



# An application of generalised simulated annealing towards the simultaneous modelling and clustering of glaucoma

Mohd Zairul Mazwan Bin Jilani<sup>1</sup>  · Allan Tucker<sup>1</sup> · Stephen Swift<sup>1</sup>

Received: 1 August 2017 / Revised: 10 August 2018 / Accepted: 25 April 2019 / Published online: 14 May 2019  
© The Author(s) 2019

## Abstract

Optimisation methods are widely used in complex data analysis, and as such, there is a need to develop techniques that can explore huge search spaces in an efficient and effective manner. Generalised simulated annealing is a continuous optimisation method which is an advanced version of the commonly used simulated annealing technique. The method is designed to search for the global optimum solution and avoid being trapped in local optima. This paper presents an application of a specially adapted generalised simulated annealing algorithm applied to a discrete problem, namely simultaneous modelling and clustering of visual field data. Visual field data is commonly used in managing glaucoma, a disease which is the second largest cause of blindness in the developing world. The simultaneous modelling and clustering is a model based clustering technique aimed at finding the best grouping of visual field data based upon prediction accuracy. The results using our tailored optimisation method show improvements in prediction accuracy and our proposed method appears to have an efficient search in terms of convergence point compared to traditional techniques. Our method is also tested on synthetic data and the results verify that generalised simulated annealing locates the optimal clusters efficiently as well as improving prediction accuracy.

**Keywords** Generalised simulated annealing · Visual field · Glaucoma · Optimisation

---

✉ Mohd Zairul Mazwan Bin Jilani  
1338913@alumni.brunel.ac.uk

Allan Tucker  
allan.tucker@brunel.ac.uk

Stephen Swift  
stephen.swift@brunel.ac.uk

<sup>1</sup> Brunel University London, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK

## 1 Introduction

In recent years data has been collected at a vastly increasing rate. This data is typically complex in nature and high in both dimensionality and size. To cope with this data explosion there is a need for efficient algorithms that can effectively “keep up”.

Generalised simulated annealing (GSA) is an improved version of the simulated annealing (SA) algorithm, proposed by Tsallis and Stariolo (1996) and Xiang et al. (2017). 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 et al. 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 (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 the CSA since the jumps are frequently local, but can occasionally be quite long. The cooling of temperature in the method is much faster which makes the search more efficient.

Later in 1988, a generalisation of 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 (Andricioaei and Straub 1996; Brooks and Verdini 1988; dos R Correia et al. 2005; Sutter et al. 1995; Xiang et al. 2017) have often 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 (2000) have shown that the GSA algorithms are relatively efficient in Thomsons 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. Another recent study by Menin and Bauch (2017) have used GSA in finding the source of an outbreak (patient zero). They found that GSA is able to identify the location and time of infection of patient zero with good accuracy. Meanwhile Feng et al. (2018) presented a case study of dynamic land

using cellular automata simulation with GSA. Their GSA algorithm in the problem took the least time in the optimisation process compared to other heuristics search methods (particle swarm optimisation and genetic algorithm).

The positive findings from the previous studies on GSA have inspired this study to further investigate the compatibility of GSA algorithm with SMC to efficiently (with high accuracy and fast convergence) cluster and classify high dimensional data such as visual fields. Our work is an extension of the work done by Jilani et al. (2016) where SMC was experimented on visual field data using simple optimisation methods such as hill climbing to predict glaucoma progression. Since there is very little work has been done in this problem, taking advantage of GSA that has been proven an efficient method in literature, it is an open research opportunity.

We note that most of the studies have applied GSA towards solving continuous problems (Advanced Glaucoma Intervention Study Investigators 1994); hence within this paper we present a GSA algorithm for discrete optimisation. Our hypothesis underpinning this research is that the GSA method finds the optimum solution more efficiently than SA. Efficiency is measured in two ways, firstly we use prediction accuracy and secondly we look at algorithmic convergence, i.e. the speed an algorithm locates a high quality solution.

This paper is organised into the following sections: Sect. 2 details the background of the data used in this study. Section 3 describes the application of GSA algorithm towards simultaneous modelling and clustering. In Sect. 4, details regarding the experimental setup are discussed. Section 5 describes the results of our experiments and results are tabulated for comparison. The results are discussed in Sect. 6 and conclusion and future work are presented in Sect. 7.

## 2 Background

### 2.1 Glaucoma

Glaucoma is the second leading cause of blindness disease in the world (Quigley and Broman 2006) and is an irreversible disease that can seriously affect a sufferers quality of life (McKean-Cowdin et al. 2007). The disease is typically caused by a build-up of intraocular pressure in the eye that causes damage to the optic nerve, which delivers image signals to the brain (Quigley 2011). In clinical practice, glaucoma is commonly managed by performing a visual field test which captures the sensitivity of the eye to light (Kristie Draskovic 2016; Westcott et al. 1997).

Much research has been done to study glaucoma deterioration such as statistical analysis and data modelling using visual field data as presented by Pavlidis et al. (2013), Swift and Liu (2002) and Ceccon et al. (2014). However, there is less work on analysing and modelling visual field data using significant clustering arrangements of the data.

Since there is no gold standard established in analysing glaucoma data (Schulzer et al. 1994), applying machine learning techniques to find patterns of glaucoma deterioration in visual field data helps medical practitioners to provide appropriate treatments, i.e. the prediction of potential future visual field loss (Swift and Liu 2002; Tucker and

Garway-Heath 2010). Additionally, this piece of work is a novel approach of high-dimensional data analysis in which a continuous optimisation method is applied in a discrete problem finding the best clustering arrangement of visual field locations.

## 2.2 Visual field data

Visual field data is commonly used in glaucoma management. It is collected via a non-invasive clinical procedure by means of computerised automated perimetry (Heijl and Patella 2002). Typically 54 or 76 points (including the blind spot) are tested to determine a patient's eye sensitivity to light (Fig. 1). A diagnosis of glaucoma using VF data is often quantified by the Advance Glaucoma Intervention Studies (AGIS) score (Advanced Glaucoma Intervention Study Investigators 1994). This metric is used to classify the severity of a glaucoma patient's condition. In this study we aim to predict future values of this AGIS metric from previously recorded VF data observations, i.e. VF data recorded at a time  $T$  is used to predict the AGIS score of the next visit ( $T + 1$ ) (Advanced Glaucoma Intervention Study Investigators 1994; Sacchi et al. 2013). In addition to envisage the level of glaucoma deterioration in the next VF test, predicting AGIS with high accuracy prediction could equip physician to prepare an appropriate treatment for patient. Table 1 shows the AGIS score used in this study that have been classified into 3 categories for an efficient classification (Jilani et al. 2016; Sacchi et al. 2013). In clinical practices, the AGIS score is computed to determine patient glaucoma stage and most analyses in research use six nerve fibre bundle (6NFB), which was established by Garway-Heath et al. (2000), for data modelling and prediction.

