,195	29,195	86.2658	0.0410
8	86.8354	9	84,053	84,053	86.8354	0.0049
9	86.3924	10	62,481	62,481	86.3924	0.0150
10	86.4557	10	15,357	15,357	86.4557	0.0086
11	86.6456	10	59,775	59,775	86.6456	0.0009
12	87.0253	6	31,151	31,151	87.0253	-0.0235
13	86.5190	10	43,387	43,387	86.5190	-0.0271
14	86.3291	15	34,978	34,978	86.3291	-0.0361
15	87.4684	8	27,132	27,132	87.4684	-0.0211
16	86.5823	14	31,877	31,877	86.5823	0.0013
17	87.2152	8	21,155	21,155	87.2152	0.0435
18	86.5823	12	28,441	28,441	86.5823	-0.0260
19	86.2658	12	47,267	47,267	86.2658	0.0638
20	86.7089	12	36,152	36,152	86.7089	-0.0019
21	86.6456	7	64,381	64,381	86.6456	-0.0401
22	86.6456	9	44,328	44,328	86.6456	0.0205
23	86.3924	15	66,914	66,914	86.3924	-0.0429
24	86.5823	12	42,332	42,332	86.5823	0.0379
25	86.6456	11	87,255	87,255	86.6456	0.0185
Average	**86.6329**		**47,304**	**47,304**	**86.6354**	**0.0005**

Data : Visual Field Data (Sampled Every Iteration)
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.01
Fitness Tolerance 10 fold : 1.568740349
Fitness Tolerance 2 fold : 0.4526

	10 Folds CV					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	87.8408	7	71,671	4,078	87.8408	-0.0098
2	87.4891	7	67,076	1,294	87.7217	0.0189
3	87.9830	9	23,414	3,648	87.9830	0.0006
4	87.6354	7	56,596	2,799	87.6354	-0.0371
5	87.9576	5	37,415	6,178	87.9576	-0.0157
6	87.8045	7	93,792	1,515	87.8970	-0.0107
7	87.7277	7	79,107	4,092	87.8160	-0.0045
8	87.3958	11	61,669	1,983	87.4961	-0.0008
9	87.7379	9	1,765	1,539	87.7379	0.0590
10	87.6398	9	35,117	2,774	87.7336	-0.0263
11	87.7336	4	15,226	2,995	87.7336	-0.0109
12	88.0592	9	3,102	1,578	88.0592	0.0433
13	87.8415	8	8,253	2,933	87.8415	-0.0195
14	88.3841	7	14,190	4,623	88.3841	-0.0244
15	87.8873	12	90,428	2,772	87.8873	-0.0002
16	87.7615	9	68,839	3,172	87.8126	0.0616
17	88.3445	9	70,840	3,126	88.3445	-0.0270
18	87.4873	8	33,910	736	87.8712	0.0220
19	87.8111	8	98,630	5,411	87.8111	0.0093
20	88.1670	7	92,702	2,131	88.1670	-0.0323
21	88.1363	7	76,685	1,968	88.1641	-0.0157
22	88.5357	8	40,761	3,352	88.5357	0.0404
23	88.2470	3	84,027	2,767	88.2470	0.0100
24	87.4572	9	83,961	1,999	87.4572	-0.0334
25	88.0219	12	33,200	1,534	88.0219	0.0170

Data : Visual Field Data (Sampled Every Iteration)
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.01
Fitness Tolerance 10 fold : 1.568740349
Fitness Tolerance 2 fold : 0.4526

	2 Folds CV with 10 Repeats					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	86.2595	9	79,048	18,490	86.2816	-0.0135
2	86.4082	6	20,935	9,839	86.5285	-0.0055
3	86.7532	8	23,591	12,905	86.7532	-0.0223
4	86.5823	8	96,685	29,924	86.5823	0.0229
5	86.6329	6	29,099	29,098	86.6329	0.0065
6	86.3196	6	33,910	23,945	86.3956	0.0200
7	86.3354	5	81,284	12,344	86.4335	0.0619
8	86.9652	7	42,382	42,381	86.9652	-0.0057
9	86.5791	8	50,395	39,671	86.5791	0.0446
10	86.6582	7	60,520	24,502	86.6582	-0.0232
11	86.9209	3	33,257	16,827	86.9209	-0.0231
12	86.3418	9	86,303	34,217	86.5063	-0.0377
13	86.8924	8	97,472	97,471	86.8924	-0.0071
14	86.4146	7	57,463	23,031	86.4146	-0.0080
15	86.2342	8	99,164	18,783	86.6076	-0.0160
16	86.8228	7	55,049	55,048	86.8228	-0.0269
17	86.8766	7	90,488	13,828	86.8766	0.0266
18	86.5475	7	96,718	38,201	86.6171	0.0438
19	86.8165	9	82,384	82,383	86.8165	0.0104
20	86.3038	12	59,535	48,806	86.3038	-0.0134
21	87.1646	6	56,803	56,802	87.1646	0.0471
22	86.7310	5	95,410	38,590	86.8386	0.0077
23	86.8766	6	85,305	24,671	86.8924	-0.0012
24	86.9430	7	57,304	57,303	86.9430	-0.0169
25	86.4589	7	50,987	23,166	86.6108	0.0266

Data : Visual Field Data (Sampled Every Iteration)
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.01
Fitness Tolerance 10 fold : 1.568740349
Fitness Tolerance 2 fold : 0.4526

	No Cross Validation					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	87.2152	4	73,469	73,469	87.2152	-0.0106
2	86.8354	6	73,684	73,684	86.8354	0.0217
3	86.4557	7	49,990	49,990	86.4557	-0.0013
4	86.6456	9	96,367	96,367	86.7722	-0.0177
5	86.9620	6	43,647	43,647	86.9620	-0.0233
6	86.2025	7	98,483	98,483	86.6456	-0.0286
7	87.0253	9	68,208	68,208	87.0253	-0.0071
8	86.8987	6	67,776	67,776	86.8987	-0.0155
9	86.7722	11	27,195	27,195	86.7722	0.0218
10	86.5190	9	94,594	94,594	86.7722	-0.0199
11	86.3291	9	89,840	89,840	86.3291	0.0377
12	86.2025	6	91,568	91,568	86.7722	-0.0026
13	86.3924	8	82,606	82,606	86.7722	-0.0372
14	87.2785	8	37,973	37,973	87.2785	-0.0159
15	86.2658	7	95,188	95,188	86.3924	-0.0143
16	86.7089	9	40,104	40,104	86.7089	-0.0073
17	86.7089	9	88,903	88,903	86.7089	-0.0173
18	86.9620	8	13,939	13,939	87.0253	0.0311
19	86.7089	14	96,128	96,128	86.7089	-0.0155
20	85.8861	8	97,279	97,279	86.3924	-0.0074
21	86.4557	7	97,167	97,167	86.5823	0.0161
22	86.7089	7	61,021	61,021	86.7089	-0.0242
23	86.5190	7	97,629	97,629	86.5190	-0.0077
24	87.1519	3	87,750	87,750	87.1519	0.0352
25	86.5190	8	77,694	77,694	86.8354	-0.0171

Prediction Accuracy Results – Synthetic Data

Data : Synthetic Data
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance Synthetic 10 fold : 1.181885086
Fitness Tolerance Synthetic 2fold : 0.464052274

	10 Folds CV					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	98.3783	6	53,303	21,338	98.3783	0.7345
2	98.4322	5	50,803	19,598	98.4322	0.9086
3	98.4322	5	51,776	19,467	98.4322	0.9086
4	98.4322	5	54,216	20,381	98.4322	0.9086
5	98.4322	5	53,421	16,939	98.5479	0.9086
6	98.2960	4	56,780	16,655	98.3989	0.9136
7	98.4322	5	49,717	23,961	98.4322	0.9086
8	98.3989	4	87,567	19,829	98.4322	0.8785
9	98.4322	5	50,115	17,548	98.4632	0.9086
10	98.1630	6	61,088	19,119	98.4322	0.7401
11	98.4322	5	52,717	19,672	98.5623	0.9086
12	98.1926	4	38,227	20,128	98.1938	0.8631
13	98.4322	5	48,786	19,279	98.4322	0.9086
14	98.2894	5	77,696	23,299	98.3433	0.8268
15	98.4322	5	51,147	17,101	98.4322	0.9086
16	98.3846	6	51,438	18,025	98.3989	0.8294
17	98.4322	5	51,962	18,798	98.4322	0.9086
18	98.2557	5	50,777	18,491	98.2902	0.8215
19	98.0407	6	43,370	20,368	98.4322	0.7176
20	98.2400	5	42,821	18,863	98.4322	0.9086
21	98.4322	5	50,868	17,685	98.4322	0.9086
22	98.3183	5	41,555	23,143	98.4322	0.8136
23	98.4322	5	50,300	19,793	98.4322	0.9086
24	98.4632	4	42,417	19,906	98.4632	0.8785
25	98.3208	4	70,145	20,174	98.4322	0.8785
Average	**98.3571**		**53,320**	**19,582**	**98.4209**	**0.8683**

Data : Synthetic Data
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance Synthetic 10 fold : 1.181885086
Fitness Tolerance Synthetic 2fold : 0.464052274

