Adaptive Melanoma Diagnosis Using Evolving Clustering, Ensemble and Deep Neural Networks

Teck Yan Tan[1], Li Zhang[1], and Chee Peng Lim[2]

[1]Computational Intelligence Research Group
Department of Computer and Information Sciences
Faculty of Engineering and Environment
University of Northumbria
Newcastle, NE1 8ST, UK

[2]Institute for Intelligent Systems Research and Innovation
Deakin University
Waurn Ponds, VIC 3216, Australia

Email: {teck.tan; li.zhang}@northumbria.ac.uk;
chee.lim@deakin.edu.au

Abstract.
In this research, we propose a variant of the Particle Swarm Optimization (PSO) algorithm, namely hybrid learning PSO (HLPSO), for skin lesion segmentation and classification. HLPSO combines diverse search mechanisms including modified Firefly Algorithm (FA) operations, a new spiral research action, probability distributions, crossover, and mutation procedures to diversify and improve the original PSO algorithm. It is used in conjunction with the K-Means clustering algorithm to enhance lesion segmentation. Its cost function takes both intra-class and inter-class variations into account to increase scalability. Two lesion classification systems are formulated based on HLPSO. In the first system, HLPSO is used to devise evolving convolutional neural networks (CNN) with optimized topologies and hyper-parameters for lesion classification. In the second system, shape and colour features, as well as texture features extracted using the Kirsch operator and Shift Local Binary Patterns are used to produce an initial discriminative lesion representation. HLPSO is then used to identify the most significant components of each feature vector for ensemble lesion classification. Evaluated using several skin lesion data sets, both systems depict superior capabilities in lesion segmentation, deep CNN architecture generation, and discriminative feature selection for ensemble lesion classification, and outperform a number of advanced PSO and FA variants, classical search methods, as well as other related models on skin lesion classification significantly. HLPSO also yields better performances over other classical and advanced search methods in solving a number of benchmark tasks related to mathematical landscapes and those in the complex CEC 2014 test suite.

Keywords: Skin Lesion Segmentation and Classification, Feature Selection, Clustering, Evolutionary Algorithm, Evolving Convolutional Neural Network and Ensemble Classifier.

1. INTRODUCTION

Skin lesion classification is a challenging task owing to the visual similarity between benign and melanoma lesions. Robust and accurate skin lesion segmentation is an important step leading to reliable melanoma diagnosis. Owing to the indiscernible boundary between the tumour and skin as well as considerable physical variations in different cases, there are many challenges pertaining to segmentation of skin lesions. In this research, we propose a variant of the Particle Swarm Optimization (PSO) algorithm in conjunction with clustering, ensemble and deep learning models for scalable lesion segmentation and classification, in an attempt to address the abovementioned problems. The proposed PSO model combines diverse search mechanisms including modified Firefly Algorithm (FA) operations, a new spiral research action, probability distributions, crossover, and mutation procedures to diversify and improve the original PSO algorithm.

Conventional clustering algorithms such as K-Means (KM) or Fuzzy C-means (FCM) show impressive performance for image segmentation and outlier detection [1, 2]. However they are sensitive to centroid or membership initialization, and tend to be trapped in local optima [3, 4]. In this research, we employ the proposed PSO variant in conjunction with the KM algorithm to enhance lesion segmentation. Specifically, the proposed PSO variant is employed to fine-tune the KM cluster centres, where both inter- and intra-cluster measures are taken into account for lesion segmentation.

Two lesion classification systems are subsequently developed. The first system utilizes evolving deep convolutional neural networks (CNN) for lesion classification, where the topologies and hyper-parameters of the deep networks are devised by the proposed PSO variant. The second system is composed of three key stages, i.e. colour, shape and texture feature extraction, feature selection using the proposed PSO variant, and lesion classification using ensemble models. Besides the extraction of clinically important colour and shape features, Kirsch compass masks and a Local Binary Patterns (LBP) variant, i.e. Shift LBP (SLBP), are employed for texture feature extraction from the segmented lesions. The SLBP operator is able to overcome limitations of the original LBP operator and better deal with images with low contrast characteristics. Discriminative features are identified by the proposed PSO model for subsequent lesion classification. Two ensemble classifiers are constructed, whereby each ensemble classifier consists of two base models trained with the optimized colour and shape features as well as optimized texture features, respectively. The detailed system architecture is illustrated in Figure 1.

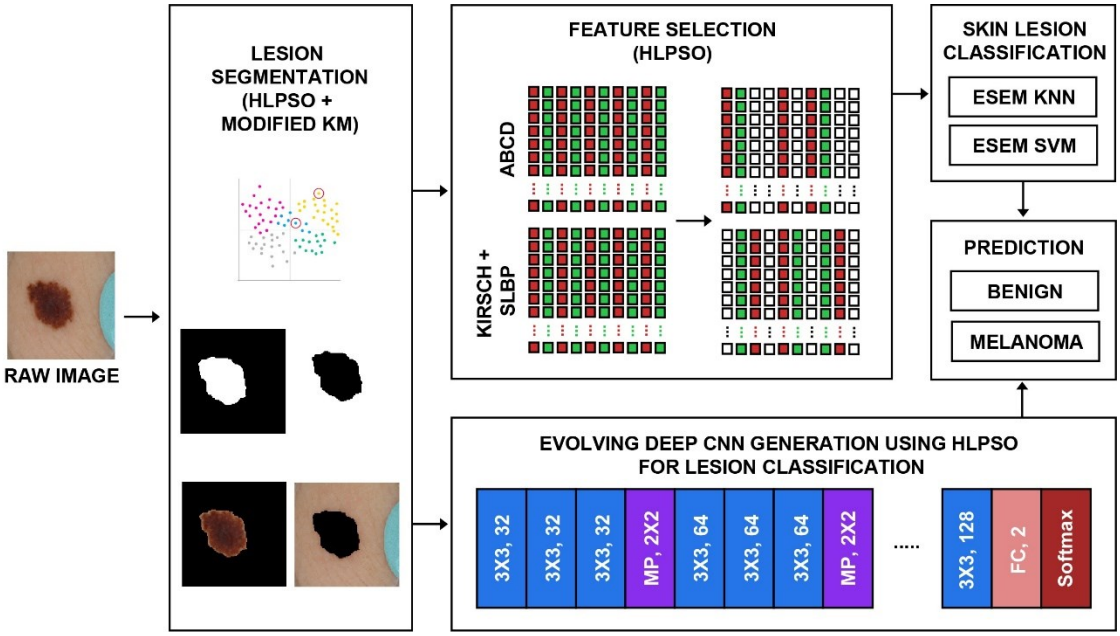

Figure 1 The proposed lesion segmentation and classification systems

The contributions of this research are summarized as follows.

1. A new PSO variant, known as hybrid learning PSO (HLPSO), is proposed for discriminant lesion segmentation, feature selection, and evolving deep CNN model generation for lesion classification. HLPSO employs local search strategies (probability distributions, crossover and mutation operations) and global search mechanisms (a new spiral search action, random and average FA search actions) to overcome stagnation of the original PSO algorithm.
2. HLPSO is firstly employed for optimization of the KM cluster centroids to derive more accurate segmentation of lesion and skin. The new KM variant takes both between- and within-class differences into account for lesion segmentation. It also employs a novel alternative strategy for the within-class measure, where the largest distance between each centre and its corresponding members is used to measure compactness of the clusters. The separation scales of the clusters are also measured by the between-class factors. HLPSO in conjunction with the KM variant is employed to minimize the division of within-class difference by the between-class variation to achieve robust segmentation of lesion and skin.
3. A novel texture descriptor, which incorporates Kirsch compass masks and SLBP, is proposed to derive an initial discriminative lesion representation. The proposed descriptor is capable of dealing with low contrast images by extracting spatial patterns and contrast measures of each segmented region. HLPSO

is employed for feature selection, in order to identify the most significant discriminative characteristics related to benign and malignant lesions.