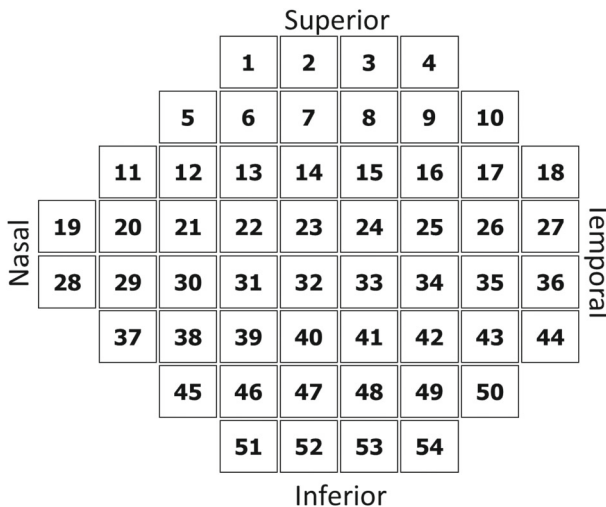


Fig. 1 The 54 locations of the visual field test for the right eye

**Table 1** AGIS score classification

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	

### 2.3 Simultaneous modelling and clustering of visual field data

Data clustering (Ghahramani 2004) techniques commonly use the data point structure to group data based on similarity such as distance between object in data. However, clustering technique such as those introduced by Banfield and Raftery (1993) have introduced methods based on modelling the data. Datasets are clustered based on object shape and structure rather than on proximity between data points. Likewise, the introduced approach that is SMC a model-based clustering method that finding clusters within data based on the best modelling result using an optimisation search method. In this paper presents the extension on the approach using GSA method in VF data. The SMC approach was used by Jilani et al. (2016) predicting the stages of glaucoma progression using VF data. The best clusters, which represent the collections of visual field locations, are searched in the data to improve prediction of glaucomatous progression as well as understanding the clusters arrangement. The motivation behind SMC is that a search for the best clustering arrangement is made, where the worth of such an arrangement is based on using the clusters to create compound variables that in turn can be used to classify accurately and predict the AGIS score. The clusters discovered in the search are hypothesised to be better than the six nerve fibre bundles (6NFB) in terms of prediction accuracy. These nerve fibre bundles are effectively a clustering arrangement that has been clinically determined through observing the physiological layout of the retina and optic nerves. Currently there is no conclusive evidence that this layout provides the best predictive accuracy, although there is some work on modelling the data using them (Sacchi et al. 2013).

## 3 Method

### 3.1 Data pre-processing

The VF data consists of 13,739 records of 1580 patients obtained from Moorfields Eye Hospital London (United Kingdom). The AGIS score was provided with the data and the data were prepared into a time series record which each record has the next visit AGIS ( $T + 1$ ) score. As such, the cleaning dataset after excluded the most recent record and consists of 12,159 records of VF data<sup>1</sup>.

<sup>1</sup> Ethical approval was obtained and the records were completely anonymous.

### 3.2 Clustering

Clustering, which is also known as unsupervised learning (Duda et al. 2000), is a data analysis technique for finding similarity amongst objects and grouping similar objects into the same cluster. The objects in the cluster share the common feature, function, shape, and attribute. In classical data clustering, the distance between objects (e.g. Euclidean) is commonly used (Jain and Dubes 1988). However, this study applies model-based technique to cluster data based on classification performance.

### 3.3 Data modelling

Data modelling is a key process involved in SMC. This process is also known as supervised learning where the data are classified to a set of categories it belongs. In this study, VF data is modelled by classifying the target variable (AGIS) which indicates the stage of glaucoma disease. The VF data is classified into  $T + 1$  of AGIS which is the time series record. Since the target variable value is  $T + 1$  of the AGIS, the classification is a prediction of the patient glaucoma progression. The Naive Bayes Multinomial Updateable (NBU) classifier is used in the modelling process owing to the most efficient method in terms of runtime and has good accuracy when applied to a number of different problem domains (Jilani et al. 2016; Rennie et al. 2003; Sundar 2013; Tao and Wei-hua 2010).

### 3.4 Simultaneous modelling and clustering

The approach used in the SMC comprises of these two main analysis techniques that is clustering and modelling. The 52 VF locations are grouped into clusters randomly then the clusters are used for data modelling as the validation to the clusters. The clusters validation is measured as prediction accuracy. The SMC process uses optimisation methods searching the best cluster with prediction accuracy as the fitness function. The hypothesis underpinning this approach is that the higher prediction accuracy, the better the quality of the associated clustering arrangement. This heuristic process iterates for suitable number of iterations and searches for the best clusters using an optimisation method. The final clusters arrangements (resulting bundles) of the VF locations are captured. This SMC process is schematically illustrated in Fig. 2.

The final clusters obtained in the search using SMC are recorded and analysed using the Weighted Kappa statistic. The Weighted Kappa statistic (Table 2) (Viera and Garrett 2005) between the clustering arrangement of SMC and the 6NFB (Garway-Heath et al. 2000) (depicted in Fig. 3) is computed to measure the similarity of resulting clusters with the 6NFB. The resulting clusters are also mapped on the VF to visualise the association between VF location and bundles in order to comprehend the glaucomatous progression. By looking at the bundles of VF points discovered by the method, patterns of glaucoma progression can be understood.

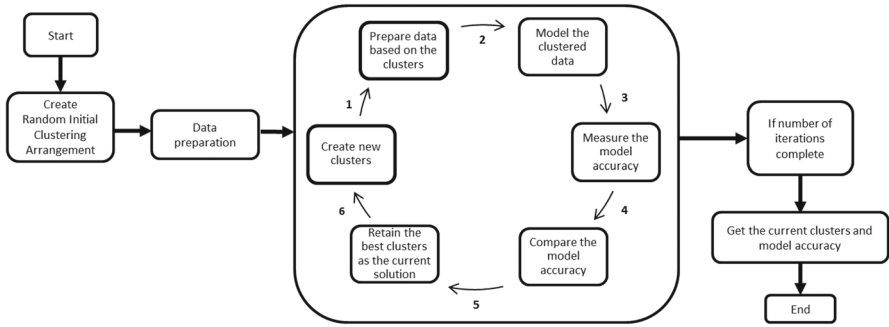


Fig. 2 High level process of simultaneous modelling and clustering

Table 2 Weighted kappa statistic metric

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

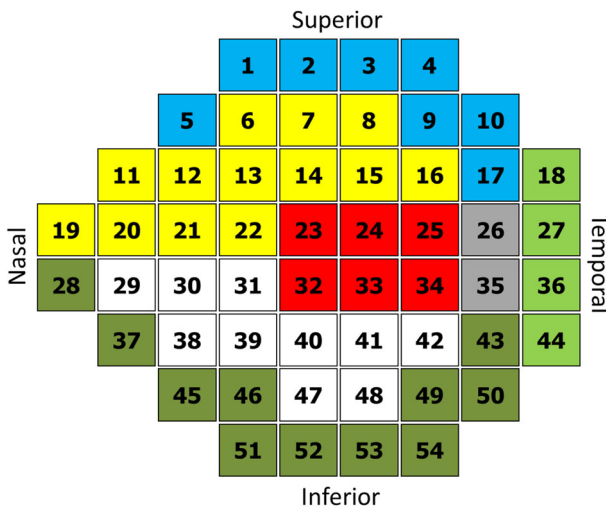


Fig. 3 The 54 locations of the visual field data mapped to the six nerve fibre bundles