	2 Folds CV with 10 Repeats					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	97.0440	3	41,354	30,980	97.0440	1.0000
2	97.0440	3	41,314	28,721	97.0440	1.0000
3	97.0440	3	38,966	33,070	97.0440	1.0000
4	97.0440	3	35,984	32,589	97.0440	1.0000
5	97.0440	3	36,853	29,214	97.0440	1.0000
6	97.0440	3	43,619	29,263	97.0440	1.0000
7	97.0440	3	44,028	28,418	97.0440	1.0000
8	97.0440	3	37,900	28,146	97.0440	1.0000
9	97.0440	3	37,422	33,716	97.0440	1.0000
10	97.0440	3	44,432	29,170	97.0440	1.0000
11	97.0440	3	46,927	30,524	97.0440	1.0000
12	97.0440	3	44,782	28,229	97.0440	1.0000
13	96.8960	4	68,585	26,905	97.0440	0.8785
14	97.0440	3	47,254	31,303	97.0440	1.0000
15	97.0440	3	47,224	27,872	97.0440	1.0000
16	97.0440	3	41,411	31,908	97.0440	1.0000
17	96.8980	4	44,166	30,136	97.0440	0.8936
18	97.0440	3	40,254	31,399	97.0440	1.0000
19	97.0440	3	43,144	28,650	97.0440	1.0000
20	97.0440	3	43,316	29,814	97.0440	1.0000
21	97.0440	3	38,490	32,316	97.0440	1.0000
22	96.7600	4	75,805	27,486	97.0440	0.8936
23	97.0440	3	45,490	30,379	97.0440	1.0000
24	97.0440	3	44,120	35,591	97.0440	1.0000
25	97.0440	3	40,166	33,327	97.0440	1.0000
Average	**97.0209**		**44,520**	**30,365**	**97.0440**	**0.9866**

Data : Synthetic Data
Method : Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
Fitness Tolerance Synthetic 10 fold : 1.181885086
Fitness Tolerance Synthetic 2fold : 0.464052274

	No Cross Validation				
Experiment	Final Fitness Value	Bundles	Convergence Point	Best Accuracy	Weighted Kappa
1	98.6000	4	42,791	98.600	0.9670
2	98.2800	5	51,132	98.600	0.7949
3	98.4400	4	51,741	98.600	0.8631
4	98.6000	4	42,935	98.600	0.9670
5	98.6000	4	44,631	98.600	0.9670
6	98.6000	4	39,769	98.600	0.9670
7	98.6000	4	42,943	98.600	0.9670
8	98.6000	4	45,555	98.600	0.9670
9	98.6000	4	43,197	98.600	0.9670
10	98.6000	4	42,772	98.600	0.9670
11	98.6000	4	41,749	98.600	0.9670
12	98.6000	4	41,274	98.600	0.9670
13	98.2400	7	51,043	98.600	0.6746
14	98.6000	4	42,107	98.600	0.9670
15	98.6000	4	46,131	98.600	0.9670
16	98.6000	4	43,296	98.600	0.9670
17	98.4400	4	54,488	98.600	0.9136
18	98.6000	4	40,595	98.600	0.9670
19	98.6000	4	40,792	98.600	0.9670
20	98.4400	4	59,368	98.600	0.9136
21	98.6000	4	43,474	98.600	0.9670
22	98.6000	4	41,826	98.600	0.9670
23	98.6000	4	41,978	98.600	0.9670
24	98.6000	4	43,530	98.600	0.9670
25	98.6000	4	41,975	98.600	0.9670
Average	**98.5536**		**44,844**	**98.600**	**0.9400**

Data : Synthetic Data
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.009
Fitness Tolerance 10 fold : 1.064513159
Fitness Tolerance 2fold : 0.392248689

	10 Folds CV					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	98.5623	5	51,995	5,135	98.5623	0.8683
2	98.2813	4	99,793	13,148	98.4632	0.8936
3	98.3098	4	99,927	5,553	98.4322	0.9381
4	98.4627	5	55,622	5,019	98.4632	0.8477
5	98.4322	5	99,993	4,572	98.5623	0.9086
6	98.4186	4	99,960	10,768	98.5623	0.8683
7	98.3098	4	99,949	2,613	98.4632	0.9381
8	98.5623	5	99,019	4,512	98.5623	0.8683
9	98.2993	4	99,437	11,002	98.5623	0.8785
10	97.9743	4	99,932	15,672	98.5623	0.8631
11	98.2960	4	99,954	4,619	98.4983	0.9136
12	98.2038	4	99,119	5,805	98.4322	0.8683
13	98.3098	4	99,707	3,006	98.4322	0.9381
14	98.3608	5	97,829	5,958	98.4269	0.8268
15	97.9339	5	89,966	10,081	98.3098	0.8268
16	98.5623	5	96,566	5,458	98.5623	0.8683
17	98.3098	4	98,685	5,790	98.4322	0.9381
18	98.5623	5	96,415	11,784	98.5623	0.8683
19	98.4627	5	83,755	10,383	98.4632	0.8477
20	98.3893	4	99,950	4,346	98.4632	0.8683
21	98.2400	5	89,775	12,522	98.5623	0.9086
22	98.1955	4	99,987	3,836	98.4322	0.9136
23	98.4627	5	71,640	5,234	98.4627	0.8477
24	98.1100	4	99,886	8,766	98.5623	0.9381
25	98.1415	4	99,989	7,618	98.5623	0.8631

Data : Synthetic Data
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.009
Fitness Tolerance 10 fold : 1.064513159
Fitness Tolerance 2fold : 0.392248689

	2 Folds CV with 10 Repeats					
Experiment	Final Fitness Value	Bundles	Convergence Point	Fit.Tolerance Convergence	Best Accuracy	Weighted Kappa
1	97.0440	3	98,887	40,487	97.0440	1.0000
2	97.0440	3	92,600	49,316	97.0440	1.0000
3	96.8660	4	99,978	50,279	97.0440	0.9670
4	97.0440	3	98,398	32,793	97.0440	1.0000
5	97.0440	3	98,933	58,807	97.0440	1.0000
6	97.0440	3	98,685	46,787	97.0440	1.0000
7	97.0440	3	99,062	54,063	97.0440	1.0000
8	97.0440	3	92,059	45,542	97.0440	1.0000
9	97.0440	3	90,056	45,804	97.0440	1.0000
10	97.0440	3	95,483	36,971	97.0440	1.0000
11	97.0440	3	95,814	43,407	97.0440	1.0000
12	97.0440	3	98,278	62,507	97.0440	1.0000
13	97.0440	3	94,925	24,407	97.0440	1.0000
14	97.0440	3	98,449	35,491	97.0440	1.0000
15	97.0440	3	97,358	51,124	97.0440	1.0000
16	97.0440	3	99,287	27,252	97.0440	1.0000
17	97.0440	3	99,243	46,432	97.0440	1.0000
18	97.0440	3	95,902	26,885	97.0440	1.0000
19	97.0440	3	96,471	51,020	97.0440	1.0000
20	97.0440	3	98,714	70,774	97.0440	1.0000
21	97.0440	3	97,235	52,322	97.0440	1.0000
22	97.0440	3	96,953	34,078	97.0440	1.0000
23	97.0440	3	99,180	45,559	97.0440	1.0000
24	97.0440	3	97,370	44,552	97.0440	1.0000
25	97.0440	3	98,104	54,455	97.0440	1.0000
Average	**97.0369**		**97,097**	**45,245**	**97.0440**	**0.9987**

Data : Synthetic Data
Method : Generalised Simulated Annealing
Classifier : Naïve Bayes Updateable
Iteration : 100,000
QA : 0.009
Fitness Tolerance 10 fold : 1.064513159
Fitness Tolerance 2fold : 0.392248689

	No Cross Validation				
Experiment	Final Fitness Value	Bundles	Convergence Point	Best Accuracy	Weighted Kappa
1	98.6000	4	62,038	98.6000	0.9670
2	98.6000	4	55,743	98.6000	0.9670
3	98.6000	4	64,127	98.6000	0.9670
4	98.3600	4	55,447	98.6000	0.9136
5	98.6000	4	66,661	98.6000	0.9670
6	98.6000	4	92,761	98.6000	0.9670
7	98.6000	4	72,152	98.6000	0.9670
8	98.6000	4	50,996	98.6000	0.9670
9	98.6000	4	55,515	98.6000	0.9670
10	98.6000	4	96,471	98.6000	0.9670
11	98.6000	4	90,164	98.6000	0.9670
12	98.6000	4	60,935	98.6000	0.9670
13	98.6000	4	62,130	98.6000	0.9670
14	98.6000	4	96,534	98.6000	0.9670
15	98.6000	4	80,442	98.6000	0.9670
16	98.6000	4	98,626	98.6000	0.9670
17	98.6000	4	55,462	98.6000	0.9670
18	98.6000	4	89,264	98.6000	0.9670
19	98.6000	4	59,229	98.6000	0.9670
20	98.6000	4	56,978	98.6000	0.9670
21	98.6000	4	56,728	98.6000	0.9670
22	98.6000	4	98,201	98.6000	0.9670
23	98.6000	4	54,541	98.6000	0.9670
24	98.6000	4	65,656	98.6000	0.9670
25	98.6000	4	78,374	98.6000	0.9670
Average	**98.5904**		**71,007**	**98.6000**	**0.9649**

Appendix 6E Noisy Fitness Tolerance

Z-scores are expressed in terms of standard deviations from their means in normal distribution data. The *Z*-score value 1.98, which is equivalent to 97.61% data lie within the tolerance limit, is used in the computation of noisy fitness tolerance.

To obtain the noisy fitness tolerance limit, the *Z*-score equation is used:-

$$z = \frac{X - \mu}{\sigma} \qquad \textbf{(Equation 6E-1)}$$

From Equation 6E-1, *X* is the upper limit which is unknown. The limit is defined as:-

$$L = | X - \mu | \qquad \textbf{(Equation 6E-2)}$$

To obtain the value of *X*:-

$$X = (z \times \sigma) + \mu \qquad \textbf{(Equation 6E-3)}$$

In order to get the value of μ, a simulation was carried out on both data (visual field and synthetic). The simulation was run on the data (10,000 iterations) to get the distribution of fitness values (accuracy) for both 10FCV and 2FCV modelling strategy. The fitness value is called 100 times (using the same solution each iteration) for each iteration in the simulation. Computation of noisy tolerance limit for the synthetic data 10FCV is as follows:-

$z = 1.98 \qquad \mu = 89.7445 \qquad \sigma = 0.537633$

$X = (1.98 \times 0.537633) + 89.7445$

$X = 90.80901$

$L = | 90.80901 - 89.7445 |$

$L = 1.06451$

Appendix 7A Combination

A combination is the number of ways to choose a sample of a number of elements from a set of distinct objects where order does not matter and replacements are not allowed.