4. Evolving deep learning models with optimized structures and hyper-parameters are also devised by HLPSO for lesion classification. Three different types of convolutional layers are used as the base for model generation. A convolutional block is constructed for each type of convolutional layer and its associated learning layers such as batchNormalizationLayer and reluLayer. HLPSO is able to generate optimized *n* number of each type of the convolutional block as well as optimized associated weights and hyper-parameters. In comparison with the default and other CNN models generated using other PSO and FA variants and classical search models, the HLPSO-based systems depict statistically superior improvements over other deep learning models.
5. Evaluated with a combined data set and the ISIC subset, the empirical results indicate that the proposed HLPSO-based systems have superior capability of lesion segmentation, discriminative feature selection, ensemble and deep CNN based lesion classification, and they outperform a number of advanced PSO and FA variants, classical search methods, as well as other related skin lesion classification models in the literature by a significant margin. Moreover, HLPSO also shows great superiority over other methods for solving diverse benchmark problems related to mathematical landscapes and complex CEC 2014 test suite.

The paper is organized as follows. In section 2, the related studies on skin lesion segmentation and classification, and swarm intelligence algorithms for image segmentation, clustering and feature selection are presented. Ensemble classification techniques are also discussed. The proposed HLPSO variant is discussed in detail in Section 3. We present lesion segmentation using HLPSO integrated with a KM variant in Section 4, while Section 5 discusses the generation of evolving deep CNN models. Feature extraction, HLPSO-based feature selection as well as ensemble models for lesion classification are provided in Section 6. A comprehensive evaluation and future directions are presented in Sections 7 and 8, respectively.

2. RELATED WORK

In this section, we discuss a variety of state-of-the-art investigations on skin lesion segmentation and classification, ensemble classification techniques and swarm intelligence and evolutionary algorithms for hybrid clustering, feature selection and benchmark optimization.

2.1 Skin Lesion Segmentation and Related Studies

Yuan and Lo [5] proposed deep fully convolutional-deconvolutional neural networks (CDNNs) to generate binary masks for skin lesion segmentation in dermoscopic images. Pixel-wise classification to filter the input image into skin and lesion classes was conducted. The training process was used to minimize a loss function based on the Jaccard distance. Each CDNN had 29 layers, and its hyper-parameters were obtained by grid search. Upsampling and deconvolutional layers were used to restore image resolutions. RGB, Hue-Saturation-Value (HSV) and lightness in LAB spaces were considered for the segmentation tasks. An ensemble model consisting of 6 CDNNs as the base classifiers was used to obtain the final segmentation output. Their work showed superior robustness for lesion segmentation. Li and Shen [6] proposed a deep learning framework composed of two fully-convolutional residual networks (FCRN) for lesion segmentation and coarse lesion classification. The classification results were further refined using a lesion index calculation unit (LICU), which measured the importance of pixels based on their distances to the nearest border. Specifically, the two FCRNs were trained on the original as well as flipped and rotated augmented images, respectively, to generate the initial coarse maps. The distance maps generated by the LICU were convoluted multiplied with the coarse maps to yield the refined maps. The average probabilities of the refined maps were used as the final lesion categorization results. Alvarez and Iglesias [7] employed an iterative colour-based KM clustering algorithm and ensemble of regressions for skin lesion segmentation. Their system consisted of image pre-processing (e.g. removal of reflection and hair), colour clustering, feature extraction and Jaccard score calculation. The regression models were trained using training images and ground truth masks provided by the ISIC 2017 data set. Morphological operations were used as the post-processing procedure. A number of features, including region area, position, circularity, solidity, average colour, were extracted from each segmented region. Random Forest and Support Vector Regression (SVR) models were used to predict the segmentation scores, respectively, with the average of the above scores employed as the Jaccard Index. The process was repeated by increasing the number of clusters by one until there was no significant improvement in the Jaccard score.

Ashour et al. [8] proposed a clustering-based method for skin lesion segmentation known as histogram-based clustering estimation neutrosophic c-means clustering (HBCENCM). The histogram-based clustering estimation (HBCE) method was firstly proposed to identify the number of clusters based on the histogram of each image. The pre-determined number of clusters was then used to drive neutrosophic c-means (NCM) clustering for lesion

segmentation. The proposed HBCENCM model with the horizontal-vertical procedure showed great superiority over the original NCM model (without the pre-determined number of clusters) for skin lesion segmentation. Their model outperformed other deep learning models for lesion segmentation based on the evaluation of the ISIC archive data set. Moreover, fuzzy classification of pixels and histogram thresholding were employed by Garcia-Arroyo and Garcia-Zapirain [9] for lesion segmentation. Guo et al. [10] conducted skin lesion classification using neutrosophic clustering and adaptive region growing in dermoscopic images. The shearlet transform was first used to map the skin lesion images into a neutrosophic set domain. An indeterminate filter was also employed to reduce the indeterminacy of the images. The NCM clustering algorithm as well as an adaptive region growing method was used to carry out lesion segmentation. The model showed impressive performance when it was trained and evaluated using randomly selected subsets of the ISIC 2017 data. Lin et al. [11] compared U-Nets with clustering algorithms for skin lesion segmentation. The U-Net with the proposed histogram-based pre-processing technique was implemented. The model had 4 contracting steps and 4 expansion steps, and was trained over one epoch (i.e. a small number of iterations) with the original as well as flipped and rotated augmented images. In addition, for the clustering model, FCM was firstly used to cluster the images into distinctive regions of interest. KM was subsequently used to further categorize the identified clusters (i.e. the outputs of FCM) based on colour features. The group with the darkest colour features was regarded as the lesion. The U-Net model outperformed the clustering algorithms in their experiments, but with inferior trade-off between performance and computational efficiency.

On the other hand, evolutionary algorithms are used in conjunction with the clustering methods for medical image segmentation tasks. Neoh et al. [3] proposed stimulating discriminant measures (SDM) for image segmentation in microscopic blood cancer detection. The SDM model was incorporated with a Genetic Algorithm (GA) for discriminative segmentation of nucleus and cytoplasm of lymphocytes/lymphoblasts. Their model employed both between-class and within-class variations in the objective evaluation of each set of identified cluster centres. Colour, shape, and texture features were subsequently extracted from the segmented regions for abnormality detection. The GA-based SDM model achieved superior performance for leukaemia classification. A Kernel Possibilistic C-Means algorithm was used in conjunction with PSO for Magnetic Resonance (MR) brain image segmentation in Mekhmoukh and Mokrani [4]. The original FCM algorithm was sensitive to outlier and noise, and its performance relied heavily on initialization of cluster centres. The PSO algorithm was used to initialize the cluster centroids and memberships to overcome the abovementioned limitations. Outlier rejection and spatial neighbourhood information were also taken into account to enhance performance. Their extended FCM algorithm depicted impressive effectiveness in dealing with noise for MR brain image segmentation.

2.2 Skin Lesion Classification

Tan et al. [12] proposed a modified PSO model for discriminant feature selection in skin lesion classification. Their PSO model employed multiple swarm leaders and subswarm concepts. The subswarm particles followed the subswarm leader, and avoided the worst solutions in every dimension or randomly selected partial dimensions, respectively. Probability distributions and dynamic matrix representations were used to increase search diversity. Their model outperformed other search methods significantly for discriminative feature selection for a mixed data set and the PH2 database. Lopez et al. [13] employed the VGGNet model and transfer learning for skin lesion classification with dermoscopic images. The VGG16 network with 5 convolutional blocks and 3 fully connected layers was employed. Three experiments were conducted, i.e. (1) training VGGNet from scratch, (2) using a pre-trained VGGNet on ImageNet with the fully connected layers trained with the lesion data set, and (3) freezing the first four convolutional blocks with the fifth convolutional block loaded with corresponding weights obtained in experiment (1) and the fully connected layers loaded with weights obtained in experiment (2). A total of 346 training images and 150 test images were selected from the ISBI 2016 Challenge data set for evaluation. Transfer learning in Experiment (3) obtained the best performance. DeVries and Ramachandram [14] employed a multiscale Inception-v3 network for skin lesion classification. The model was pre-trained on ImageNet and fine-tuned using the ISIC 2017 data set with two different image resolutions, i.e. low resolution and high resolution centre-crop images. The extracted features for both resolutions were concatenated and passed on to the fully connected layer. The training process updated the weights of the last two inception blocks and the fully connected layers while all other layers in the network remained frozen. Augmented images were produced for each test image for evaluation. The resulting network with different training options was also produced. The combined result of these variant deep networks was used as the final classification output.