### 3.5 Generalised simulated annealing algorithm

The GSA method is distinct from SA in determining the temperature, the acceptance probability value and the selection of a neighbouring point in the search space. These parameter values regulate the methods behaviour such as convergence and cooling

rate. The acceptance probability (Eq. 1) and artificial temperature (Eq. 2) use an acceptance index ( $q_a$ ) and visiting index ( $q_v$ ) as parameters (Tsallis and Stariolo 1996). Furthermore, the visiting index ( $q_v$ ) is used in the visiting distribution equation (Eq. 4) (Tsallis and Stariolo 1996) to determine the size of change (in this study this is the number of VF locations to be re-arranged between clusters) for the next potential solution in the GSA search. The size of change (small change) in Algorithm 1 is defined as a number of moves, i.e. exchanging two variables from two different clusters. Each new better (in terms of fitness function) solution found in the search is accepted. However, if the new solution is worse, a criterion of acceptance the solution is derived by computing the acceptance probability according to Eq. 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 a number of iterations are complete. The detailed procedure of the GSA method is described in Algorithm 1.

$$p_{q^A}(x_t \rightarrow x_{t+1}) = \frac{1}{[1 + (q_A - 1)(E(x_{t+1}) - E(x_t))/T_{q_A}^A]^{1/(q_A-1)}} \tag{1}$$

where  $x_t$  is the current solution (VF clusters) of the current iteration, while  $x_{t+1}$  is the new solution for next iteration.  $E(x_t)$  is defined as the fitness value of solution  $x$ .  $T_{q_a}^A$  is the acceptance temperature at time  $t$  that decreasing from initial value to zero.

$$T_{q_v}^V(t) = T_{q_v}(1) \frac{2^{q_v-1} - 1}{(1 + t)^{q_v-1} - 1} \tag{2}$$

where  $T_{q_v}^V$  is the visiting temperature at time  $t$ .

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

As suggested by Advanced Glaucoma Intervention Study Investigators (1994)  $D$  is set to 1.

$$g_{q_v}(\Delta x_t) = \left(\frac{q_v - 1}{\pi}\right)^{D/2} \frac{\Gamma\left(\frac{1}{q_v-1} + \frac{D-1}{2}\right)}{\Gamma\left(\frac{1}{q_v-1} - \frac{1}{2}\right)} \times a \tag{4}$$

According to Tsallis and Stariolo (1996), parameter  $q_a$  and  $q_v$  are calibrated depending upon the nature of the problem where the best value range for  $q_v$  is between 1.66 to 2.70. However, they also presented that the GSA was found having faster convergence when  $q_a < 1$ . Therefore in our study, the region of exploration for the  $q_a$  variable is between 0 and 1. Our new visit value ( $q_v$ ) is represented by the number of VF locations (totalling 52 locations) to be re-arranged between the VF bundles in the next iteration of the search.



---

**Algorithm 1:** Simultaneous Modelling and Clustering Pseudo Code of Generalised Simulated Annealing
 

---

**Input** : Raw data of visual field  
 iterations = Number of iterations  
 fd = Number of fold cross validation  
 temp = initial temperature  
 $q_a$  = acceptance index  
 $q_v$  = visiting index  
 Model= NaiveBayesUpdatable

- 1 Let bundles = a random bundles of visual field points
- 2 Let data = visual field data of the bundles
- 3 Let fitness = prediction accuracy of the current bundles with fd
- 4 **for**  $i=0$  to iterations-1 **do**
- 5     Calculate smallChange (Eq. 4) new bundles = re-arrange bundles for n smallChange
- 6     Prepare data = visual field data of the new bundles
- 7     new fitness = prediction accuracy of data classification with fd
- 8     **if** new fitness > fitness **then**
- 9         fitness = new fitness
- 10        bundles = new bundles
- 11     **else**
- 12         different fitness = new fitness-fitness
- 13         Calculate probability acceptance (Eq. 1)
- 14         Let random = a uniformly distributed real random number between 0.0 and 1.0 inclusive
- 15         **if** probability > random **then**
- 16             fitness = new fitness
- 17             bundles = new bundles
- 18         **else**
- 19             **end**
- 20     **end**
- 21     Calculate temperature (Eq. 2)
- 22 **end**

**Output:** Visual field bundles, prediction accuracy

---

## 4 Experiment setup

### 4.1 GSA parameter exploration

Many studies have devised GSA algorithms and calibrate the parameters in the GSA to suit the nature of the problem at hand. 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. Whereas in this paper we use 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 e.g. fitness value is between 0 to 1, equations are manipulated using Newton Raphson method to get the  $q_v$ . As such, Eqs. 1 and 2 are simplified to derive  $T_0$ , and thus  $q_v$  is obtained.

### 4.2 Newton Raphson

The Newton Raphson technique (Akram and 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.

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 (Equation 1) and temperature equation (Eq. 2) to derive the parameters values. We define the acceptance probability at iteration zero as

$$P_0 = \frac{1}{[1 + (q_a - 1)(\Delta f)/T_0]^{1/q_a - 1}} \tag{5}$$

In order to get the initial temperature ( $T_0$ ) from this equation, some assumptions have been made. We assume that the acceptance probability is 0.4 at iteration 1 in the search. This is possible when the worst random solution obtain 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 Eq. 5 is simplified as follows,