The number of combinations of *k* objects from a set with *n* objects is termed as ${}_nC^k$.

$$ {}_nC^k = n!/[k!(n-k)!] \qquad \textbf{(Equation 6A-1)} $$

For example, the combinations of {1,2,3,4} taken *k*=2 at a time are:-

4! / [(2!)(4-2) !] = 6 subsets

{1,2}, {1,3}, {1,4}, {2,3}, {2,4}, {3,4}.

Combinations for 2 objects from a set with 25 objects are:-

25! / [(2!)(25-2)!] = 300 subsets.

Appendix 7B *kN* Experiment Results

RGFGSA xN Experiments (N : 45 variables) - Synthetic Data
x : 0.5 - 5.0
10,000 Iterations
10 Repeats
10-fold Cross Validation
QV : 1.4330930018498693 (for 100,000 iterations)

	0.5				1				1.5			
Run	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK
0	97.81	9,996	5	0.718	97.35102	9,994	6	0.787	97.37211	9,989	6	0.789
1	97.87	9,986	6	0.740	97.35708	9,990	7	0.765	98.1007	9,994	5	0.832
2	97.25	9,994	6	0.835	98.25837	9,967	3	1.000	97.91902	9,742	7	0.683
3	98.20	9,953	5	0.848	97.5307	9,938	6	0.845	98.49905	9,955	4	0.878
4	97.53	9,978	4	0.863	98.02776	9,996	4	0.967	97.56571	9,827	5	0.773
5	97.60	9,983	5	0.909	98.33835	9,990	5	0.936	96.67502	9,984	7	0.703
6	97.71	9,953	5	0.848	97.38394	9,961	6	0.819	97.3651	9,997	5	0.868
7	98.51	9,966	3	1.000	97.7384	9,982	5	0.800	97.55983	9,988	6	0.803
8	97.20	9,989	5	0.855	97.14676	9,999	6	0.798	97.27414	9,993	6	0.837
9	97.81	9,992	4	0.894	98.10487	9,993	3	1.000	98.53121	9,992	3	1.000
Avg	**97.75**	**9,979**		**0.851**	**97.72372**	**9,981**		**0.872**	**97.68619**	**9,946**		**0.817**

	2				2.5				3			
Run	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK
0	98.11	9,913	4	0.878	97.54649	9,984	4	0.894	97.90	9,987	3	1.000
1	98.08	9,991	3	1.000	97.26962	9,960	5	0.746	98.14	9,956	4	0.967
2	98.45	9,873	4	0.938	97.25279	9,867	7	0.657	97.24	9,997	5	0.933
3	98.39	9,983	3	1.000	97.4823	9,976	5	0.800	97.50	9,905	6	0.718
4	98.34	9,904	4	0.894	97.9716	9,975	4	0.938	97.90	9,999	5	0.886
5	97.45	9,996	5	0.800	96.79333	9,984	7	0.735	98.10	9,944	5	0.858
6	98.29	7,385	5	0.735	97.48397	9,994	6	0.789	97.43	10,000	6	0.806
7	97.76	10,000	4	0.938	97.88175	9,973	4	0.938	97.58	9,924	4	0.894
8	97.77	9,994	4	0.938	96.90287	9,995	6	0.723	97.73	9,966	4	0.868
9	97.99	9,966	6	0.842	97.30563	9,999	6	0.842	98.07	9,805	6	0.709
Avg	**98.06**	**9,701**		**0.896**	**97.38904**	**9,971**		**0.806**	**97.76**	**9,948**		**0.864**

	3.5				4				4.5			
Run	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK	Fitness	Fit_Conv	Bundles	WK
0	97.98	9,929	5	0.858	97.13748	9,995	4	0.878	98.47283	9225	4	0.868
1	98.00	9,974	4	0.863	97.82702	9,384	6	0.776	98.41963	9,831	4	0.868
2	97.20	9,811	8	0.621	98.06427	9,782	5	0.909	97.33479	9,920	6	0.751
3	97.31	9,980	6	0.751	97.89013	9,984	4	0.967	97.84956	9,844	5	0.795
4	97.60	9,997	5	0.816	97.75402	9,950	5	0.827	97.80279	9,980	4	0.938
5	97.70	9,998	4	0.894	97.84138	9,998	4	0.894	98.41832	9,877	4	0.967
6	98.37	9,989	3	1.000	97.85573	9,887	4	0.868	97.034	9,978	7	0.740
7	98.01	9,965	4	0.878	97.99405	9,996	4	0.967	97.01854	9,988	6	0.850
8	96.81	9,964	6	0.689	97.86694	9,997	3	1.000	97.16567	10,000	6	0.718
9	97.31	9,915	6	0.695	98.04495	9,865	4	0.894	98.18954	9,984	4	0.914
Avg	**97.63**	**9,952**		**0.807**	**97.8276**	**9,884**		**0.898**	**97.77057**	**9,863**		**0.841**

	5			
Run	Fitness	Fit_Conv	Bundles	WK
0	97.22	9,957	6	0.683
1	97.32	10,000	5	0.832
2	97.88	9,998	5	0.768
3	98.11	9,780	4	0.878
4	97.39	9,997	6	0.824
5	98.63	9,470	3	1.000
6	98.12	9,951	4	0.967
7	96.41	9,972	8	0.615
8	98.44	9,888	3	1.000
9	98.41	9,984	3	1.000
Avg	**97.79**	**9,900**		**0.857**

References

Advanced Glaucoma Intervention Study Investigators (1994) 'Advanced Glaucoma Intervention Study: 2. Visual field test scoring and reliability', *Ophthalmology*, 101(8), pp. 1445-1455.

Agostini, F.P., Soares-Pinto, D.D.O., Moret, M.A., Osthoff, C. and Pascutti, P.G. (2006) 'Generalized simulated annealing applied to protein folding studies', *Journal of computational chemistry*, 27(11), pp. 1142-1155.

Ajaz, R.H. and Hussain, L. (2015) 'Seed classification using machine learning techniques', *Journal of Multidisciplinary Engineering Science and Technology*, 2(5).

Akosa, J. (2017) 'Predictive Accuracy: A Misleading Performance Measure for Highly Imbalanced Data', *Proceedings of the SAS Global Forum.*

Akram, S. and ul Ann, Q. (2015) 'Newton Raphson Method', *International Journal of Scientific & Engineering Research*, 6, pp. 7.

Alam, F. and Pachauri, S. (2017) 'Detection using WEKA', *Advances in Computational Sciences and Technology*, 10(6), pp. 1731-1743.

Al-Jefri, M.M., Evans, R., Ghezzi, P. and Uchyigit, G. (2017) 'Using machine learning for automatic identification of evidence-based health information on the web', *Proceedings of the 2017 International Conference on Digital Health.* ACM, 167-174.

Almeida, H., Meurs, M., Kosseim, L., Butler, G. and Tsang, A. (2014) 'Machine learning for biomedical literature triage', *PloS one*, 9(12), pp. e115892.

Alpaydin, E. (2014) *Introduction to machine learning.* MIT press.

Ananya Mandal (2012) *Causes of visual impairment*. Available at: http://www.news-medical.net/health/Causes-of-visual-impairment.aspx (Accessed: December/22 2016).

Anderson, D., Drance, S. and Schulzer, M. (1998) 'The effectiveness of intraocular pressure reduction in the treatment of normal-tension glaucoma', *American Journal of Ophthalmology*, 126(4), pp. 498-505.

Andricioaei, I. and Straub, J.E. (1996) 'Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide', *PHYSICAL REVIEW-SERIES E-*, 53, pp. R3055-R3055.

Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Paliouras, G. and Spyropoulos, C.D. (2000) 'An evaluation of naive bayesian anti-spam filtering', *arXiv preprint cs/0006013*, .

Artes, P.H., Nicolela, M.T., LeBlanc, R.P. and Chauhan, B.C. (2005) 'Visual field progression in glaucoma: total versus pattern deviation analyses', *Investigative ophthalmology & visual science*, 46(12), pp. 4600-4606.

Asaoka, R., Iwase, A., Hirasawa, K., Murata, H. and Araie, M. (2014) 'Identifying "Preperimetric" Glaucoma in Standard Automated Perimetry Visual FieldsVisual Fields in Preperimetric Glaucoma', *Investigative ophthalmology & visual science*, 55(12), pp. 7814-7820.

Askarzadeh, A. (2016) 'A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm', *Computers & Structures*, 169, pp. 1-12.

Athey, S. (2017) 'Beyond prediction: Using big data for policy problems', *Science (New York, N.Y.)*, 355(6324), pp. 483-485.

Bagheri, S., Konen, W., Allmendinger, R., Branke, J., Deb, K., Fieldsend, J., Quagliarella, D. and Sindhya, K. (2017) 'Constraint handling in efficient global optimization', *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 673-680.

Balachandran, K. and ANITHA, D. (2014) 'Feature selection based on the classifier models: Performance issues in the pre-diagnosis of lung cancer', *Journal of Theoretical and Applied Information Technology*, 59(3).

Balachandran, K. and Anitha, R. (2012) 'Dimensionality reduction based on the classifier models: Performance Issues in the prediction of Lung cancer', *Software Engineering (CONSEG), 2012 CSI Sixth International Conference on*. IEEE, 1-4.

Barocas, S., Bradley, E., Honavar, V. and Provost, F. (2017) 'Big Data, Data Science, and Civil Rights', *arXiv preprint arXiv:1706.03102*, .

Basseur, M. and Goëffon, A. (2013) 'Hill-climbing strategies on various landscapes: an empirical comparison', *Proceeding of the fifteenth annual conference on Genetic and evolutionary computation conference*. ACM, 479-486.

Bebie, H., Fankhauser, F. and Spahr, J. (1976) 'Static perimetry: accuracy and fluctuations', *Acta Ophthalmologica*, 54(3), pp. 339-348.