Bi et al. [15] employed the deep residual networks (ResNet) for skin lesion segmentation and classification. The ResNet model employed residual blocks, with each block consisting of several convolutional layers, and other associated learning layers. The ResNet model was able to provide shortcut connections that were accumulated

with the output of the convolution layers. Transfer learning was subsequently conducted, where the model pre-trained on the ImageNet data set was fine-tuned using both the ISIC 2017 challenge training set and a large number of ISIC archive images for segmentation and classification tasks. A multi-scale integration method was used for lesion segmentation where different scales and rotations of test images were used as inputs. In addition, three methods were used for lesion classification, i.e. (1) a three-class lesion classification task, (2) two binary lesion classification tasks (melanoma vs others and seborrheic keratosis (SK) vs benign), and (3) an ensemble of the above two models. Schaefer et al. [16] conducted dermoscopic melanoma classification using ensemble models. Their system firstly segmented lesions using border detection algorithms. Then, colour, shape, and texture features were extracted from each segmented lesion region. A set of base models was trained on the extracted feature subspaces. Diversity of the base classifiers was subsequently evaluated. The discriminative base models were selected for ensemble model construction with the redundant ones removed. Their resulting model achieved statistically significant superiority over other ensemble classification methods. Kawahara et al. [17] employed Alexnet pretrained on ImageNet for 10-class lesion classification. The fully connected layers of AlexNet were converted to convolutional layers, in order to extract multi-scale features. The extracted deep skin features were used for training a logistic regression classifier for 10-class lesion classification. Max-pooling across the spatial dimensions was also conducted to reduce dimensionality of the input images with higher resolutions. Besides using pooled-multi-scale feature extraction, a per image normalization and pooling across augmented feature space were conducted to enhance performance.

2.3 Evolutionary and Swarm Intelligence Algorithms

Owing to their superior search capabilities, evolutionary algorithms have been widely used for solving diverse optimization problems, such as network routing, planning, job scheduling, image retrieval and classification, and design optimization. Especially, nature-inspired swarm intelligence (SI) algorithms have received great attention in recent years owing to their simplicity and flexibility. SI algorithms engage a swarm of search agents, which interact with one another and follow simple rules for the exploration of an unknown search space. Instead of adopting dictated centralized control mechanisms, the search agents conduct the search autonomously with a certain degree of randomness. Their communication and social interactions result in the occurrence of intelligent global behaviour, which is unknown to individual agents at the beginning of the search process. Such a self-organizing search process can evolve over a number of iterations to achieve converged states. To ensure the effective exploration of the search space for finding global optimality, it is important that the SI algorithms show sufficient trade-off between intensification and diversification. SI algorithms have been applied to undertaking diverse optimization problems pertaining to classification, regression, clustering and forecasting models, including discriminative feature selection, evolving deep architecture generation, parameter tuning and cluster centroid enhancement. Motivated by swarm behaviours such as bird flocking and fish schooling, respectively, PSO and FA are two of the most popular SI algorithms. Other popular SI algorithms include Cuckoo Search (CS), Bat Algorithm (BA), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Grey Wolf Optimization (GWO), Dragonfly Algorithm (DA), and Flower Pollination Algorithm (FPA) [18].

Proposed by Kennedy and Eberhart [19], the PSO model employs the personal and global best solutions to lead the search process of the particle swarm, as defined in Equations (1)-(2).

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{1}$$

$$v_{id}^{t+1} = w \times v_{id}^{t} + c_1 \times r_1 \times (p_{id} - x_{id}^{t}) + c_2 \times r_2 \times \left(p_{gd} - x_{id}^{t}\right) \tag{2}$$

where p_{gd} and p_{id} denote the swarm leader and the personal best solution of particle x_i in the d^{th} dimension, respectively. The position and velocity of particle x_i in the $(t+1)^{th}$ iteration and the d^{th} dimension are denoted as x_{id}^{t+1} and v_{id}^{t+1}, respectively. The acceleration coefficients, c_1 and c_2, are used to adjust the convergence speed, while the inertia weight, w, is used to fine-tune the impact of the previous velocity on the current one, with r_1 and r_2 as random vectors.

Unlike the PSO operation, the FA model with Levy flights [20] employs brighter neighbouring solutions, instead of the global best solution, to guide the search process, as defined in Equation (3).

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j - x_i\right) + \alpha \,\text{sign}[\text{rand} - \tfrac{1}{2}] \oplus \text{Levy} \tag{3}$$

where x_j represents a brighter neighbouring solution. The light absorption coefficient, γ, is used to influence the attractiveness rate and determine the convergence speed. Note that r_{ij} denotes the distance between the two fireflies, x_i and x_j. The attraction strength between two fireflies decreases as their distance increases.

The final term in Equation (3) indicates the randomization operation following a Levy distribution, where α represents the randomization parameter. Specifically, Levy distribution possesses an infinite second moment in comparison with a finite second moment in Gaussian distribution. It employs adjustable parameters to implement effective mutation operators. A simplified mathematical definition for Levy distribution is provided in Equation (4) [20, 21].

$$f(\sigma,\tau,\varphi) = \sqrt{\frac{\tau}{2\pi}}\,\frac{e^{\frac{-\tau}{2(\sigma-\varphi)}}}{(\sigma-\varphi)^{3/2}} \qquad (0<\varphi<\sigma<\infty) \tag{4}$$

where φ and τ denote the shift and scale parameters, respectively. Moreover, Equation (5) defines Levy distribution using Fourier transform [20, 21].

$$L(\theta,m,\delta) = e^{-\theta|m|^{\delta}} \tag{5}$$

where θ denotes a scale parameter with the range of [-1, 1], and δ represents the Levy index with the range of (0, 2]. Equation (5) can be used to implement Cauchy and Gaussian distributions when δ=1 and δ=2, respectively. The parameter δ has a significant impact on the shape of the probability distribution. A smaller setting of δ leads to a longer tail, and vice versa.

Levy flights are effective in exploring an uncertain environment consisting of an unknown large-scale search space. They are able to generate more effective offspring solutions with long jumps (especially when $0<\delta<1$) in comparison with those of Cauchy and Gaussian distributions. Therefore, Levy flights are more efficient than Gaussian and Cauchy random walk mechanisms. Besides FA, Levy flights have been embedded in other swarm intelligence algorithms such as fruit flies and spider monkeys [20, 21], and can be used to represent many real-world physical phenomena such as noise and cooling behaviours [20].

Moreover, owing to their simplicity and flexibility, PSO and FA and their variants have been employed to solve diverse real-life single-objective and multi-objective optimization problems, which include hyperspectral image classification [21], colour image segmentation [3], stock price index forecasting [22], data clustering [23] and mathematical benchmark optimization [24]. They have also been used for discriminative feature selection for facial and bodily expression recognition [25, 26, 27, 28], brain tumour [4], heart disease [29, 30], skin and blood cancer [31, 32] detection and classifier ensemble reduction [33].

Other recently proposed inspiring evolutionary models for solving benchmark functions are available in the literature. Motivated by the spiral phenomena in nature, Tamura and Yasuda [34] proposed three two-dimensional spiral search operations with different parameter settings. These spiral operations adopted the search behaviours of logarithmic spirals. Evaluated using three benchmark functions, the proposed methods outperformed PSO. The use of randomness in the spiral mechanisms was proposed as their future work. An extension of the abovementioned work implemented a multi-dimensional spiral search action [35], which was built based on rotation matrices defined in a multi-dimensional space. This multi-dimensional spiral operation showed enhanced performance in comparison with those of other search methods for solving diverse mathematical landscapes.