---

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

---

$$T(N) = T_0 \left[ \frac{2q_v - 1}{(1 + N)^{q_v - 1} - 1} \right] \tag{7}$$

where  $s = q_v - 1$ ,  $N$  is number of iteration

$$T_N((1 + N)^s - 1) = T_0(2^s - 1)$$

$$T_N(1 + N)^s - T_N = T_0 2^s - T_0$$

$$T_N(1 + N)^s - T_0 2^s = T_N - T_0$$

$$\frac{T_N(1 + N)^s}{T_N - T_0} - \frac{T_0 2^s}{T_N - T_0} = 1 \tag{8}$$

Within Eq. 11, the  $T_0$  value is obtained from Eq. 9 by selecting the  $q_a$  (values ranged between  $-0.5$  to  $1.6$ ). With  $T_0$  and  $N$  (number of iterations,  $100,000$ ) in hand, the Eq. 11 is used to derive the correspondent value of  $q_v$  where  $T_N$  is the final temperature set as  $0.001$ . Equation 11 is solved by means the Newton Raphson method. We know that  $ab^x + cd^x = 1$  is non-linear. Therefore, Eq. 11 is written as  $a(1 + N)^s + c2^s = 1$ , where  $a = \frac{T_n}{T_n - T_0}$  and  $\frac{-T_0}{T_n - T_0}$ . Thus, the non-linear function of Eq. 11 is,

$$f(x) = a(1 + N)^s + c2^s - 1 \tag{9}$$

$$f(x) = ab^x + cd^x - 1 \tag{10}$$

and the derivative of the function  $f(x)$  is,

$$f'(x) = ab^x \ln(b) + cd^x \ln(d) \tag{11}$$

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  $q_v$ .

### 4.3 Experiment strategy

The experiments were run on both VF and synthetic data [multivariate normal (MVN) generated data] which consists of  $12,159$  and  $1500$  records respectively. However, in order to avoid bias in the VF experiments, the data is resampled every iteration resulting in  $1580$  records.

A synthetic dataset was constructed as a verification tool for the SMC approach. Three samples ( $2500$  in length) are generated from three multivariate normal distributions ( $15$  variables each). 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  $2500$  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. For example, if the first  $15$  variables of a given row have a larger average than the second or third set of  $15$  variables, then the class label is “1”. That is, if the first  $15$  variables of a given row have a larger average than the second or third set of  $15$  variables, then the class label is “1”, if the second  $15$  variable average is the largest then the class label is “2”, for the third  $15$  variables “3” (as depicted in Figure 4). The process (data generation) is repeated until the distribution of the three classes are approximately equal. The expectation is that the SMC approach will be able to reverse engineer the original structure of the underlying data generation process, i.e. three clusters of  $15$  variables each.

The GSA and SA algorithms were used in the experiments presented in this paper. The experiments were modelled using three modelling strategies that is  $10$  folds cross validation ( $10FCV$ ) (Zhang 1993),  $2$  folds cross validation ( $2FCV$ ) with  $10$  repeats and no fold cross validation ( $NoCV$ ). The classifier used in modelling is NBU. Each of the experiment strategies was run for  $25$  times. Statistical values such as mean, maximum, minimum of the prediction accuracy and convergence point of the search are observed.

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

Fig. 4 Visualisation of synthetic dataset for defining target variable class

These results are tabulated and compared between the optimisation methods and the modelling strategies.

### 4.4 Simulated annealing

The SA algorithm has been selected as a comparison method with GSA since previous work (Jilani et al. 2016; Vincent et al. 2017) have shown that SA is a very competitive search method and can outperform a number of standard methods (random mutation hill climbing and random restart hill climbing). We have also used the same strategy for determining the SA algorithms parameters (initial temperature and cooling rate) from previous work (Swift et al. 2004). Within our formal experiment on SA, the initial temperature (Eq. 15) obtained from a simulation in which the summation of different fitness function evaluations (Eq. 13) is computed from a number of iterations. The simulation was run for 5000 iterations and upon obtaining the initial temperature, the formal experiment of SA resumes for another 95,000 iterations to complete 100,000 iterations as per the GSA experiment. The cooling rate of SA is then computed using Eq. 14.

$$T_0 = \frac{f}{n} \tag{12}$$

where n is 5% from the formal experiment of SA.

$$f = \sum_{i=1}^n \Delta f_i = \sum_{i=1}^n |f' - f| \tag{13}$$

$$T_{iter} = \lambda^{iter} T_0 \tag{14}$$

#### 4.5 Generalised simulated annealing parameters

In the GSA optimisation method, preliminary experiments were conducted to find the suitable value for the  $q_a$  and  $q_v$  index. Experiments were run for 10,000 iterations (10 repeats) on both VF and synthetic data with 10-fold cross validation. Values range between 0.1 and 0.5 (Tsallis and Stariolo 1996) were applied in exploration experiments to determine the  $q_a$ . Tables 3 and 4 show the corresponding fitness value (predictive accuracy) and convergence point of the search to the  $q_a$  for VF and synthetic data respectively.

Our hypothesis is that the GSA method will converge earlier than SA, thus as tabulated in Table 3 and 4,  $q_a$  values with the minimum averaged convergence were chosen for GSA experiments. Therefore,  $q_a = 0.010$  and  $q_a = 0.009$  were used in our experiments for VF data and synthetic data respectively. Thereafter, the correspondent values of  $q_v$  for VF data and synthetic data were calculated using the Newton Raphson.

**Table 3** VF data parameter fitness value

QA	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	5521.4
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	87.45	10000.9

**Table 4** Synthetic data parameter fitness value

QA	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

## 5 Results

This section presents initial experiments and the main result of this work. The main results include the model prediction accuracy, Weighted Kappa statistic (WK), convergence point and resultant clusters. Statistical values of the results for each modelling strategy are tabulated for comparison.

### 5.1 Initial experiment

Initial experiments (25 samples) were conducted to evaluate the classification performance on visual field data using 6NFB clustering arrangement. In this set of experiment, the data are prepared in the 6NFB clustering arrangement for classification. Additionally, another experiment (25 samples) was conducted with the same purpose, which the data were clustered first using K-Means before classification. Both experiments used Multinomial Naive Bayes Updatable classifier. The results from these experiments are obtained to be the benchmark for our study. The results are presented in following sub-sections:

**Table 5** Prediction accuracy of visual field data using 6NFB (sampled data)

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** Prediction accuracy of visual field data using K-Means resultants clusters (sampled data)

Modelling strategy/result	10FCV Accuracy (%)	2FCV Accuracy (%)	NoFCV Accuracy (%)
Minimum	82.49	82.91	83.17
Maximum	85.89	85.45	86.27
Average	84.30	84.22	84.47

*6NFB experiment* From Table 5, the average predictive accuracies of visual field data using the 6NFB are 83.27%, 83.90% and 83.65% for 10FCV, 2FCV and NoFCV respectively. On average result, 2FCV is slightly higher than the other two modelling strategies. However, the experiment recorded maximum accuracy of 85.31% with 10FCV.

*K-Means clustering experiment* Meanwhile, predictive accuracies from K-Means clustering arrangement experiment are consistently higher (slightly) than the 6NFB. On average, the predictive accuracies are 84.30%—10FCV, 84.22%—2FCV and 84.47%—NoFCV as shown in Table 6. The NoFCV modelling strategy recorded the highest predictive accuracy in the experiment with 86.27%.