Bengio, Y. and Grandvalet, Y. (2004) 'No unbiased estimator of the variance of k-fold cross-validation', *Journal of machine learning research*, 5(Sep), pp. 1089-1105.

Bhargava, N., Sharma, G., Bhargava, R. and Mathuria, M. (2013) 'Decision tree analysis on j48 algorithm for data mining', *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, 3(6).

Birch, M.K., Wishart, P.K. and O'Donnell, N.P. (1995) 'Determining progressive visual field loss in serial Humphrey visual fields', *Ophthalmology*, 102(8), pp. 1227-1235.

Bizios, D., Heijl, A. and Bengtsson, B. (2007) 'Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms', *Journal of glaucoma*, 16(1), pp. 20-28.

Boden, C., Blumenthal, E.Z., Pascual, J., McEwan, G., Weinreb, R.N., Medeiros, F. and Sample, P.A. (2004) 'Patterns of glaucomatous visual field progression identified by three progression criteria', *American Journal of Ophthalmology*, 138(6), pp. 1029-1036.

Bohachevsky, I.O., Johnson, M.E. and Myron L. Stein (1986) 'Generalized Simulated Annealing for Function Optimization', *Technometrics*, 28(3), pp. 209-217.

Bolón-Canedo, V., Sánchez-Maroño, N. and Alonso-Betanzos, A. (2013) 'A review of feature selection methods on synthetic data', *Knowledge and information systems*, 34(3), pp. 483-519.

Bonyadi, M.R. and Michalewicz, Z. (2016) 'Analysis of stability, local convergence, and transformation sensitivity of a variant of the particle swarm optimization algorithm', *IEEE Transactions on Evolutionary Computation*, 20(3), pp. 370-385.

Bose, I. and Chen, X. (2009) 'Hybrid models using unsupervised clustering for prediction of customer churn', *Journal of Organizational Computing and Electronic Commerce*, 19(2), pp. 133-151.

Bourne, R.R. (2006) 'The optic nerve head in glaucoma', *Community eye health / International Centre for Eye Health*, 19(59), pp. 44-45.

Bowd, C. and Goldbaum, M.H. (2008) 'Machine learning classifiers in glaucoma', *Optometry and vision science : official publication of the American Academy of Optometry*, 85(6), pp. 396-405.

Brinkman, G., Vance, G., Hannigan, M.P. and Milford, J.B. (2006) 'Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for PM2. 5 exposure data', *Environmental science & technology*, 40(6), pp. 1892-1901.

Broadway, D.C. (2012) 'Visual field testing for glaucoma - a practical guide', *Community eye health,* 25(79-80), pp. 66-70.

Brooks, D.G. and Verdini, W.A. (1988) 'Computational experience with generalized simulated annealing over continuous variables', *American Journal of Mathematical and Management Sciences,* 8(3-4), pp. 425-449.

Brooks, S.P. and Morgan, B.J. (1995) 'Optimization using simulated annealing', *The Statistician,* , pp. 241-257.

Browne, M.W. (2000) 'Cross-validation methods', *Journal of mathematical psychology,* 44(1), pp. 108-132.

Brünger, A., Krukowski, A. and Erickson, J.W. (1990) 'Slow-cooling protocols for crystallographic refinement by simulated annealing', *Acta Crystallographica Section A: Foundations of Crystallography,* 46(7), pp. 585-593.

Brusini, P. and Johnson, C.A. (2007) 'Staging functional damage in glaucoma: review of different classification methods', *Survey of ophthalmology,* 52(2), pp. 156-179.

Bryan, K., Cunningham, P. and Bolshakova, N. (2005) 'Biclustering of expression data using simulated annealing', *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on.* IEEE, 383-388.

Bryan, S.R., Vermeer, K.A., Eilers, P.H., Lemij, H.G. and Lesaffre, E.M. (2013) 'Robust and censored modeling and prediction of progression in glaucomatous visual fields', *Investigative ophthalmology & visual science,* 54(10), pp. 6694-6700.

Burgansky-Eliash, Z., Wollstein, G., Chu, T., Ramsey, J.D., Glymour, C., Noecker, R.J., Ishikawa, H. and Schuman, J.S. (2005) 'Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study', *Investigative ophthalmology & visual science,* 46(11), pp. 4147-4152.

Busetti, F. (2003) 'Simulated annealing overview', *World Wide Web URL www.geocities.com/francorbusetti/saweb.pdf,* .

Campbell, L.R., Dahlberg, S., Dorward, R., Gerhard, J., Grubb, T., Purcell, C. and Sagan, B.E. (2016) 'Restricted growth function patterns and statistics', *arXiv preprint arXiv:1605.04807,* .

Cao, H., Sen, P.K., Peery, A.F. and Dellon, E.S. (2016) 'Assessing agreement with multiple raters on correlated kappa statistics', *Biometrical Journal,* 58(4), pp. 935-943.

Capó, M., Pérez, A. and Lozano, J.A. (2017) 'An efficient approximation to the K-means clustering for massive data', *Knowledge-Based Systems*, 117, pp. 56-69.

Caprioli, J., Mock, D., Bitrian, E., Afifi, A.A., Yu, F., Nouri-Mahdavi, K. and Coleman, A.L. (2011) 'A method to measure and predict rates of regional visual field decay in glaucoma', *Investigative ophthalmology & visual science*, 52(7), pp. 4765-4773.

Casabianca, J.M. and Junker, B.W. (2016) 'Multivariate normal distribution', *Handbook of Item Response Theory, Volume Two: Statistical Tools*, 21, pp. 35.

Ceccon, S., Garway-Heath, D.F., Crabb, D.P. and Tucker, A. (2014) 'Exploring early glaucoma and the visual field test: Classification and clustering using bayesian networks', *IEEE journal of biomedical and health informatics*, 18(3), pp. 1008-1014.

Ceccon, S., Garway-Heath, D., Crabb, D. and Tucker, A. (2012) 'Non-stationary clustering bayesian networks for glaucoma', *Proceedings of the Workshop on machine Learning for Clincial Data Analysis, ICML 2012*.

Chan, K., Lee, T., Sample, P.A., Goldbaum, M.H., Weinreb, R.N. and Sejnowski, T.J. (2002) 'Comparison of machine learning and traditional classifiers in glaucoma diagnosis', *IEEE Transactions on Biomedical Engineering*, 49(9), pp. 963-974.

Chang, T., Hong, M., Liao, W. and Wang, X. (2016) 'Asynchronous distributed ADMM for large-scale optimization—part I: algorithm and convergence analysis', *IEEE Transactions on Signal Processing*, 64(12), pp. 3118-3130.

Charnay, C., Lachiche, N. and Braud, A. (2015) 'Construction of complex aggregates with random restart hill-climbing', in *Inductive Logic Programming*. Springer, pp. 49-61.

Chawla, N.V., Japkowicz, N. and Kotcz, A. (2004) 'Special issue on learning from imbalanced data sets', *ACM Sigkdd Explorations Newsletter*, 6(1), pp. 1-6.

Cheng, Y. and Church, G.M. (2000) 'Biclustering of expression data.', *Ismb.* , 93-103.

Christian Nordqvist (2016) *Glaucoma: Symptoms, Causes, Treatments*. Available at: http://www.medicalnewstoday.com/articles/9710.php (Accessed: 29 December 2016).

Clements, M.P. and Hendry, D.F. (2002) 'Modelling methodology and forecast failure', *The Econometrics Journal*, 5(2), pp. 319-344.

Crabb, D.P., Edgar, D.F., Fitzke, F.W., McNaught, A.I. and Wynn, H.P. (1995) 'New approach to estimating variability in visual field data using an image processing technique', *The British journal of ophthalmology,* 79(3), pp. 213-217.

Dada, T., Dave, V. and Mithal, N. (2009) 'Medical Management of Glaucoma', *Journal of Current Glaucoma Practice,* 3(3), pp. 13-17.

De Keyser, M., De Belder, M. and De Groot, V. (2017) 'Quality of life in glaucoma patients after selective laser trabeculoplasty', *International journal of ophthalmology,* 10(5), pp. 742.

Dekkers, A. and Aarts, E. (1991) 'Global optimization and simulated annealing', *Mathematical Programming,* 50(1), pp. 367-393.

Diab, D.M. and El Hindi, K.M. (2017) 'Using differential evolution for fine tuning naïve Bayesian classifiers and its application for text classification', *Applied Soft Computing,* 54, pp. 183-199.

dos R Correia, E., Nascimento, V.B., de Castilho, C.M., Esperidiao, A.S., Soares, E.A. and de Carvalho, V.E. (2005) 'The generalized simulated annealing algorithm in the low energy electron diffraction search problem', *Journal of Physics: Condensed Matter,* 17(1), pp. 1.

Drance, S.M. (1969) 'The early field defects in glaucoma', *Investigative ophthalmology & visual science,* 8(1), pp. 84-91.

Dreiseitl, S., Ohno-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H. and Binder, M. (2001) 'A comparison of machine learning methods for the diagnosis of pigmented skin lesions', *Journal of Biomedical Informatics,* 34(1), pp. 28-36.

Edelman, G.M. and Gally, J.A. (2001) 'Degeneracy and complexity in biological systems', *Proceedings of the National Academy of Sciences of the United States of America,* 98(24), pp. 13763-13768.

Enshaeifar, S., Hoseinitabatabaei, S., Ahrabian, A. and Barnaghi, P. (2017) 'Pattern Identification for State Prediction in Dynamic Data Streams', .

Ester, M., Kriegel, H., Sander, J. and Xu, X. (1996) 'A density-based algorithm for discovering clusters in large spatial databases with noise.', *Kdd.* , 226-231.

Fageeri, S.O., Ahmed, S.M.M., Almubarak, S.A. and Mu'azu, A.A. (2017) 'Eye refractive error classification using machine learning techniques', *Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017 International Conference on.* IEEE, 1-6.

Falkenauer, E. (1998) *Genetic algorithms and grouping problems.* John Wiley & Sons, Inc.

Falkenauer, E. and Delchambre, A. (1992) 'A genetic algorithm for bin packing and line balancing', *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on.* IEEE, 1186-1192.