2.4 Hybrid Clustering and Classification Models

There are diverse hybrid evolutionary clustering and classification models proposed in the literature utilising meta-heuristic methods integrated with clustering and classification models to further enhance performance. Thangavel et al. [36] proposed a clustering model incorporating PSO with Simulated Annealing (SA) for the identification of biclusters in gene expression data. SA was used to enhance the stagnant particles to avoid premature convergence. Evaluated using three medical data sets, the proposed clustering model showed competitive performance, which extracted biclusters with higher coherence and larger volume. Hatamlou [37] proposed an optimization model, namely Heart, for clustering analysis. Two search strategies, i.e. the movement of blood molecules and the pump action of the heart, were proposed to simulate the contraction and expansion actions of the heart. The blood molecule movement operation enabled the molecules to move towards the heart to increase global exploration, whereas the pump action empowered dynamic local exploitation around the best solution, i.e. the heart. The model was evaluated using a set of six UCI benchmark data sets and outperformed KM and three other classical search methods, i.e. PSO, gravitational search algorithm (GSA), and big bang-big crunch algorithm (BB-BC), respectively. Farzi and Kianian [38] proposed a k-Medoid based clustering model for attributed graph generation. Their model was effective in tackling Political Blogs and DBLP bibliography data sets in comparison with other baseline methods. In addition, a meta-heuristic passive mine detection and classification model, based on the magnetic anomaly, measurement height, and soil type, was proposed by

Yilmaz et al. [39]. Firstly, the Multi-layer Perceptron (MLP) with backpropagation learning mechanism was used to determine the effectiveness of different combinations of the input variables. Subsequently, the identified best feature combinations were used to perform mine detection and classification using traditional as well as enhanced k-Nearest Neighbour (KNN) models. Specifically, weight coefficient optimization for the input attributes was performed using GA in their enhanced KNN model to improve performance. A set of 32 classification models was constructed based on the traditional and enhanced KNN methods pertaining to the number of different neighbours and different distance metrics. Their findings indicated that the meta-heuristic KNN method integrated with the fuzzy distance metrics produced the best mine detection and classification performance among all the system prototypes.

Nerurkar et al. [40] proposed a new cost function for the original PSO model for clustering analysis. Their fitness function took both intra- and inter-cluster distances into account, with a linear computational cost. The PSO clustering model integrated with this new cost function outperformed other evolutionary and non-evolutionary clustering methods with pure intra-cluster measures for the evaluation of six artificial benchmark data sets. Nananukul [41] proposed an enhanced KM clustering model for solving production inventory and distribution problems. The cost function of their clustering model took into account the holding costs of customers and demand patterns. A reactive Tabu search algorithm was employed to further improve the initial solution identified by the KM clustering model. Evaluation using complex real-life customer data indicated the superiority of the proposed model. A comprehensive survey of diverse clustering models was provided in Xu and Tian [42].

2.5 Ensemble Classification and Feature Selection Techniques

Classifier ensemble construction using a modified FA model has been conducted by Zhang et al. [33]. Their proposed FA model employed the neighbouring and global best experiences to lead the attractiveness operation whereas the local and global worst experiences have been used for the guidance of an evading mechanism. The FA model was subsequently used to identify the smallest optimal sets of diversified base evaluators for ensemble classifier construction while maintaining classification performance. Their work outperformed the full-sized ensemble models as well as the ensemble classifiers devised by other classical and advanced search methods significantly, for the evaluation of several UCI data sets. Moyano et al. [43] proposed an evolutionary approach, namely Evolutionary Multi-label Ensemble (EME), for the automatic generation of ensemble classifiers. Their ensemble construction took the imbalance, dimensionality and relationships among the class labels into account. Specifically, each base model was dedicated to the prediction of a subset of classes. Both the classification performance and the occurrences of each class label were taken into account for ensemble model fitness evaluation. A mutation operator was also employed to identify the relationship of different classes by favouring more related combinations of labels. The experimental results for the evaluation of 14 data sets indicated the superiority of their EME model over a set of baseline classical and ensemble-based algorithms. Ribeiro et al. [44] performed wavenet ensemble construction for short-term load forecasting. Two base model generation techniques were adopted, i.e. bootstrapping and cross-validation integrated with the inputs decimation algorithm, to ensure the generation of diverse base regressors. Especially, the former focused on manipulating data samples while the latter focused on manipulating feature dimensions. The base model selection was conducted by using both forward selection and pruning by ranking. Subsequently several ensemble aggregation techniques, i.e. simple mean, mode, median, and stacking, were used for ensemble construction. Their model outperformed other prediction methods, such as a MLP and a regression tree, for the evaluation of two real data sets. Moreover, ensemble classier construction techniques were also studied by Thomas et al. [45] for proactive quality monitoring and control. They employed four types of base classifiers for ensemble construction, i.e. KNN, MLP, Decision Tree (DT), and Support Vector Machine (SVM). Bagging was used for the generation of training data sets in order to increase the diversity of the base models. A series of 19 ensemble classifiers was constructed based on the diversity and accuracy of the base models. These ensemble models were composed of not only purely the KNN, DT, SVM or MLP evaluators, but also the combinations of different types of the base classifiers. The ensemble model constructed by using the four different types of base classifiers achieved the minimum global misclassification rate, while the one constructed by purely using the NN base classifiers obtained a better non-detection rate. Guo et al. [46] proposed heterogeneous ensemble-assisted multi-objective evolutionary algorithms (MOEAs) for solving computationally expensive optimization problems. Instead of using Gaussian processes (GPs), their model employed a heterogeneous ensemble for fitness evaluation. To increase ensemble diversity, two radial basis function networks (RBFNs) and a least square SVM model, trained with different input features (selected/extracted subsets of features or full feature sets), were used for base evaluator generation. Evaluated using two well-known test suites, i.e. DTLZ and WFG, their model achieved better trade-off between performance and computational efficiency in comparison with the GP-assisted MOEAs.

Besides the above ensemble model construction investigations, there are also diverse research studies focusing on SI and evolutionary algorithm based feature selection. De Souza et al. [47] proposed a Binary Crow Search Algorithm (BCSA) for feature selection. A V-shaped transfer function was used to transform the positions of crows from continuous values into binary ones. The classical CSA model employs an awareness probability factor to balance between intensification and diversification. It also achieves a fast convergence rate by memorizing the best known positions of the crows as well as adopting the flight length search parameter. Evaluated using six UCI benchmark data sets, their BCSA model showed great discriminative capabilities for feature selection and outperformed other search methods such as Binary Bat Algorithm (BBA), Binary PSO (BPSO), Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), in terms of classification performance and computational cost. Song et al. [48] proposed an evolutionary multi-objective ensemble learning (EMOEL) algorithm for electricity consumption prediction. The EMOEL model was used to devise the number of the hidden neurons in the Extreme Learning Machines (ELMs), and in the meanwhile to identify the feature extraction methods, the sizes of the time windows as well as the most optimal input feature channels (e.g. representing energy consumption data and environmental factors). A variety of prediction models recommended by the identified optimized solutions in the Pareto front were aggregated to yield ensemble prediction where the Differential Evolution (DE) algorithm was used to optimize the combination coefficients for the building of the ensemble. Experimental studies using a real-world electricity consumption data set ascertained the efficiency of the EMOEL model in comparison with other search methods. Zhang et al. [30] proposed a modified FA model for feature selection. Their FA variant employed two swarm leaders with similar fitness scores but remote in positions to guide the search process. Specifically the mean vector of the two swarm leaders and optimal neighbouring solutions were further enhanced using the SA model. These enhanced local and global signals as well as the chaotic search parameters were then used to lead the attractiveness operation. The aforementioned two swarm leaders were also used to divert weak solutions to optimal regions with fast convergence rates. The empirical and statistical results for 29 classification and 11 regression data sets indicated the superiority and robustness of their FA model over other search methods. Bushehri and Zarchi [49] employed Probabilistic Neural Network (PNN) for the self-care problem classification for children with physical and motor disability, where the GA was used to conduct feature optimization. A comprehensive survey of SI and evolutionary algorithm based feature selection was provided by Xie et al. [50].

3. THE PROPOSED HLPSO MODEL

In this research, we propose the HLPSO model to perform image segmentation, feature selection, and optimized deep CNN model generation. HLPSO incorporates diverse search strategies, including modified FA operations, a spiral search action, random walks, crossover and mutation procedures, to diversify the original PSO operation. Moreover, an extended KM clustering algorithm is proposed for lesion segmentation, whose centroids are further enhanced using HLPSO with both intra- and inter-cluster variance measures. Two lesion classification systems are formulated. Firstly, HLPSO is used to optimize the structure and hyper-parameters of deep CNNs for lesion classification. Secondly, HLPSO is used for feature selection in ensemble lesion classification. Algorithm 1 shows the pseudo-code of HLPSO.