## 5.2 Predictive accuracy: visual field data

Both methods predict the VF data better using the 10FCV strategy with 88.48 % (Table 7) and 88.54% (Table 8) than 2FCV. The results of VF experiments show that the GSA method improves the prediction accuracy better than SA in all modelling strategies where the highest accuracy is 87.89% (average) in 10-fold cross validation (Table 8). From the result we note that the average accuracy of 2FCV 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 with 10 repeats (0.004 - Table 8). We compute the confidence intervals (95%) of the accuracy in these 25 experiments, Tables 9 and Table 10 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 88.002% and 87.765%.

**Table 7** Prediction accuracy of SA for VF data

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 8** Prediction accuracy of GSA for VF 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 9** Confidence interval of SA for VF data

Modelling strategy /confidence interval	10FCV	2FCV	NoCV
Upper limit	87.924	86.556	86.742
Lower limit	87.591	86.341	86.523

**Table 10** Confidence interval of GSA for VF data

Modelling strategy/confidence interval	10FCV	2FCV	NoCV
Upper limit	88.002	86.737	86.788
Lower limit	87.765	86.530	86.518

### 5.3 Predictive accuracy: synthetic data

Tables 11 and 12 show the prediction accuracy of the synthetic data by modelling strategy for the SA and GSA method respectively. Prediction accuracy results in the synthetic data experiments were found higher in both methods and modelling strategies compared to visual field data (98.55%—SA and 98.59%—GSA). The synthetic

**Table 11** Prediction accuracy of SA for synthetic data

Modelling strategy/result	10FCV Accuracy (%) (WK)	2FCV Accuracy (%) (WK)	NoFCV 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 12** Prediction accuracy of GSA for synthetic data

Modelling strategy/result	10FCV Accuracy (%) (WK)	2FCV Accuracy (%) (WK)	NoFCV 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 13** Confidence interval of SA for synthetic data

Modelling strategy/confidence interval	10FCV	2FCV	NoCV
Upper limit	98.400	97.0475	98.594
Lower limit	98.315	96.994	98.513

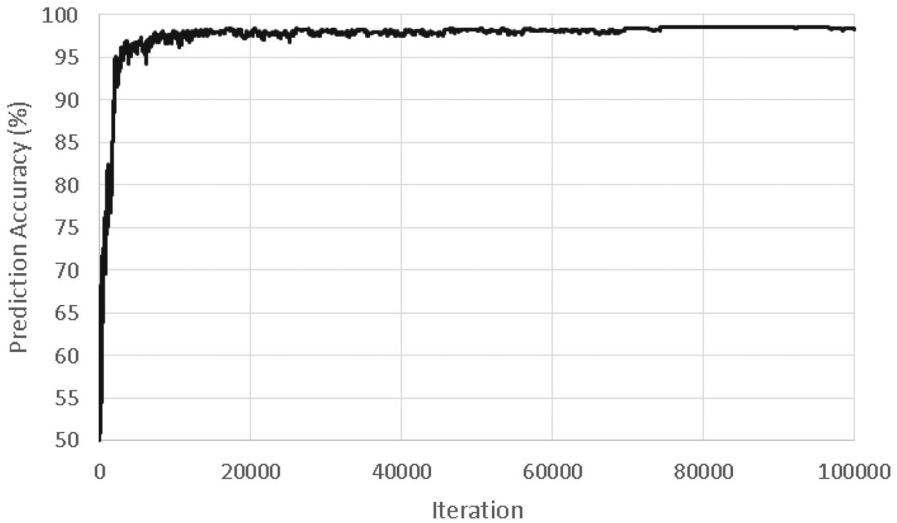
**Table 14** Confidence interval of GSA for synthetic data

Modelling strategy/confidence interval	10FCV	2FCV	NoCV
Upper limit	98.393	97.051	98.609
Lower limit	98.259	97.023	98.571

results do not show any pattern in prediction accuracy in both methods and modelling strategies. However, the SA 10FCV outperforms the GSA 10FCV. Whilst with 2FCV and NoFCV, the GSA method accuracies are higher than the SA method. The significant difference between these 2 methods on synthetic data can be seen in the WK values. It can be concluded that the GSA recorded almost perfect WK (near 1.0) which agrees with the expected synthetic clusters of the 45 variables. Table 13 (SA) and Table 14 (GSA) show the 95% confidence intervals of the accuracy for the synthetic dataset. Confidence interval in the synthetic dataset was found has higher range in NoCV strategy in both methods (SA and GSA).

#### 5.4 Convergence point

GSA methods are often described as having a very fast convergence in literature (Bohachevsky et al. 1986; Penna 1995; Tsallis and Stariolo 1996). Thus, the iteration point that the search has converged was captured. Since the fitness function of both datasets is noisy (as depicted in Fig. 5), we established a rule to determine the convergence point by calculating a noisy fitness tolerance. We define a noisy fitness function as one that returns different fitness values each time it is evaluated on the same solution. This will occur in this work due to both cross validation (different cross validation folds) and to sampling (a random pair of VF 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 (Ratray and Shapiro 1998). From the graph shown in Fig. 5, the search



**Fig. 5** Convergence point of the synthetic data with the GSA method (10FCV)

**Table 15** Standard deviation of the data

Method/modelling strategy	10FCV		2FCV 10 Repeats		No FCV	
	VF	Syn.	VF	Syn.	VF	Syn.
SA	0.77	0.59	0.18	0.23	n.a	n.a
GSA	0.79	0.54	0.23	0.20	n.a	n.a

has converged before reaching the 10,000th iteration, however within the convergence line, the fitness value (prediction accuracy) varies. This shows that with the GSA rule (acceptance probability) for searching for the best solution, poorer solutions with a certain degree of change are still being accepted. In order to cater for the inherent noisy nature of the fitness function an investigation into tolerance limits was conducted. We first ensure the fitness values in both data are normally distributed. We have shown that the fitness values are normal distributed (with  $p$ -value 0.49) and Table 15 shows the standard deviation of the data by methods and strategies.

From Table 15, we found that the standard deviation of the fitness within the 10-fold experiments is very high. 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. The noisy fitness tolerance limits were calculated with z-score value 1.98 (Weiss 2013) which equivalent to 97.61% of the data lies under the defined limits (Table 16). The noisy fitness tolerance limits tabulated in Table reftab:14 are used to determine whether the change of fitness value is a significant change or not. At any point of the search that has change in the fitness value ( $\Delta F$ ) within the limit are considered has no significant change. The different of fitness values is calculated using Eq. 15.