Fang, L., Chen, P. and Liu, S. (2007) 'Particle swarm optimization with simulated annealing for TSP', *Proceedings of the 6th Conference on 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases.* , 206-210.

Fang, R., Pouyanfar, S., Yang, Y., Chen, S. and Iyengar, S. (2016) 'Computational health informatics in the big data age: a survey', *ACM Computing Surveys (CSUR),* 49(1), pp. 12.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters,* 27(8), pp. 861-874.

Fechtner, R.D. and Weinreb, R.N. (1994) 'Mechanisms of optic nerve damage in primary open angle glaucoma', *Survey of ophthalmology,* 39(1), pp. 23-42.

Fernández, A., del Río, S., Chawla, N.V. and Herrera, F. (2017) 'An insight into imbalanced Big Data classification: outcomes and challenges', *Complex & Intelligent Systems,* , pp. 1-16.

Finkelstein, J. (2017) 'Machine learning approaches to personalize early prediction of asthma exacerbations', *Annals of the New York Academy of Sciences,* 1387(1), pp. 153-165.

Fister Jr, I., Yang, X., Fister, I., Brest, J. and Fister, D. (2013) 'A brief review of nature-inspired algorithms for optimization', *arXiv preprint arXiv:1307.4186,* .

Fitzke, F.W., Crabb, D.P., McNaught, A.I., Edgar, D.F. and Hitchings, R.A. (1995) 'Image processing of computerised visual field data', *The British journal of ophthalmology,* 79(3), pp. 207-212.

Fitzke, F.W., Hitchings, R.A., Poinoosawmy, D., McNaught, A.I. and Crabb, D.P. (1996) 'Analysis of visual field progression in glaucoma', *The British journal of ophthalmology,* 80(1), pp. 40-48.

Flammer, J., Drance, S., Augustiny, L. and Funkhouser, A. (1985) 'Quantification of glaucomatous visual field defects with automated perimetry.', *Investigative ophthalmology & visual science,* 26(2), pp. 176-181.

Foster, K.R., Koprowski, R. and Skufca, J.D. (2014) 'Machine learning, medical diagnosis, and biomedical engineering research-commentary', *Biomedical engineering online,* 13(1), pp. 94.

Foster, P.J. and Johnson, G.J. (2001) 'Glaucoma in China: how big is the problem?', *The British journal of ophthalmology,* 85(11), pp. 1277-1282.

Fraley, C. and Raftery, A.E. (2002) 'Model-based clustering, discriminant analysis, and density estimation', *Journal of the American statistical Association,* 97(458), pp. 611-631.

Frank, E. and Bouckaert, R. (2006) 'Naive bayes for text classification with unbalanced classes', *Knowledge Discovery in Databases: PKDD 2006,* , pp. 503-510.

Franklin, L.R. (2005) 'Exploratory experiments', *Philosophy of Science,* 72(5), pp. 888-899.

Friedman, J.H. (1997) 'On bias, variance, 0/1—loss, and the curse-of-dimensionality', *Data mining and knowledge discovery,* 1(1), pp. 55-77.

Friedman, N., Geiger, D. and Goldszmidt, M. (1997) 'Bayesian network classifiers', *Machine Learning,* 29(2-3), pp. 131-163.

Fukui, T., Sato, S. and Takahashi, A. (2016) 'Estimating Style Weights of Mutual Funds by Monte Carlo Filter with Generalized Simulated Annealing', .

Fushiki, T. (2011) 'Estimation of prediction error by using K-fold cross-validation', *Statistics and Computing,* 21(2), pp. 137-146.

Gagneur, J., Friedel, C., Heun, V., Zimmer, R. and Rost, B. (2017) 'Bioinformatics advances biology and medicine by turning big data troves into knowledge', *Informatik-Spektrum,* 40(2), pp. 153-160.

Garey, M.R. and Johnson, D.S. (1979) 'Computers and intractability: a guide to the theory of NP-completeness. 1979', *San Francisco, LA: Freeman,* 58.

Garway-Heath, D.F., Poinoosawmy, D., Fitzke, F.W. and Hitchings, R.A. (2000) 'Mapping the visual field to the optic disc in normal tension glaucoma eyes', *Ophthalmology,* 107(10), pp. 1809-1815.

Gasch, A.T., Wang, P. and Pasquale, L.R. (2000) 'Determinants of glaucoma awareness in a general eye clinic', *Ophthalmology,* 107(2), pp. 303-308.

George-Nektarios, T. (2013) 'Weka classifiers summary', *Athens University of Economics and Bussiness Intracom-Telecom, Athens*, .

Glaucoma Research Foundation (2014) *Symptoms of Open-Angle Glaucoma*. Available at: http://www.glaucoma.org/glaucoma/symptoms-of-primary-open-angle-glaucoma.php (Accessed: December/22 2016).

Glaucoma Research Foundation (2013) *Five Common Glaucoma Tests*. Available at: http://www.glaucoma.org/glaucoma/diagnostic-tests.php (Accessed: December/19 2016).

Goffe, W.L., Ferrier, G.D. and Rogers, J. (1994) 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics*, 60(1-2), pp. 65-99.

Gravina, D., Liapis, A. and Yannakakis, G.N. (2017) 'Surprise Search for Evolutionary Divergence', *arXiv preprint arXiv:1706.02556*, .

Haley, M.J. (1986) *The field analyzer primer.* Allergan Humphrey.

Hamming, R.W. (1950) 'Error detecting and error correcting codes', *Bell Labs Technical Journal*, 29(2), pp. 147-160.

Han, T.D. and Abdelrahman, T.S. (2017) 'Use of Synthetic Benchmarks for Machine-Learning-Based Performance Auto-Tuning', *Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2017 IEEE International.* IEEE, 1350-1361.

Hayreh, S.S. (2011) 'Structure of the Optic Nerve', in *Ischemic Optic Neuropathies.* Springer, pp. 7-34.

He, H. and Garcia, E.A. (2009) 'Learning from imbalanced data', *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp. 1263-1284.

He, Y., Pan, W. and Lin, J. (2006) 'Cluster analysis using multivariate normal mixture models to detect differential gene expression with microarray data', *Computational Statistics & Data Analysis*, 51(2), pp. 641-658.

Heijl, A., Leske, M.C., Bengtsson, B., Hyman, L., Bengtsson, B. and Hussein, M. (2002) 'Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial', *Archives of Ophthalmology*, 120(10), pp. 1268-1279.

Heijl, A., Lindgren, G. and Olsson, J. (1987) 'A package for the statistical analysis of visual fields', *Seventh International Visual Field Symposium, Amsterdam, September 1986.* Springer, 153-168.

Heijl, A. and Patella, V.M. (2002) *Essential perimetry: The field analyzer primer.* Carl Zeiss Meditec.

Heimberg, G., Bhatnagar, R., El-Samad, H. and Thomson, M. (2016) 'Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing', *Cell systems,* 2(4), pp. 239-250.

Helsgaun, K. (2009) 'General k-opt submoves for the Lin–Kernighan TSP heuristic', *Mathematical Programming Computation,* 1(2-3), pp. 119-163.

Henson, D., Spenceley, S.E. and Bull, D. (1997) 'Artificial neural network analysis of noisy visual field data in glaucoma', *Artificial Intelligence in Medicine,* 10(2), pp. 99-113.

Herland, M., Khoshgoftaar, T.M. and Wald, R. (2014) 'A review of data mining using big data in health informatics', *Journal of Big Data,* 1(1), pp. 2.

Hill, E.J., Roberts, C., Franklin, J.M., Enescu, M., West, N., MacGregor, T.P., Chu, K.Y., Boyle, L., Blesing, C., Wang, L.M., Mukherjee, S., Anderson, E.M., Brown, G., Dutton, S., Love, S.B., Schnabel, J.A., Quirke, P., Muschel, R., McKenna, W.G., Partridge, M. and Sharma, R.A. (2016) 'Clinical Trial of Oral Nelfinavir before and during Radiation Therapy for Advanced Rectal Cancer', *Clinical cancer research : an official journal of the American Association for Cancer Research,* 22(8), pp. 1922-1931.

Hoffman, K.L., Padberg, M. and Rinaldi, G. (2013) 'Traveling salesman problem', in *Encyclopedia of operations research and management science.* Springer, pp. 1573-1578.

Hoffmann, J. (2001) 'FF: The fast-forward planning system', *AI magazine,* 22(3), pp. 57.

Hsu, C., Chang, C. and Lin, C. (2003) 'A practical guide to support vector classification', .

Hutchinson, J.K. (2012) 'Optic neuropathies: glaucomatous vs. non-glaucomatous', *18th annual glaucoma report.Rev Optom,* 149, pp. 58.

Iester, M., Capris, P., Pandolfo, A., Zingirian, M. and Traverso, C.E. (2000) 'Learning effect, short-term fluctuation, and long-term fluctuation in frequency doubling technique', *American Journal of Ophthalmology,* 130(2), pp. 160-164.

Jacobson, S.H. and Yücesan, E. (2004) 'Analyzing the performance of generalized hill climbing algorithms', *Journal of Heuristics,* 10(4), pp. 387-405.

Jain, A.K., Murty, M.N. and Flynn, P.J. (1999) 'Data clustering: a review', *ACM computing surveys (CSUR)*, 31(3), pp. 264-323.

James C. Tsai (2016) *High Eye Pressure and Glaucoma*. Available at: http://www.glaucoma.org/gleams/high-eye-pressure-and-glaucoma.php (Accessed: 10 January 2017).

Jay, J.L. and Murdoch, J.R. (1993) 'The rate of visual field loss in untreated primary open angle glaucoma', *The British journal of ophthalmology*, 77(3), pp. 176-178.

Jebb, A.T., Parrigon, S. and Woo, S.E. (2017) 'Exploratory data analysis as a foundation of inductive research', *Human Resource Management Review*, 27(2), pp. 265-276.