Specifically, after initialization of the swarm, all particles are ranked based on their fitness values. The overall population is divided into two subswarms, where the first subswarm (subswarm 1) consists of 50% top ranked particles while the second subswarm (subswarm 2) contains the remaining particles. Three probability distributions, i.e. Gaussian, Cauchy, and Levy distributions, are used to further improve the promising solutions in subswarm 1, whereas newly proposed search mechanisms including two neighbouring search strategies, a spiral search operation, as well as the PSO action, are used to explore the search space in subswarm 2. Both subswarms are combined after a number of iterations. Another set of top ranked promising solutions is identified. The crossover and mutation operators are used to generate offspring solutions using the top ranked particles as parents. The search process iterates until the termination criteria are met. In the following sub-sections, we describe each key proposed mechanism comprehensively.

Algorithm 1: Pseudo-Code of the Proposed HLPSO Model	
1.	**Start**
2.	Initialize a population of particles randomly;
3.	Evaluate the population;
4.	Sort the population based on fitness values and identify $gBest$;
5.	**While** (!Stagnation) {
6.	Divide the population into two sub-swarms (with the top 50% particles stored in subswarm 1 and the remaining 50% in subswarm 2);
7.	**For** (each particle in sub-swarm 1) do {

8.	Conduct long jumps using Levy flights/Gaussian/Cauchy distributions as in Equations (7-9);
9.	Update the $pBest$ if the new position developed by long jumps achieves better fitness;
10.	} **End For**
11.	**For** (each particle in sub-swarm 2) do {
12.	Randomly select one of the following operations to update the particle position;
13.	1. Conduct the original PSO operation as defined in Equations (1)-(2);
14.	2. Conduct the modified FA operation using a randomly selected brighter neighbouring firefly as defined in Equation (10);
15.	3. Conduct the modified FA operation using the average of all brighter neighbouring fireflies as defined in Equation (13);
16.	4. Conduct the spiral search action as defined in Equation (14);
17.	Update the $pBest$ if the new position developed by any of the above randomly selected operations achieves better fitness;
18.	} **End For**
19.	Combine two sub-swarms;
20.	Sort the overall swarm based on the fitness values and identify $gBest$ and the top 20% particles;
21.	Generate a set of offspring particles (e.g. 20% of the swarm) using the crossover operation and the top ranked particles as parents;
22.	Generate another set of offspring particles (e.g. 20% of the swarm) by mutating the above newly generated offspring solutions using the crossover operation;
23.	Replace the bottom 40% particles with the new offspring solutions generated using the crossover and the mutation operations as in lines 21-22;
24.	Sort the population based on the fitness scores and identify the $gBest$;
25.	} **Until** (Stagnation)
26.	Output $gBest$;
27.	**End**

3.1 Local Exploitation Using Gaussian, Cauchy, and Levy Distributions

As mentioned earlier, the top 50% particles of the swarm are selected to generate subswarm 1. These particles are further enhanced using local search operations such as the Gaussian, Cauchy, and Levy distributions. Equation (6) defines these random walk strategies.

$$x_{id}^{t+1} = x_{id}^{t} + (Ub_d - Lb_d) \times \varepsilon \tag{6}$$

where ε denotes the random search mechanisms, i.e. Gaussian, Cauchy, and Levy distributions. Ub_d and Lb_d denote the upper and lower boundaries of the solution vectors in the dth dimension. The proposed HLPSO model randomly selects any of these three probability distributions for position updating in subswarm 1. These random walk strategies increase local exploitation of the promising solutions in subswarm 1.

As mentioned in Section 2.3, Gaussian, Cauchy, and Levy distributions are efficient random walk strategies. They possess unique search characteristics. The combination of these search operations enables local exploitation with sufficient diversity. Equation (7) defines the local search exploitation using the Gaussian distribution.

$$x_{id}^{t+1} = x_{id}^{t} + (Ub_d - Lb_d) \times g(k, h) \tag{7}$$

where $g(k, h)$ represents the Gaussian distribution while k and h denote the mean and the standard deviation of the distribution, respectively. If the new offspring solution x_{id}^{t+1} generated using the Gaussian distribution has a better fitness score than that of the current solution vector, x_{id}^{t}, it is used to substitute the current individual. Otherwise the current solution x_{id}^{t} is passed on to the next generation.

Similarly, Equation (8) defines the search operation using Cauchy distribution for further enhancement of the current decision vector.

$$x_{id}^{t+1} = x_{id}^{t} + (Ub_d - Lb_d) \times \vartheta(z, v) \tag{8}$$

where $\vartheta(z, v)$ denotes the Cauchy distribution with z and v signifying the location parameter and the scaling factor, respectively. If this random walk mechanism is able to produce a fitter solution, x_{id}^{t+1}, this new offspring solution substitutes the current decision vector. Otherwise, the current individual x_{id}^{t} is used to guide the search in the next iteration.

Levy distribution is also used to conduct local exploitation of the particles in the subswarm. Levy flights are more effective in exploring an unknown large-scale search space with longer jumps in comparison with those of Cauchy and Gaussian distributions. Equation (9) denotes the search action using the Levy distribution.

$$x_{id}^{t+1} = x_{id}^{t} + (Ub_d - Lb_d) \times L(\theta, m, \delta) \tag{9}$$

Where $L(\theta, m, \delta)$ denotes the Levy distribution and θ and δ represent the scale parameter and the Levy index, respectively. If the fitness score of the new offspring solution x_{id}^{t+1} yielded by the Levy distribution is better than that of the current particle x_{id}^{t}, it is used to replace the current solution. Otherwise, the current decision vector remains intact and passed on to the next iteration.

The above Gaussian, Cauchy and Levy search mechanisms are randomly selected to conduct effective and diversified local exploitation of the particles in subswarm 1.

3.2 The Proposed Modified FA Operations

The remaining 50% particles are used to form subswarm 2. Several search mechanisms are used to diversify the search process in subswarm 2. Specifically, we include random and average FA operations and a new spiral search action to overcome stagnation of the original PSO operation. The motivations of using FA operations are as follows. Unlike the original PSO algorithm that employs the personal and global best solutions to lead the search process, the FA model enables each firefly to follow neighbouring brighter solutions, in order to better explore the search space. As such, enhanced FA models are used to provide distinctive alternative search behaviours to diversify the search procedure.

In the original FA operation, each individual carries out the search process by following all promising neighbouring solutions as indicated in Equation (3). The search process becomes computationally costly as the problem dimension increases. Instead of using all brighter neighbouring fireflies, the first modified FA operation uses a randomly selected promising neighbouring solution for position updating, as defined in Equation (10), to improve computational efficiency and search diversity.

$$x_i = x_i + \beta \times \left(x_j' - x_i\right) + \alpha' \,\text{sign}[\text{rand} - \tfrac{1}{2}] \oplus \text{Levy} \tag{10}$$

$$\beta = (1 - \beta_{min}) \times e^{-\gamma r_{ij}^2} + \beta_{min} \tag{11}$$

$$\alpha' = \alpha' \times \left(\frac{10^{-4}}{0.9}\right)^{\frac{1}{maxi\_Gen}} \tag{12}$$

where x_j' denotes a randomly selected neighbouring firefly with a higher light intensity. Equation (11) defines a comparatively stronger attractiveness impact to accelerate convergence, where $\beta_{min} = 0.2$. In Equation (12), the randomized parameter α' is adaptive, and is used to fine-tune the randomized step over iterations.

Instead of using a randomly selected brighter neighbouring solution, the second modified FA operation takes the average position of all brighter neighbouring fireflies into account for position updating. As defined in Equation (13), this modified FA search mechanism employs the mean of all brighter individuals to lead the search process.

$$x_i = x_i + \beta \times \left(x_j'' - x_i\right) + \alpha' \,\text{sign}[\text{rand} - \tfrac{1}{2}] \oplus \text{Levy} \tag{13}$$

where x_j'' represents the average position of all brighter neighbouring fireflies.

In comparison with the original FA action where each firefly is guided by each of the brighter neighbouring solutions, both proposed search strategies perform movement towards either a randomly selected brighter firefly or the average position of all the brighter neighbouring solutions to accelerate convergence.