**Table 16** Noisy fitness tolerance

Method/modelling strategy	10FCV		2FCV 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 17** Average convergence point of VF data

Modelling strategy/method	10FCV	2FCV 10 repeats	NoFCV
SA	11,643.16	19,938.04	47,304
GSA	2839.88	34,889.04	73,928.08

$$\Delta F = F_{t+1} - F_t \quad (15)$$

Table 17 tabulates the average of convergence point of the individual experiments for visual field data. It shows that the GSA method has converged fastest in 10FCV that is in average at iteration 2839. This also can be shown in Fig. 6 for the 25 experiment repeats convergence point in comparison with the SA method. However, for 2FCV modelling strategy, the SA method is far outperformed the GSA method with average 19,938 (GSA: 34,889). The synthetic data convergence point results (Table 18) are found consistent with the VF data. The best convergence point for the synthetic data was the GSA method in 10FCV (7328).

To support these convergence point results, we run simulations (25 samples) on the algorithms for SA and GSA on visual field data. The simulations comprised of the same experiments property as the main experiments including 100,000 iterations and Naive Bayes Multinomial Updatable classifier with 10FCV. The simulations were conducted to capture the runtime (in s) for the algorithms. In these simulations, the total effort of the experiments (entire SMC) and the effort for classification were captured. Classification effort includes data preparation within the iteration loops of the experiment. The algorithm effort in terms of runtime is obtained by subtracting the total effort and classification effort. The reason being is that based on our observation, classification effort in SMC is not part of the algorithm (SA and GSA) process and classification effort is dependent with number of moves which is more complex with high number of moves. From Table 19, our simulation experiments reveal that algorithmic computation time is taken by the fitness function where the GSA algorithm only took 1.67% (meanwhile SA took 6.40%) from the total effort of SMC. This figure indicates that GSA is 4.74% more efficient than SA.

## 5.5 Resultant clusters

The clusters size of each experiment was captured to see the range of clusters size searched by the algorithms. We note that VF data experiments with the SA method tend to get lower clusters size when the number of fold cross validation is decreasing. There

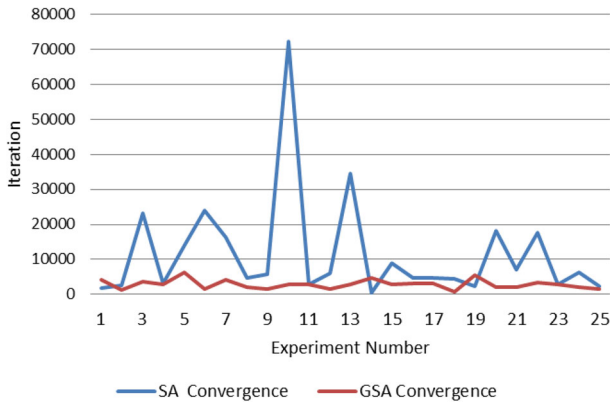


Fig. 6 Convergence point of VF data For 10FCV

Table 18 Average convergence point of synthetic data

Modelling strategy/method	10FCV	2FCV 10 repeats	NoFCV
SA	19,582.4	30,365.04	44,843.68
GSA	7328	45,244.56	71,007

Table 19 Algorithm effort runtime

Method	Total experiment runtime (s)	Classification effort runtime (s)	Algorithm effort (%)
SA	41,860.86	39,180.37	6.40
GSA	128,120.47	125,985.66	1.67

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 the GSA method has range between 3–12 (minimum–maximum) clusters size for both 10FCV and 2FCV with 10 repeats, and 3–14 clusters size 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 range clusters size are 4–6 (minimum–maximum), 3–4, and 4–7 in SA method for 10FCV, 2FCV with 10 repeats, and NoFCV respectively. Knowing that the correct cluster size for the synthetic data is 3 clusters, we found that the GSA method searches the clusters better than the SA method with the range of clusters size are 4–5, (minimum–maximum), 3–4, and 4–4 for 10FCV, 2FCV with 10 repeats, 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 Discussion

Empirical observation on the experiments found that GSA appears to be an effective algorithm to find good solutions (with high WK) in the synthetic data with 2FCV modelling strategy as far as the reverse-engineer is concerned. In terms of prediction accuracy (synthetic dataset), there was less significant difference (with only 0.04%) between SA and GSA, where NoFCV and 10FCV have better prediction than 2FCV. Having slightly higher accuracy in 10FCV and NoFCV can be due to a bias element and overfitting (Varma and Simon 2006).

However, in the real data experiments, WK statistics indicate poor agreement between the resultant clusters and the 6NFB. WK metric is not the main criteria to indicate a quality solution (clustering arrangement) because the 6NFB is just a common practice being used in clinical. Furthermore, the motivation of this research is to find any other clustering arrangements in visual field data that can highly predicts the AGIS score using an advanced optimisation method in SMC. Our finding is that 10FCV appears to be the best modelling strategy in SMC as this strategy consistently produced higher accuracy in both methods, where GSA slightly outperformed SA. This finding corresponds to Kohavi (1995).

Additionally, from our initial experiments results for K-Means and the 6NFB classification, SMC with GSA has improved 3.90% and 3.42% prediction accuracy (average results) from the 6NFB and K-Means respectively.

In terms of algorithm efficiency for SMC, GSA has proven in our experiments to be more efficient than SA as the convergence points are best recorded in GSA with 10FCV (both datasets). The GSA method was 8.80% faster than the SA in visual field data, while it was 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). Whilst, our analysis results on algorithms effort runtime have

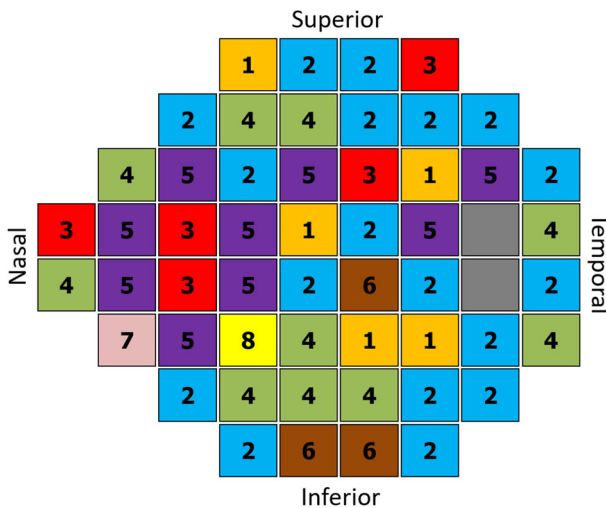


Fig. 7 Convergence point of VF data For 10FCV

shown that GSA only took 1.67% from the total effort of the SMC process (meanwhile SA used up 6.40% from the total SMCs runtime). This shows that GSA is 4.74% more efficient than SA in terms of SMCs runtime.