Jennifer Skillen (2007) *Understanding Visual Fields*. Available at: http://www.ssc.education.ed.ac.uk/courses/VI&multi/vnov072i.html (Accessed: December/19 2016).

Jiao, L. and Wang, L. (2000) 'A novel genetic algorithm based on immunity', *IEEE Transactions on Systems, Man, and Cybernetics-part A: systems and humans*, 30(5), pp. 552-561.

John Berdahl (2016) *Glaucoma: Types, Symptoms, Diagnosis And Treatment*. Available at: http://www.allaboutvision.com/conditions/glaucoma.htm (Accessed: December/29 2016).

John, G.H. and Langley, P. (1995) 'Estimating continuous distributions in Bayesian classifiers', *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 338-345.

Jung, J., Jayakrishnan, R. and Park, J.Y. (2016) 'Dynamic Shared-Taxi Dispatch Algorithm with Hybrid-Simulated Annealing', *Computer-Aided Civil and Infrastructure Engineering*, 31(4), pp. 275-291.

Kale, B., Siravuri, H.V., Alhoori, H. and Papka, M.E. (2017) 'Predicting Research that will be Cited in Policy Documents', *arXiv preprint arXiv:1706.04140*, .

Katz, J., Gilbert, D., Quigley, H.A. and Sommer, A. (1997) 'Estimating progression of visual field loss in glaucoma', *Ophthalmology*, 104(6), pp. 1017-1025.

Kelley, C.T. (2003) *Solving nonlinear equations with Newton's method*. Siam.

Kerrigan-Baumrind, L.A., Quigley, H.A., Pease, M.E., Kerrigan, D.F. and Mitchell, R.S. (2000) 'Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons', *Investigative ophthalmology & visual science,* 41(3), pp. 741-748.

Khanmohammadi, S., Adibeig, N. and Shanehbandy, S. (2017) 'An improved overlapping k-means clustering method for medical applications', *Expert Systems with Applications,* 67, pp. 12-18.

Khaw, P.T., Wells, A.P. and Lim, K.S. (2003) 'Surgery for glaucoma', *The British journal of ophthalmology,* 87(4), pp. 517.

Kiang, M.Y. (2003) 'A comparative assessment of classification methods', *Decision Support Systems,* 35(4), pp. 441-454.

Kibriya, A.M., Frank, E., Pfahringer, B. and Holmes, G. (2004) 'Multinomial Naive Bayes for Text Categorization Revisited.', *Australian Conference on Artificial Intelligence.* Springer, 488-499.

Kim, S.J., Cho, K.J. and Oh, S. (2017) 'Development of machine learning models for diagnosis of glaucoma', *PloS one,* 12(5), pp. e0177726.

Kirkpatrick, S., Gelatt, C.D.,Jr and Vecchi, M.P. (1983) 'Optimization by simulated annealing', *Science (New York, N.Y.),* 220(4598), pp. 671-680.

Klaver, C.C., Wolfs, R.C., Vingerling, J.R., Hofman, A. and de Jong, P.T. (1998) 'Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam Study', *Archives of Ophthalmology,* 116(5), pp. 653-658.

Klein, P.N. and Young, N.E. (2010) 'Approximation algorithms for NP-hard optimization problems', *Algorithms and theory of computation handbook.* Chapman & Hall/CRC, 34-34.

Kodinariya, T.M. and Makwana, P.R. (2013) 'Review on determining number of Cluster in K-Means Clustering', *International Journal,* 1(6), pp. 90-95.

Kohavi, R. (1995) 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *Ijcai.* , 1137-1145.

Kojima, K. (1971) 'Is there a constant fitness value for a given genotype? NO!', *Evolution,* 25(2), pp. 281-285.

Krichene, A. and Krichene, A. (2017) 'Using a naive Bayesian classifier methodology for loan risk assessment: Evidence from a Tunisian commercial bank', *Journal of Economics, Finance and Administrative Science,* 22(42), pp. 3-24.

Kriegel, H., Kröger, P. and Zimek, A. (2009) 'Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering', *ACM Transactions on Knowledge Discovery from Data (TKDD),* 3(1), pp. 1.

Kristie Draskovic, John J. McSoley (2016) *Automated Perimetry: Visual Field Deficits in Glaucoma and Beyond*. Available at: https://www.reviewofoptometry.com/ce/automated-perimetry-visual-field-deficits-in-glaucoma-and-beyond (Accessed: 17 January 2016).

Landhuis, E. (2017) 'Neuroscience: big brain, big data', *Nature,* 541(7638), pp. 559-561.

Lang, K.J. (2016) 'Hill climbing beats genetic search on a boolean circuit synthesis problem of koza's', *Proceedings of the Twelfth International Conference on Machine Learning.* , 340-343.

Lee, D.A. and Higginbotham, E.J. (2005) 'Glaucoma and its treatment: a review.', *American journal of health-system pharmacy,* 62(7).

Lee, H., Kim, H., Kim, M. and Kim, J. (2016) 'A fast convergence LLL algorithm with fixed-complexity for SIC-based MIMO detection', *Information Networking (ICOIN), 2016 International Conference on.* IEEE, 439-441.

Leung, K.M. (2007) 'Naive bayesian classifier', *Polytechnic University Department of Computer Science/Finance and Risk Engineering,* .

Lilliefors, H.W. (1967) 'On the Kolmogorov-Smirnov test for normality with mean and variance unknown', *Journal of the American statistical Association,* 62(318), pp. 399-402.

Lim, A., Rodrigues, B. and Zhang, X. (2006) 'A simulated annealing and hill-climbing algorithm for the traveling tournament problem', *European Journal of Operational Research,* 174(3), pp. 1459-1478.

Lin, S. (1965) 'Computer solutions of the traveling salesman problem', *Bell System Technical Journal,* 44(10), pp. 2245-2269.

Liu, X., Cheng, G. and Wu, J.X. (1994) 'Identifying the measurement noise in glaucomatous testing: an artificial neural network approach', *Artificial Intelligence in Medicine,* 6(5), pp. 401-416.

Lodi, A., Martello, S. and Vigo, D. (2002) 'Heuristic algorithms for the three-dimensional bin packing problem', *European Journal of Operational Research,* 141(2), pp. 410-420.

Loechner, J. (2016) *90% Of Today's Data Created In Two Years*. Available at: https://www.mediapost.com/publications/article/291358/90-of-todays-data-created-in-two-years.html (Accessed: 22 December 2017).

M. Z. M. B. Jilani, A. Tucker and S. Swift (2016) 'Simultaneous Modelling and Clustering of Visual Field Data', *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS).* , 213-218.

Matwin, S. and Sazonova, V. (2012) 'Direct comparison between support vector machine and multinomial naive Bayes algorithms for medical abstract classification', *Journal of the American Medical Informatics Association,* 19(5), pp. 917-917.

McAfee, A. and Brynjolfsson, E. (2012) 'Big data: the management revolution', *Harvard business review,* 90(10), pp. 60-68.

McKean-Cowdin, R., Varma, R., Wu, J., Hays, R.D., Azen, S.P. and Los Angeles Latino Eye Study Group (2007) 'Severity of visual field loss and health-related quality of life', *American Journal of Ophthalmology,* 143(6), pp. 1013-1023.

Meer, K. (2007) 'Simulated annealing versus metropolis for a TSP instance', *Information Processing Letters,* 104(6), pp. 216-219.

Meilă, M. and Heckerman, D. (2001) 'An experimental comparison of model-based clustering methods', *Machine Learning,* 42(1-2), pp. 9-29.

Menin, O.H. and Bauch, C.T. (2017) 'Solving the patient zero inverse problem by using generalized simulated annealing', *Physica A: Statistical Mechanics and its Applications,* .

Menin, O.H., Martinez, A.S. and Costa, A. (2016) 'Reconstruction of bremsstrahlung spectra from attenuation data using generalized simulated annealing', *Applied Radiation and Isotopes,* 111, pp. 80-85.

Metwally, A., Agrawal, D. and El Abbadi, A. (2005) 'Efficient computation of frequent and top-k elements in data streams', *International Conference on Database Theory.* Springer, 398-412.

Mikelberg, F.S. (1986) 'Visual field examination in glaucoma.', *Optometry & Vision Science,* 63(6), pp. 477-482.

Mirjalili, S. and Gandomi, A.H. (2017) 'Chaotic gravitational constants for the gravitational search algorithm', *Applied Soft Computing*, 53, pp. 407-419.

Mitchell, M., Holland, J.H. and Forrest, S. (1994) 'When will a genetic algorithm outperform hill climbing', *Advances in neural information processing systems.* , 51-58.

Mojica, O. and Bassrei, A. (2015) 'Application of the Generalized Simulated Annealing Algorithm to the Solution of 2D Gravity Inversion of Basement Relief', *3rd Latin American Geosciences Student Conference.*

Morales, E., de Leon, John Mark S, Abdollahi, N., Yu, F., Nouri-Mahdavi, K. and Caprioli, J. (2016) 'Enhancement of Visual Field Predictions with Pointwise Exponential Regression (PER) and Pointwise Linear Regression (PLR)', *Translational vision science & technology*, 5(2), pp. 12-12.

Moret, M.A., Pascutti, P.G., Bisch, P.M. and Mundim, K.C. (1998) 'Stochastic molecular optimization using generalized simulated annealing', *Journal of computational chemistry*, 19(6), pp. 647-657.

Nakamura, K. and Hong, B. (2016) 'Fast-convergence superpixel algorithm via an approximate optimization', *Journal of Electronic Imaging*, 25(5), pp. 053035-053035.

Nayak, J., Acharya, R., Bhat, P.S., Shetty, N. and Lim, T. (2009) 'Automated diagnosis of glaucoma using digital fundus images', *Journal of medical systems*, 33(5), pp. 337.

NHS Choices (2016) *Causes of glaucoma*. Available at: http://www.nhs.uk/Conditions/Glaucoma/Pages/Causes.aspx (Accessed: December/19 2016).

Nonnemaker, J. and Baird, H.S. (2009) 'Using synthetic data safely in classification', *bold*, 2, pp. 76.