3.3 The Proposed Spiral Search Operation

Motivated by MFO [51] and PSO [19], a new logarithmic spiral search operation is proposed, as defined in Equation (14). It guides the search process using both personal and global best solutions. This spiral research mechanism is able to divert each weak solution to optimal regions in a comparatively small number of iterations.

$$x_i = (mean(g_{best} + p_{best}) - x_i) \times e^{bt} \cos(2\pi t) + mean(g_{best} + p_{best}) \quad (14)$$

$$t = (\alpha' - 1) \times rand + 1 \quad (15)$$

where p_{best} and g_{best} represent the personal and global best solutions, respectively. A constant variable, b, is used to indicate the shape of the logarithmic spiral while $t \in$ [-1, 1] designates the closeness of the current individual to mean of the personal and global best solutions in the next iteration, where -1 and 1 denote the closest and farthest to mean of the personal and global best solutions, respectively. The value of t as defined in Equation (15) is randomly generated for each dimension of each individual, with α' defined in Equation (12).

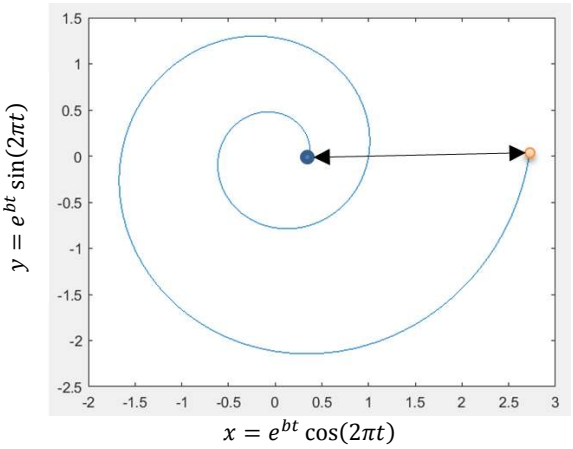

Figure 2 An example of the logarithmic spiral search trajectory, where the x axis denotes the step parameter $e^{bt} \cos(2\pi t)$ defined in Equation (14) with $t \in$ [-1, 1] and b=1, while the y axis denotes $e^{bt} \sin(2\pi t)$. (Both coordinates work together for spiral search path generation.) The blue and orange circles represent the mean of the personal and global best solutions and the current individual, respectively

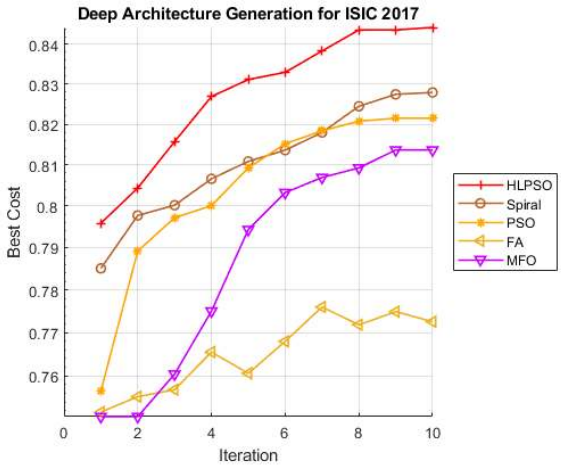

Figure 3 Convergence curves of the proposed spiral search operation, HLPSO, and other search methods over a set of 30 runs for deep architecture generation for the ISIC data set

An example of the logarithmic spiral search trajectory is illustrated in Figure 2, where the x axis denotes the step parameter, $e^{bt} \cos(2\pi t)$, defined in Equation (14). As illustrated in the x axis in Figure 2, the range of the step parameter $e^{bt} \cos(2\pi t)$ is [-1.67, 2.75], with $t \in$ [-1, 1] and b=1. Given a positive step parameter, the spiral search operation explores the search space between the mean of the personal and global best solutions and the current individual. On the other hand, a negative step parameter leads the search operation to explore the region outside of the space between the mean of the personal and global best solutions and the current individual. In other words, a random exploration of the search region within and outside of the area between the current individual and the mean of the local and global optimal solutions by following a spiral distribution is initiated. Figure 3 shows the convergence curve of the proposed spiral search operation in comparison with those of HLPSO (full version), PSO, FA and MFO over a set of 30 runs for deep architecture generation for the ISIC data

set. Since the spiral operation is guided by both global and personal best experiences, it shows faster convergence in comparison with those of PSO, MFO, and FA, and enables the weak solutions to converge to the optimal regions in a small number of iterations.

The HLPSO model selects the two modified FA operations, the spiral search and the PSO actions randomly in each iteration to diversify the search process and increase global exploration in the second subswarm.

3.4 The Crossover and Mutation Operations

After a number of iterations, both subswarms are merged. A new set of top 20% particles is identified. Offspring solutions are generated using the crossover and mutation operators with these top ranked particles as parents. Specifically, the crossover operators are firstly used to generate offspring solutions. Overall, three crossover strategies are included in HLPSO, i.e. single point, double point, and uniform crossover operations [52]. A roulette wheel selection mechanism is used to determine which crossover operation is used for offspring generation. When a single point crossover is applied, the exchange of genes between two parent chromosomes is conducted by copying the genes from the beginning to a random crossover point from one parent, with the rest copied from another parent. When the double point crossover is performed, the string from the beginning to a first random crossover point is copied from one parent, and the genes from the first to the second random crossover points is copied from another parent, with the rest copied from the first parent. For the uniform crossover, a pseudo-random scalar integer vector, μ, is generated. The following equation is used for offspring generation.

$$Offspring = \mu \times x_l + (1 - \mu) \times x_b \tag{16}$$

where x_l and x_b represent two promising parent chromosomes.

The top 20% particles are selected in pairs to perform the crossover operations. The crossover operation produces a set (i.e. 20% of the original population) of new offspring particles. Subsequently, these newly generated offspring solutions are mutated by altering randomly selected genes. Another set (i.e. 20% of the original population) of mutated offspring is produced. These promising offspring solutions from crossover and mutation operations are employed to replace the bottom 40% particles in the swarm.

The above proposed local and global search operations of HLPSO increase intensification and diversification of the search process to overcome stagnation of the original PSO algorithm. HLPSO is subsequently used for skin lesion segmentation, feature selection and optimized deep network generation. Different objective functions are used to evaluate the swarm, as detailed in the following sections.

4. LESION SEGMENTATION

The proposed HLPSO model is first integrated with the KM clustering algorithm for lesion segmentation. Segmentation is conducted in lightness (i.e. the L^* component) of the CIELAB colour space. Several pre-processing steps are initially applied to remove noise (e.g. hair removal as well as contrast and sharpness enhancement). In the conventional KM algorithm, random initialization of the centroids is very important to the clustering performance. However, the search process could be trapped in local optima [3, 4]. To achieve more scalable lesion segmentation, HLPSO is used to further improve the centroids obtained by the KM algorithm for lesion segmentation. Algorithm 2 shows the proposed lesion segmentation algorithm.

Algorithm 2: Pseudo-Code of the Proposed Lesion Segmentation Algorithm	
1.	**Start**
2.	Initialize a population of particles randomly;
3.	Initialize the KM parameters (e.g. number of clusters);
4.	Conduct KM clustering to obtain the initial cluster centres;
5.	Assign the initial centres obtained from KM to the first particle;
6.	Conduct fitness evaluation of the swarm and identify the personal and global best solutions;
7.	**While** (!Stagnation) {
8.	Conduct the proposed HLPSO model for position and velocity updating;
9.	Perform fitness evaluation of the overall swarm;
10.	Update the personal and global best solutions;
11.	} **Until** (Stagnation)
12.	Use the global best solution, $gBest$, as the cluster centres to group pixels in the image into different clusters.
13.	Visualize the pixels of each cluster;
14.	Calculate Jaccard score and store the segmented skin and lesion images;
15.	**End**

As indicated in Algorithm 2, KM is used to obtain the initial cluster centres. We consider two clusters, i.e. lesion and skin regions, for the segmentation task. A swarm of 20 particles is randomly initialized to represent a set of possible cluster centres, with each particle denoting a possible set of centroids. The initial centroids identified by KM are assigned to the first particle as the seed. The swarm is subsequently evaluated. Unlike traditional KM where the pure within-class variance is considered, the fitness evaluation function, as defined in Equations (18)-(22), takes both within- and between-class variance into account. HLPSO is used to perform the search of the optimal centroids in the pixel space. The final global best solution (i.e. the most optimal centroid) is used to cluster the pixels into the lesion and skin classes. The segmented lesion clusters are visualized with the Jaccard index calculated.