Apart from the analysis of the experiments results on prediction, convergence and resultant clusters size, we extended our analysis to visualise the resultant clusters by mapping the VF locations on the VF grid map. Mapping the VF locations on the grid is to comprehend the pattern of visual loss suggested by our algorithms. The highest prediction accuracy of the GSA was selected for this analysis. Therefore, the resultant clusters from the GSA method in 10FCV (88.54% accurate with 8 clusters) are visualised in the 54 locations VF grid map (Fig. 7). From the visualisation of the clusters, we found that the larger clusters size appears on the periphery of vision. This can be seen from the Fig. 7 that clusters 2 and 4 are on the periphery of the VF 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.

## 7 Conclusion

In this paper we have presented SMC that has been tailored with the GSA algorithm to solve discrete optimisation problems. The method is arguably more efficient in solving problems that have multiple local optima, in that it is less likely to get trapped near a sub-optimal point in the search space (Tsallis and Stariolo 1996). Our proposed algorithm in SMC for the problem has shown improvement in predictive accuracy by 3.90% and 3.42% compared to the 6NFB and K-Means respectively. Furthermore, GSA appears to be compatible with SMC in terms of efficacy where our evaluation on convergence points have shown significantly faster (with 10FCV) than SA. Also, we have discovered that GSA uses little effort in the SMC process compared to SA (evaluated in algorithms runtime).

The visiting distribution ( $q_v$ ) in the GSA method makes it distinct from the other annealing methods. This parameter is responsible for controlling how the search space is explored, allowing the method to locate high quality clustering arrangements of the VF data in an efficient manner. The findings conclude that the GSA algorithm outperforms the SA method in terms of convergence point, which the algorithm effectively locates the best clusters of VF as well as improving the predictive accuracy and an efficient method in the search.

The effectiveness of the algorithm is also proven in our synthetic data experiments where the correct clusters underlying the data were found in 24 out of the 25 experiments (with a WK of 1.0) with 2FCV modelling strategy. Furthermore, the WK results of the other 2 strategies (10FCV and NoFCV) of GSA also outperformed SA in the synthetic data.

Based on the positive finding from our study, a number of research opportunities can be considered in future work. Application of SMC using an advanced optimisation method and datasets manipulation can be further explored to reduce noisy fitness found in this work. This includes applying an operator for solution selection in the algorithm that could possibly reduce noisy fitness and by segmenting the data according to

early, moderate and advance glaucoma. Secondly more extensive experiments are needed to explore the visiting index and probability index of the GSA algorithm. We note that the exploration of these parameters could be delegated to an automatic configuration algorithm, for example Lopez-Ibez et al. (2016) and Hoos (2012). The range of parameters for GSA can be established for a certain domain of problems based on a dataset's nature and the choice of fitness function as a reference. Furthermore, improving algorithmic performance and developing a tool for clinicians would help in providing better treatments for glaucoma patients and help in understanding the nature of visual deterioration of the disease.

**Acknowledgements** This work is in part funded by a Ph.D. bursary from the Majlis Amanah Rakyat (Malaysian government). We would also like to thank Moorefields Eye Hospital London, UK for granting the use of the visual field data used in this project.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Advanced Glaucoma Intervention Study Investigators: Advanced glaucoma intervention study: 2. visual field test scoring and reliability. *Ophthalmology* **10**(8), 1445–1455 (1994). [https://doi.org/10.1016/S0161-6420\(94\)31171-7](https://doi.org/10.1016/S0161-6420(94)31171-7)
- Agostini, F.P., SoaresPinto, D.D.O., Moret, M.A., Osthoff, C., Pascutti, P.G.: Generalized simulated annealing applied to protein folding studies. *J. Comput. Chem.* **27**(11), 1142–1155 (2006)
- Akram, S., Ann, Q.U.: Newton raphson method. *Int. J. Sci. Eng. Res.* **6**(7), 1748–1752 (2015)
- Andricioaei, I., Straub, J.E.: Generalized simulated annealing algorithms using Tsallis statistics: application to conformational optimization of a tetrapeptide. *Phys. Rev. Ser. E* **53**, R3055–R3055 (1996)
- Banfield, J., Raftery, A.: Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**(3), 803–821 (1993). <https://doi.org/10.2307/2532201>
- Bohachevsky, I.O., Johnson, M.E., Stein, M.L.: Generalized simulated annealing for function optimization. *Technometrics* **28**(3), 209–217 (1986). <http://www.jstor.org/stable/1269076>
- Brooks, D.G., Verdini, W.A.: Computational experience with generalized simulated annealing over continuous variables. *Am. J. Math. Manag. Sci.* **8**(3–4), 425–449 (1988)
- Ceccon, S., Garway-Heath, D.F., Crabb, D.P., Tucker, A.: Exploring early glaucoma and the visual field test: classification and clustering using bayesian networks. *J. Biomed. Health Inform.* **18**(3), 1008–1014 (2014)
- Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern classification*, 2nd edn. Wiley, Chichester, New York (2000)
- dos R Correia, E., Nascimento, V.B., de Castilho, C.M., Esperidiao, A.S., Soares, E.A., de Carvalho, V.E.: The generalized simulated annealing algorithm in the low energy electron diffraction search problem. *J. Phys. Condens. Matter* **17**(1), 1 (2005)
- Feng, Y., Liu, Y., Tong, X.: Comparison of metaheuristic cellular automata models: a case study of dynamic land use simulation in the Yangtze River Delta. *Comput. Environ. Urban Syst.* **70**, 138–150 (2018)
- Garway-Heath, D.F., Poinoosawmy, D., Fitzke, F.W., Hitchings, R.A.: Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* **107**(10), 1809–1815 (2000)
- Ghahramani, Z.: *Unsupervised Learning*. *Advanced Lectures on Machine Learning*, pp. 72–112. Springer, Berlin (2004)
- Heijl, A., Patella, V.M.: *Essential Perimetry: The Field Analyzer Primer*. Carl Zeiss Meditec, Jena (2002)
- Hoos, H.H.: Programming by optimization. *Commun. ACM* **55**(2), 70–80 (2012)
- Jain, A.K., Dubes, R.C.: *Algorithms for Clustering Data*. Prentice-Hall, Inc, Upper Saddle River (1988)