Nouri-Mahdavi, K., Hoffman, D., Coleman, A.L., Liu, G., Li, G., Gaasterland, D. and Caprioli, J. (2004) 'Predictive factors for glaucomatous visual field progression in the Advanced Glaucoma Intervention Study', *Ophthalmology*, 111(9), pp. 1627-1635.

Nouri-Mahdavi, K., Hoffman, D., Gaasterland, D. and Caprioli, J. (2004) 'Prediction of visual field progression in glaucoma', *Investigative ophthalmology & visual science*, 45(12), pp. 4346-4351.

Nouri-Mahdavi, K., Hoffman, D., Ralli, M. and Caprioli, J. (2007) 'Comparison of methods to predict visual field progression in glaucoma', *Archives of Ophthalmology,* 125(9), pp. 1176-1181.

O'Neil, M.A. and Burtscher, M. (2015) 'Rethinking the parallelization of random-restart hill climbing: a case study in optimizing a 2-opt TSP solver for GPU execution', *Proceedings of the 8th Workshop on General Purpose Processing using GPUs.* ACM, 99-108.

Otori, Y., Takahashi, G., Urashima, M., Kuwayama, Y. and Quality of Life Improvement Committee (2017) 'Evaluating the Quality of Life of Glaucoma Patients Using the State Trait Anxiety Inventory', *Journal of glaucoma,* .

Özgür, A. and Erdem, H. (2017) 'The impact of using large training data set KDD99 on classification accuracy', *PeerJ Preprints,* 5, pp. e2838v1.

Papadimitriou, C.H. and Steiglitz, K. (1982) *Combinatorial optimization: algorithms and complexity.* Courier Corporation.

Park, E., Chang, H.J. and Nam, H.S. (2017) 'Use of Machine Learning Classifiers and Sensor Data to Detect Neurological Deficit in Stroke Patients', *Journal of medical Internet research,* 19(4), pp. e120.

Pascolini, D. and Mariotti, S.P. (2012) 'Global estimates of visual impairment: 2010', *The British journal of ophthalmology,* 96(5), pp. 614-618.

Patil, T.R. and Sherekar, S. (2013) 'Performance analysis of Naive Bayes and J48 classification algorithm for data classification', *International Journal of Computer Science and Applications,* 6(2), pp. 256-261.

Pavlidis, S., Swift, S., Tucker, A. and Counsell, S. (2013) 'The Modelling of Glaucoma Progression through the Use of Cellular Automata', *International Symposium on Intelligent Data Analysis.* Springer, 322-332.

Pendergast, S.D. and McCuen, B.W. (1996) 'Visual field loss after macular hole surgery', *Ophthalmology,* 103(7), pp. 1069-1077.

Peng, W., Chen, J. and Zhou, H. (2009) 'An implementation of ID3-decision tree learning algorithm', *From web.arch.usyd.edu.au/wpeng/DecisionTree2.pdf Retrieved date: May,* 13.

Penna, T.J. (1995) 'Traveling salesman problem and Tsallis statistics', *Physical Review E,* 51, pp. 1.

Phung, S.L., Bouzerdoum, A. and Nguyen, G.H. (2009) 'Learning pattern classification tasks with imbalanced data sets'.

Pradeep Ramulu (2017) *Standard Automated Perimetry*. Available at: http://eyewiki.aao.org/Standard_Automated_Perimetry (Accessed: 28 March 2017).

Quigley, H.A. and Broman, A.T. (2006) 'The number of people with glaucoma worldwide in 2010 and 2020', *The British journal of ophthalmology*, 90(3), pp. 262-267.

Quigley, H.A. (2011) 'Glaucoma', *The Lancet*, 377(9774), pp. 1367-1377.

Radcliffe, N.J. and Surry, P.D. (1994) 'Fitness Variance of Formae and Performance Prediction.', *FOGA*. Citeseer, 51-72.

Ramrattan, R.S., Wolfs, R.C., Panda-Jonas, S., Jonas, J.B., Bakker, D., Pols, H.A., Hofman, A. and de Jong, P.T. (2001) 'Prevalence and causes of visual field loss in the elderly and associations with impairment in daily functioning: the Rotterdam Study', *Archives of Ophthalmology*, 119(12), pp. 1788-1794.

Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R. (2003) 'Tackling the poor assumptions of naive bayes text classifiers', *ICML*. Washington DC), 616-623.

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G.P. and Mariotti, S.P. (2004) 'Global data on visual impairment in the year 2002', *Bulletin of the World Health Organization*, 82(11), pp. 844-851.

Ritch, R. and Schlötzer-Schrehardt, U. (2001) 'Exfoliation syndrome', *Survey of ophthalmology*, 45(4), pp. 265-315.

Ritch, R., Schlötzer-Schrehardt, U. and Konstas, A.G. (2003) 'Why is glaucoma associated with exfoliation syndrome?', *Progress in retinal and eye research*, 22(3), pp. 253-275.

Rizzo, M.I., Greco, A., De Virgilio, A., Gallo, A., Taverniti, L., Fusconi, M., Conte, M., Pagliuca, G., Turchetta, R. and de Vincentiis, M. (2017) 'Glaucoma: recent advances in the involvement of autoimmunity', *Immunologic research*, 65(1), pp. 207-217.

Rokach, L. and Maimon, O. (2005) 'Clustering methods', *Data mining and knowledge discovery handbook*, , pp. 321-352.

Romeijn, H.E. and Smith, R.L. (1994) 'Simulated annealing for constrained global optimization', *Journal of Global Optimization*, 5(2), pp. 101-126.

Ronan, T., Qi, Z. and Naegle, K.M. (2016) 'Avoiding common pitfalls when clustering biological data', *Science signaling*, 9(432), pp. re6.

Rudolph, G. (1994) 'Convergence analysis of canonical genetic algorithms', *IEEE Transactions on Neural Networks*, 5(1), pp. 96-101.

Rutenbar, R. (1989) 'Simulated annealing algorithms: an overview', *Circuits and Devices Magazine, IEEE*, 5(1), pp. 19-26.

Sacchi, L., Tucker, A., Counsell, S., Garway-Heath, D. and Swift, S. (2014) 'Improving predictive models of glaucoma severity by incorporating quality indicators', *Artificial Intelligence in Medicine*, 60(2), pp. 103-112.

Selman, B. and Gomes, C.P. (2006) 'Hill-climbing search', *Encyclopedia of Cognitive Science*, 81, pp. 82.

Shavlik, J.W., Mooney, R.J. and Towell, G.G. (1991) 'Symbolic and neural learning algorithms: An experimental comparison', *Machine Learning*, 6(2), pp. 111-143.

Shih, Y. (1999) 'Families of splitting criteria for classification trees', *Statistics and Computing*, 9(4), pp. 309-315.

Signor, S.A., Arbeitman, M.N. and Nuzhdin, S.V. (2016) 'Gene networks and developmental context: the importance of understanding complex gene expression patterns in evolution', *Evolution & development*, 18(3), pp. 201-209.

Sivapriya, C. and Latha, P. (2017) 'Optic Nerve Head Segmentation for Early Diagnosis of Glaucoma Based on Active Contour Model', *International Journal of Engineering Science*, 5177.

Skiścim, C.C. and Golden, B.L. (1983) 'Optimization by simulated annealing: A preliminary computational study for the tsp', *Proceedings of the 15th conference on Winter Simulation-Volume 2*. IEEE Press, 523-535.

Sleath, B., Sayner, R., Vitko, M., Carpenter, D.M., Blalock, S.J., Muir, K.W., Giangiacomo, A.L., Hartnett, M.E. and Robin, A.L. (2017) 'Glaucoma patient-provider communication about vision quality-of-life', *Patient education and counseling*, 100(4), pp. 703-709.

Spry, P.G.D. and Johnson, C.A. (2002) 'Identification of Progressive Glaucomatous Visual Field Loss', *Survey of ophthalmology*, 47(2), pp. 158-173.

ST APOR, K. (2006) 'Support vector clustering algorithm for identification of glaucoma in ophthalmology', *bulletin of the polish academy of sciences technical sciences*, 54(1).

Stylianou, A. and Talias, M.A. (2017) 'Big data in healthcare: a discussion on the big challenges', *Health and Technology*, 7(1), pp. 97-107.

Sullivan-Mee, M., Halverson, K.D., Saxon, G.B., Saxon, M.C., Shafer, K.M., Sterling, J.A., Sterling, M.J. and Qualls, C. (2005) 'The relationship between central corneal thickness-adjusted intraocular pressure and glaucomatous visual-field loss', *Optometry-Journal of the American Optometric Association*, 76(4), pp. 228-238.

Sun, Y., Kamel, M.S., Wong, A.K. and Wang, Y. (2007) 'Cost-sensitive boosting for classification of imbalanced data', *Pattern Recognition*, 40(12), pp. 3358-3378.

Sundar, P.P. (2013) 'A Comparative Study For Predicting Students Academic Performance using Bayesian Network Classifiers', *IOSR Journal of Engineering (IOSRJEN) e-ISSN*, , pp. 2250-3021.

Sutter, J.M., Dixon, S.L. and Jurs, P.C. (1995) 'Automated descriptor selection for quantitative structure-activity relationships using generalized simulated annealing', *Journal of chemical information and computer sciences*, 35(1), pp. 77-84.

Swift, S. and Liu, X. (2002) 'Predicting glaucomatous visual field deterioration through short multivariate time series modelling', *Artificial Intelligence in Medicine*, 24(1), pp. 5-24.

Swift, S., Tucker, A., Crampton, J. and Garway-Heath, D. (2007) 'An improved restricted growth function genetic algorithm for the consensus clustering of retinal nerve fibre data', *Proceedings of the 9th annual conference on Genetic and evolutionary computation.* ACM, 2174-2181.

Swift, S., Tucker, A., Vinciotti, V., Martin, N., Orengo, C., Liu, X. and Kellam, P. (2004) 'Consensus clustering and functional interpretation of gene-expression data', *Genome biology*, 5(11), pp. R94.