KM identifies the cluster centres by minimizing the sum of squared distances between all objects in each cluster and their centre, as defined in Equation (17) [53]. In other words, the objective function evaluation of KM only takes within-class variation into account.

$$J_{KM} = \sum_{i=1}^{C} \sum_{o_j \in C_i} \left\| o_j - c_i \right\|^2 \tag{17}$$

where $c_i = \frac{1}{N_i} \sum_{o_j \in C_i} o_j$.

In Equation (17), $o_j \in C_i$ denotes an object o_j belonging to cluster C_i, and c_i represents the centroid (i.e. the centre point) of cluster C_i. $\left\| o_j - c_i \right\|$ denotes the Euclidean distance between o_j and c_i. Centroid c_i is calculated using the mean of objects in the cluster. Therefore, the cluster centroids can be easily affected, if very large or small values are present in the clusters (such as noisy lesion images), to skew the within-cluster evaluation. Moreover, as indicated in Equation (17), the objective function purely focuses on within-cluster variation for the discriminant measure. It does not take the between-cluster variation into account.

In order to overcome the limitation of the original partitioning quality measure of KM, we propose a new fitness evaluation for each particle (i.e. each set of possible centroids) by considering both within- and between- cluster variations. A variant of within-cluster discriminant measure is also employed, in order to better evaluate the compactness of clusters and avoid skew of the mean for noisy data as discussed earlier. Equations (18)-(19) define the modified within-cluster measure.

$$d_i = \underset{o_j \in C_i}{\arg\max} \left\| o_j - c_i \right\| \tag{18}$$

$$LW_v = \sum_{i=1}^{C} (d_i) \tag{19}$$

As indicated in Equations (18)-(19), the within-class measure is conducted by using the sum of the maximum distance between the cluster members and the corresponding centre. The maximum distance between the members and its centre represents the maximum variation to the mean. There is no any other data object in the cluster exceeding such a limit. Such a discriminant measure is able to better measure cluster compactness in comparison with the original objective function defined in Equation (17).

We also define the between-class discriminant measure as follows. The separation of two clusters is usually determined by the boundary data sample distributions of the two clusters, as indicated in Kuo and Landgrebe [54]. Motivated by Neoh et al. [3], we employ the smallest distance between a centre and members of another cluster to measure the separation of a pair of clusters. Equations (20)-(22) define the inter-cluster discriminant measure.

Given a pair of clusters (e.g. Cluster 1 and Cluster 2), the minimum distances between all the members in Cluster 1 and the centre of Cluster 2, and vice versa, are calaculated, as follows.

$$Com_{1,2} = \min \left(\text{dist}\left(c_2, o_i^{(1)}\right) \right), \quad \forall i \in \{1,2,\dots,N_1\} \tag{20}$$

$$Com_{2,1} = \min \left(\text{dist}\left(c_1, o_k^{(2)}\right) \right), \quad \forall k \in \{1,2,\dots,N_2\} \tag{21}$$

The between-class variantion is obtained by selecting the smaller value between the two minimum distances. The inter-cluster variation therefore indicates the separation scale of the two clusters.

$$LB_v = \min\left(Com_{1,2}\ ,\ Com_{2,1}\right) \tag{22}$$

The fitness evaluation of each particle (i.e. each possible set of centroids) is calculated as division of within-class variation by the between-class variation, as shown in Equation (23).

$$F(x_i) = \frac{LW_v}{LB_v} \tag{23}$$

Therefore, lesion segmentation using HLPSO aims to minimize the above fitness function. Smaller LW_v and larger LB_v indicate higher intra-cluster similarity and larger inter-cluster difference, respectively. The HLPSO-based KM clustering model in combination with the intra- and inter-cluster variance measures depicts superior robustness and scalability in lesion segmentation, and is insensitive to noise and centroid initialization. Ensemble and deep CNN models are subsequently used to perform lesion classification using the segmented lesion sub-images.

5. GENERATION OF EVOLVING DEEP CONVOLUTIONAL NEURAL NETWORKS

After lesion segmentation, HLPSO is used to generate optimized deep CNNs for benign/malignant lesion classification. The CNN models achieve superior capabilities in undertaking diverse computer vision tasks. Its performance relies strongly on the selection of effective architectures that fit the tasks. The current architecture design and parameter selection for deep learning models mainly rely on trial and error. Therefore, when the problem complexity scales up, the manual design of optimal topologies and hyper-parameters becomes a difficult task. In order to overcome such barriers, we employ HLPSO to automatically identify optimal topologies and hyper-parameters of CNNs for skin lesion classification.

In CNNs, hyper-parameters such as the learning rate and weight decay have significant impact on the network performance. A high learning rate leads to sub-optimum performance whereas a low learning rate results in a slow training process. Weight decay (i.e. the regularization factor) also plays a very important role in reducing overfitting. In addition, the structures and integration of different convolutional layers, the filter sizes and the number of filters play significant roles in achieving optimal performance. Therefore, HLPSO is employed to optimize the key factors, i.e. the topologies of different convolutional blocks (including different types of convolutional layers and their associated layers, weights, and the number of filters), learning rate, and weight decay to yield an optimal CNN model for lesion classification.

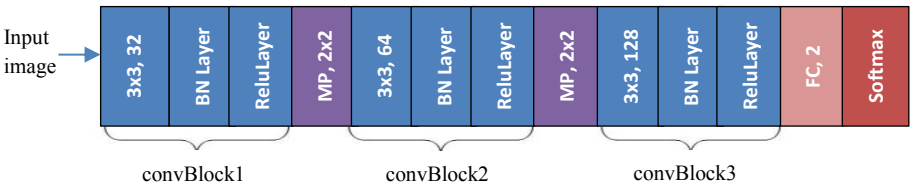

Figure 4 The structure of the initial CNN model

The evolving process of the deep learning models can be conducted in a number of different ways, e.g. starting from scratch or a default basic CNN structure. In this research, we start the model evolving process with a basic CNN structure, as shown in Figure 4. This basic CNN model is also used as the default model for comparison. It contains three types of convolutional blocks. Each convolutional block (convBlock) consists of one convolutional layer, one batchNormalizationLayer, and one reluLayer. As shown in Figure 4, there are three types of convolutional layers, i.e. one with n number of filters (for convBlock 1), another with $2 \times n$ number of filters (for convBlock 2), and the third with $4 \times n$ number of filters (for convBlock 3). These convolutional blocks are preceded by the image input layer, and are followed by maxpooling, fully connected, softmax, and classification layers.

Table 1 Optimized factors

Optimized factors	Strategy	Search range
Learning rate	Directly optimized	[1e-4, 5e-1]
Weight decay	Directly optimized	[1e-5, 5e-2]
Depth of each convolutional block	Directly optimized	[1, 3]
Initial number of filters	Indirectly optimized, calculated using Equation (24)	[18, 32]
Weights for each convolutional block and the fully connected layer	Indirectly optimized	[-1, 1]

The CNN model classifies the input data samples into two classes, i.e. benign and melanoma. The initial basic CNN model shown in Figure 4 has depth=1 for each type of convolutional blocks. The generation of the evolving deep learning model starts with the abovementioned network structure, and identifies two hyper-parameters, and the optimal depths of different convolutional blocks and their associated weights and number of filters. The directly optimized operators as well as those indirectly optimized factors using HLPSO are listed in Table 1.