- Jilani, M.Z.M.B., Tucker, A., Swift, S.: Simultaneous modelling and clustering of visual field data. In: 2016 IEEE 29th international symposium on computer-based medical systems (CBMS), pp 213–218 (2016). <https://doi.org/10.1109/CBMS.2016.66>. ID: 1
- Kelley, C.T.: Solving Nonlinear Equations with Newton's Method, vol. 1. SIAM, Philadelphia (2003)
- Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Ijcai*, vol. 14, pp. 1137–1145. Stanford, CA (1995)
- Kristie Draskovic, J.J.M.: Automated perimetry: visual field deficits in glaucoma and beyond. <https://www.reviewofoptometry.com/ce/automated-perimetry-visual-field-deficits-in-glaucoma-and-beyond>. Accessed 17 Jan 2016
- Lpez-Ibez, M., Dubois-Lacoste, J., Cceres, L.P., Birattari, M., Stltze, T.: The irace package: iterated racing for automatic algorithm configuration. *Oper. Res. Perspect.* **3**, 43–58 (2016)
- McKean-Cowdin, R., Varma, R., Wu, J., Hays, R.D., Azen, S.P., Group, L.A.L.E.S.: Severity of visual field loss and health-related quality of life. *Am. J. Ophthalmol.* **143**(6), 1013–1023 (2007)
- Menin, O.H., Bauch, C.T.: Solving the patient zero inverse problem by using generalized simulated annealing. *Phys. A Stat. Mech. Appl.* **490**, 1513–1521 (2018)
- Mojica, O., Bassrei, A.: Application of the generalized simulated annealing algorithm to the solution of 2d gravity inversion of basement relief. In: 3rd Latin American Geosciences Student Conference (2015)
- Moret, M.A., Pascutti, P.G., Bisch, P.M., Mundim, K.C.: Stochastic molecular optimization using generalized simulated annealing. *J. Comput. Chem.* **19**(6), 647–657 (1998)
- Pavlidis, S., Swift, S., Tucker, A., Counsell, S.: The modelling of glaucoma progression through the use of cellular automata. In: *International Symposium on Intelligent Data Analysis*, pp. 322–332. Springer (2013)
- Penna, T.J.: Traveling salesman problem and Tsallis statistics. *Phys. Rev. E* **51**, 1 (1995)
- Quigley, H.A.: Glaucoma. *The Lancet* **377**(9774), 1367–1377 (2011). [https://doi.org/10.1016/S0140-6736\(10\)61423-7](https://doi.org/10.1016/S0140-6736(10)61423-7)
- Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. *The British journal of ophthalmology* **90**(3), 262–267 (2006). DOI 90/3/262 [pii]. LR: 20130905; JID: 0421041; CIN: Br J Ophthalmol. 2006 Mar;90(3):253-4. PMID: 16488934; OID: NLM: PMC1856963; ppublish
- Rattray, M., Shapiro, J.: Noisy fitness evaluation in genetic algorithms and the dynamics of learning. In: *FOGA*, pp. 117–139 (1996)
- Rennie, J.D., Shih, L., Teevan, J., Karger, D.R.: Tackling the poor assumptions of naive bayes text classifiers. In: *ICML*, vol. 3, pp. 616–623. Washington, DC (2003)
- Sacchi, L., Tucker, A., Counsell, S., Garway-Heath, D., Swift, S.: Improving predictive models of glaucoma severity by incorporating quality indicators. *Artif. Intell. Med.* **60**(2), 103–112 (2014)
- Schulzer, M., Airaksinen, P.J., Alward, W.L., Amyot, M., Anderson, D.R., Balazzi, G., Blondeau, P., Cashwell, L.F., Cohen, J.S., Desjardins, D.: Errors in the diagnosis of visual field progression in normal-tension glaucoma. *Ophthalmology* **101**(9), 1589–1595 (1994)
- Selman, B., Gomes, C.P.: Hill-climbing search. *Encycl. Cogn. Sci.* **81**, 82 (2006)
- Sundar, P.P.: A comparative study for predicting students academic performance using Bayesian network classifiers. *IOSR J. Eng.* **3**(2), 37–42 (2013)
- Sutter, J.M., Dixon, S.L., Jurs, P.C.: Automated descriptor selection for quantitative structure-activity relationships using generalized simulated annealing. *J. Chem. Inf. Comput. Sci.* **35**(1), 77–84 (1995)
- Swift, S., Liu, X.: Predicting glaucomatous visual field deterioration through short multivariate time series modelling. *Artif. Intell. Med.* **24**(1), 5–24 (2002)
- Swift, S., Tucker, A., Vinciotti, V., Martin, N., Orenco, C., Liu, X., Kellam, P.: Consensus clustering and functional interpretation of gene-expression data. *Genome Biol.* **5**(11), R94 (2004)
- Szu, H., Hartley, R.: Fast simulated annealing. *Phys. Lett. A* **122**(3), 157–162 (1987)
- Tao, W., Wei-hua, L.: Naive bayes software defect prediction model. In: 2010 International Conference on Computational Intelligence and Software Engineering (CiSE), pp. 1–4. IEEE (2010)
- Tovey, C.A.: Hill climbing with multiple local optima. *SIAM J. Algebraic Discrete Methods* **6**(3), 384–393 (1985)
- Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**(1–2), 479–487 (1988)
- Tsallis, C., Stariolo, D.A.: Generalized simulated annealing. *Phys. A Stat. Mech. Appl.* **233**(1), 395–406 (1996)
- Tucker, A., Garway-Heath, D.: The pseudotemporal bootstrap for predicting glaucoma from cross-sectional visual field data. *IEEE Trans. Inf. Technol. Biomed.* **14**(1), 79–85 (2010)



- Varma, S., Simon, R.: Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **7**(1), 1 (2006)
- Viera, A.J., Garrett, J.M.: Understanding interobserver agreement: the kappa statistic. *Fam Med* **37**(5), 360–363 (2005)
- Vincent, F.Y., Redi, A.P., Hidayat, Y.A., Wibowo, O.J.: A simulated annealing heuristic for the hybrid vehicle routing problem. *Appl. Soft Comput.* **53**, 119–132 (2017)
- Weiss, N.A.: *Introductory Statistics*: Pearson New International Edition. Pearson. ID: 527015
- Westcott, M.C., McNaught, A.I., Crabb, D.P., Fitzke, F.W., Hitchings, R.A.: High spatial resolution automated perimetry in glaucoma. *The British journal of ophthalmology* **81**(6), 452–459 (1997). LR: 20140617; JID: 0421041; OID: NLM: PMC1722228; ppublish
- Xiang, Y., Gong, X.: Efficiency of generalized simulated annealing. *Phys. Rev. E* **62**(3), 4473 (2000)
- Xiang, Y., Gubian, S., Martin, F.: Generalized simulated annealing. In: Peyvandi, H. (ed.) *Computational Optimization in Engineering-Paradigms and Applications*, pp. 25–46. InTech (2017). <https://doi.org/10.5772/62604>
- Xiang, Y., Gubian, S., Suomela, B., Hoeng, J.: Generalized simulated annealing for global optimization: the gensa package. *R Journal* **5**(1), 13–28 (2013). <https://doi.org/10.32614/RJ-2013-002>
- Xiang, Y., Sun, D., Fan, W., Gong, X.: Generalized simulated annealing algorithm and its application to the Thomson model. *Phys. Lett. A* **233**(3), 216–220 (1997)
- Zhang, P.: Model selection via multifold cross validation. *Ann. Stat.* **21**(1), 299–313 (1993)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.