Szu, H. and Hartley, R. (1987) 'Fast simulated annealing', *Physics letters A*, 122(3), pp. 157-162.

Tao, W. and Wei-hua, L. (2010) 'Naive bayes software defect prediction model', *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on.* IEEE, 1-4.

Tasoulis, D., Plagianakos, V. and Vrahatis, M. (2004) 'Unsupervised clustering of bioinformatics data', *European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, Eunite.* , 47-53.

Taylor, M. and Mildenberger, T. (2017) 'Extending electronic length frequency analysis in R', *Fisheries Management and Ecology,* 24(4), pp. 330-338.

Thylefors, B. and Negrel, A.D. (1994) 'The global impact of glaucoma', *Bulletin of the World Health Organization,* 72(3), pp. 323-326.

Tielsch, J.M., Sommer, A., Katz, J., Royall, R.M., Quigley, H.A. and Javitt, J. (1991) 'Racial variations in the prevalence of primary open-angle glaucoma: the Baltimore Eye Survey', *Jama,* 266(3), pp. 369-374.

Tovée, M.J. (2008) *An Introduction to the Visual System.* Second edn. Cambridge, UK: Cambridge University Press.

Tovey, C.A. (1985) 'Hill climbing with multiple local optima', *SIAM Journal on Algebraic Discrete Methods,* 6(3), pp. 384-393.

Tran, B., Xue, B. and Zhang, M. (2016) 'Genetic programming for feature construction and selection in classification on high-dimensional data', *Memetic Computing,* 8(1), pp. 3-15.

Tsai, Y. and Chatterjee, A. (2017) 'Comprehensive, Quantitative Crack Detection Algorithm Performance Evaluation System', *Journal of Computing in Civil Engineering,* 31(5), pp. 04017047.

Tsallis, C. (1988) 'Possible generalization of Boltzmann-Gibbs statistics', *Journal of statistical physics,* 52(1-2), pp. 479-487.

Tsallis, C. and Stariolo, D.A. (1996) 'Generalized simulated annealing', *Physica A: Statistical Mechanics and its Applications,* 233(1), pp. 395-406.

Tucker, A., Crampton, J. and Swift, S. (2005) 'Rgfga: An efficient representation and crossover for grouping genetic algorithms', *Evolutionary computation,* 13(4), pp. 477-499.

Tucker, A., Vinciotti, V., Liu, X. and Garway-Heath, D. (2005) 'A spatio-temporal Bayesian network classifier for understanding visual field deterioration', *Artificial Intelligence in Medicine,* 34(2), pp. 163-177.

Turpin, A., Frank, E., Hall, M., Witten, I. and Johnson, C. (2001) 'Determining progression in glaucoma using visual fields', *Advances in Knowledge Discovery and Data Mining,* , pp. 136-147.

VanBuren, J., Oleson, J.J., Zamba, G.K. and Wall, M. (2016) 'Integrating independent spatio-temporal replications to assess population trends in disease spread', *Statistics in medicine,* 35(28), pp. 5210-5221.

Varma, S. and Simon, R. (2006) 'Bias in error estimation when using cross-validation for model selection', *BMC bioinformatics,* 7(1), pp. 1.

Varty, Z. (2017) 'Simulated Annealing Overview', .

Veneziano, D., Di Bella, S., Nigita, G., Laganà, A., Ferro, A. and Croce, C.M. (2016) 'Non-coding RNA: Current Deep Sequencing Data Analysis Approaches and Challenges', *Human mutation,* .

Viera, A.J. and Garrett, J.M. (2005) 'Understanding interobserver agreement: the kappa statistic', *Fam Med,* 37(5), pp. 360-363.

Viswanathan, A.C., Fitzke, F.W. and Hitchings, R.A. (1997) 'Early detection of visual field progression in glaucoma: a comparison of PROGRESSOR and STATPAC 2', *The British journal of ophthalmology,* 81(12), pp. 1037-1042.

Vizarim, N.P., Carlone, M., Verga, L.G. and Venegas, P.A. (2017) 'Critical Forces at Fractional Matching Fields in Superconducting Thin Films with Triangular Pinning Lattice', *Materials Research,* (AHEAD), pp. 0-0.

Wan, C. and Freitas, A.A. (2017) 'An empirical evaluation of hierarchical feature selection methods for classification in bioinformatics datasets with gene ontology-based features', *Artificial Intelligence Review,* , pp. 1-40.

Wandell, B.A., Dumoulin, S.O. and Brewer, A.A. (2007) 'Visual field maps in human cortex', *Neuron,* 56(2), pp. 366-383.

Wang, M., Li, B., Zhang, G. and Yao, X. (2017) 'Population Evolvability: Dynamic Fitness Landscape Analysis for Population-based Metaheuristic Algorithms', *IEEE Transactions on Evolutionary Computation,* .

Wang, Y. and Hajli, N. (2017) 'Exploring the path to big data analytics success in healthcare', *Journal of Business Research,* 70, pp. 287-299.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007) 'Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy', *Applied and Environmental Microbiology,* 73(16), pp. 5261-5267.

Weinreb, R.N. and Khaw, P.T. (2004) 'Primary open-angle glaucoma', *The Lancet,* 363(9422), pp. 1711-1720.

Whitley, D., Sutton, A.M., Ochoa, G. and Chicano, F. (2014) 'The component model for elementary landscapes and partial neighborhoods', *Theoretical Computer Science,* 545, pp. 59-75.

Wolpert, D.H. and Macready, W.G. (1997) 'No free lunch theorems for optimization', *IEEE transactions on evolutionary computation,* 1(1), pp. 67-82.

Woods, C.T. and Laederach, A. (2017) 'Classification of RNA structure change by 'gazing'at experimental data', *Bioinformatics,* 33(11), pp. 1647-1655.

World Health Organization (2007) 'Global Initiative for the Elimination of Avoidable Blindness: action plan 2006-2011', .

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B. and Philip, S.Y. (2008) 'Top 10 algorithms in data mining', *Knowledge and information systems,* 14(1), pp. 1-37.

Xiang, Y. and Gong, X. (2000) 'Efficiency of generalized simulated annealing', *Physical Review E,* 62(3), pp. 4473.

Xiang, Y., Sun, D., Fan, W. and Gong, X. (1997) 'Generalized simulated annealing algorithm and its application to the Thomson model', *Physics Letters A,* 233(3), pp. 216-220.

Xiang, Y., Gubian, S. and Martin, F. (2017) 'Generalized Simulated Annealing', in *Computational Optimization in Engineering-Paradigms and Applications.* InTech.

Xiang, Y., Gubian, S., Suomela, B. and Hoeng, J. (2013) 'Generalized Simulated Annealing for Global Optimization: The GenSA Package.', *R Journal,* 5(1).

Xie, L., Draizen, E.J. and Bourne, P.E. (2017) 'Harnessing big data for systems pharmacology', *Annual Review of Pharmacology and Toxicology,* 57, pp. 245-262.

Yang, M., Yang, J., Zhang, Q., Niu, Y. and Li, J. (2013) 'Classification of retinal image for automatic cataract detection', *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on.* IEEE, 674-679.

Yang, X. (2009) 'Firefly algorithms for multimodal optimization', *International symposium on stochastic algorithms.* Springer, 169-178.

Yang, Y. and Chen, W. (2016) 'Taiga: performance optimization of the C4. 5 decision tree construction algorithm', *Tsinghua Science and Technology,* 21(4), pp. 415-425.

Yang, Z., Huo, H. and Fang, T. (2017) 'Automatically finding the number of clusters based on simulated annealing', *Journal of Shanghai Jiaotong University (Science),* 22(2), pp. 139-147.

Yau, S.S., Wang, J., Niknejad, A., Lu, C., Jin, N. and Ho, Y.K. (2003) 'DNA sequence representation without degeneracy', *Nucleic acids research,* 31(12), pp. 3078-3080.

Yeung, K.Y., Fraley, C., Murua, A., Raftery, A.E. and Ruzzo, W.L. (2001) 'Model-based clustering and data transformations for gene expression data', *Bioinformatics,* 17(10), pp. 977-987.

Yoon, K. and Kwek, S. (2007) 'A data reduction approach for resolving the imbalanced data issue in functional genomics', *Neural Computing and Applications,* 16(3), pp. 295-306.

Zhang, D., Chow, C., Liu, A., Zhang, X., Ding, Q. and Li, Q. (2017) 'Efficient evaluation of shortest travel-time path queries through spatial mashups', *GeoInformatica,* , pp. 1-26.

Zhang, H., Kang, Y., Zhu, Y., Zhao, K., Liang, J., Ding, L., Zhang, T. and Zhang, J. (2017) 'Novel naïve Bayes classification models for predicting the chemical Ames mutagenicity', *Toxicology in Vitro,* 41, pp. 56-63.

Zhang, J., Wang, S., Chen, L. and Gallinari, P. (2017) 'Multiple Bayesian discriminant functions for high-dimensional massive data classification', *Data Mining and Knowledge Discovery,* 31(2), pp. 465-501.

Zhang, P. (1993) 'Model selection via multifold cross validation', *The Annals of Statistics,* , pp. 299-313.

Zhang, W., Feng, H., Wu, H. and Zheng, X. (2017) 'Accounting for tumor purity improves cancer subtype classification from DNA methylation data', *Bioinformatics,* , pp. btx303.

Zhang, X., Dastiridou, A., Francis, B.A., Tan, O., Varma, R., Greenfield, D.S., Schuman, J.S., Huang, D. and Advanced Imaging for Glaucoma Study Group (2017) 'Comparison of Glaucoma Progression Detection by Optical Coherence Tomography and Visual Field', *American Journal of Ophthalmology,* .

Zhao, Y., Healy, B.C., Rotstein, D., Guttmann, C.R., Bakshi, R., Weiner, H.L., Brodley, C.E. and Chitnis, T. (2017) 'Exploration of machine learning techniques in predicting multiple sclerosis disease course', *PloS one,* 12(4), pp. e0174866.