The optimized network depth shown in Table 1 is shared by the three different convolutional blocks, i.e. the optimized CNN model has the same depth of three different types of convolutional blocks. Based on the identified optimal network depth, the initial number of filters, σ, for convolutional block 1 is calculated using Equation (24).

$$\sigma = round\left(\frac{C}{\sqrt{BlockDepth}}\right) \tag{24}$$

where $BlockDepth$ denotes the optimized depth of each convolutional block and $C = 32$ is a constant variable to determine the overall network flexibility and the size of the network parameters. The $BlockDepth$ variable determines the depth of the networks. Specifically, since the network adopts the same depth, i.e. $BlockDepth$, for three different types of convolutional blocks, the total number of convolutional layers in the network is $BlockDepth \times 3$. Moreover, in this study, we intend to make the number of filters in each convolutional layer proportional to $\frac{1}{\sqrt{BlockDepth}}$. In this way, the networks with different optimized block depths may have a similar number of network parameters and require similar computational cost in each iteration. To enhance network flexibilities, convBlock 1 utilizes σ number of filters, while convBlock 2 and convBlock 3 employ comparatively higher numbers of filters, i.e. 2σ and 4σ numbers of filters, respectively. A larger setting of the constant variable C can also be employed to obtain a larger initial number of filters σ to increase the network flexibility and the number of network parameters. As shown in Table 1, the search of the optimized block depth, $BlockDepth$, has the search range of [1, 3]. This search range can be further extended to generate deeper networks.

HLPSO is used to optimize the factors shown in Table 1. Each particle represents a possible CNN structure and the two recommended hyper-parameters. We split each skin lesion data set into 60:20:20 as the training, validation, and test sets. The fitness evaluation of each particle is calculated using the confusion matrix measure (i.e. the mean of diagonal results in the confusion matrix) of each identified CNN model for the validation set. The setting of HLPSO-based deep network generation is as follows, i.e. dimension=3 (learning rate, weight decay, and depth of each convolutional block), and the maximum number of function evaluations=population (30) × iterations (10). As illustrated in Table 1, the search boundaries for each optimized element are provided. As an example, the hyper-parameters, such as learning rate and weight decay, are assigned to the search ranges of [1e-4, 5e-1] and [1e-5, 5e-2], respectively. The block depth has the search range of [1, 3]. The initial number of filters (i.e. the number of filters for convBlock 1) is generated using Equation (24) based on the optimized block depth. Therefore, the number of filters for convBlock 1 is in the range of [18, 32], while the numbers of filters for convBlock 2 and convBlock 3 are 2 and 4 times of those of convBlock 1, respectively. As such, the numbers of filters for convBlock 2 and convBlock 3 are in the ranges of [36, 64] and [72, 128], respectively.

Owing to complexity of the training process, we resize the input image to 32×32 to achieve the best trade-off between performance and computational efficiency. The generated evolving CNNs are compared with those devised by other search methods and the default structure shown in Figure 4 (where network depth=1 for each block, learning rate=0.01 (MATLAB default setting) and regularization factor=0.0001 (MATLAB default setting)). Besides the abovementioned optimized explicit factors, millions of weights associated with different depths of the three types of convolutional layers and their associated other learning layers, and the fully connected layer are also optimized. We discuss the detailed evaluation results of the optimized deep CNNs using two skin lesion data sets in Section 7.

6. FEATURE SELECTION & ENSEMBLE LESION CLASSIFICATION

Besides the first deep CNN based lesion classification system, the second system employs HLPSO-based feature selection and ensemble classifiers for lesion classification. After segmenting the skin lesions, clinically important shape and colour features are extracted from the lesion and skin regions according to the Asymmetry, Border, Colour, Diameter, and Enlargement (ABCDE) guideline. A total of 14 shape and 133 colour features are extracted. The shape features include area, extent, solidity, perimeter, asymmetry index, compactness index, border irregularity index, etc, while the colour features include colour ratio, entropy, variance, Principal

Component Analysis variance, Tamura coarseness, differences in lightness, Chroma, colour and Hue, etc [32]. A new texture descriptor, combining Kirsch Compass Masks [55] with a LBP variant, i.e. SLBP [56], is also used for feature extraction. The Kirsch compass kernel is a non-linear edge detector, which identifies the maximum edge strength in eight pre-determined directions, namely north, north west, west, south west, south, south east, east, and north east. Specifically, it takes a single mask and rotates it to the above eight major compass orientations at 45-degree intervals. The Kirsch operator is applied to the segmented lesion regions to extract the edges through all eight compass directions. The edge magnitude of the Kirsch operator is calculated as the maximum magnitude across all directions [55].

The Kirsch operator is able to provide additional edge information for the segmented lesions, in order to further improve lesion boundaries. Texture features are subsequently extracted from both the segmented lesions and edge responses generated by the Kirsch operator (see Figure 5).

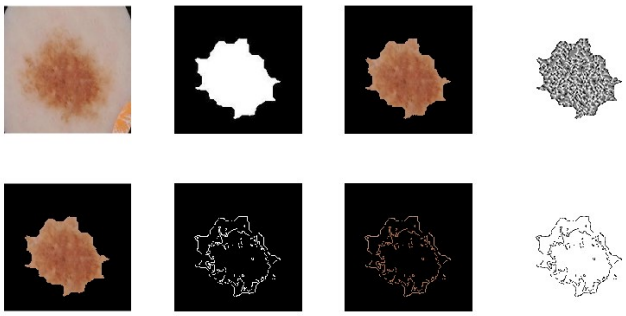

Figure 5 Pipeline processing for texture feature extraction (Top: 1. the original image, 2. GT (or generated) mask, 3. segmented lesion sub-image, 4. extracted texture features for the whole segmented lesion region. Bottom: 1. segmented lesion sub-image, 2. generated Kirsch compass masks, 3. edge response produced by the Kirsch operator, 4. extracted texture features for the detected edges.)

We employ SLBP for texture description. While the original LBP operator is efficient in extracting local texture features, it tends to lose neighbourhood contrast and global discriminative information. The SLBP descriptor is able to generate a number of binary string outputs for each pixel position to overcome such limitation. It is therefore more efficient in detecting fuzzy characteristics of the pixel neighbourhood by generating a distribution of binary patterns. It is used in this research to extract texture features from the segmented lesions, as well as edge responses generated by the Kirsch operator. These two texture feature vectors are concatenated to form the final texture outputs. Figure 5 shows the pipeline processing for texture generation. A texture feature vector of 1062 dimensions is obtained for each lesion image.

After extracting a variety of shape, colour and texture features, HLPSO is used to identify the most significant characteristics from each feature vector. The cost function for discriminative feature selection is defined in Equation (25) [21, 25]. Both classification performance and the number of selected features are taken into account in the fitness evaluation to measure the significance of each selected feature subset. In this case, we employ Geometric Mean (GM) as the performance indicator for discriminative feature selection. HLPSO is used to maximize the classification performance and minimize the number of selected features.

$$f(x) = w_{GM} \times GM_x + w_f \times (number\_feature_x)^{-1} \qquad (w_f = 1 - w_{GM}) \tag{25}$$

where w_{GM} and w_f denote the weights for GM performance and the number of selected features, respectively. Since classification performance has a higher priority in fitness evaluation, w_{GM} has a higher score (e.g. 0.9) than that (e.g. 0.1) of w_f.

HLPSO-based feature selection is conducted separately for shape and colour as well as texture features. Two ensemble classifiers are formulated for benign/melanoma lesion classification. Each ensemble classifier consists of two base models dedicated to the optimized shape and colour as well as texture features, respectively. We employ KNN and SVM as the base classifiers in the two ensemble models, respectively. To ensure a fair comparison, the default settings of the KNN and SVM classifiers are used for the evaluation of the optimized feature subsets identified by each search method. The detailed evaluation results for ensemble lesion classification are discussed in Section 7.

7. EXPERIMENTAL STUDIES

We employ two dermoscopic skin lesion data sets for evaluation of the CNN and ensemble classification models. The first data set is a combined data set which consists of dermoscopic lesion images extracted from two well-known databases, i.e. Dermofit Image Library [57] and PH2 [58]. It is composed of 270 benign and 214 melanoma lesion images. The benign subset comprises 190 and 80 images from Dermofit Image Library and PH2, whereas the malignant subset comprises 174 and 40 images from Dermofit Image Library and PH2 respectively. The second data set is the ISIC 2017 data set [59]. A total of 550 (275 benign and 275 melanoma) images from ISIC 2017 are randomly selected for evaluating lesion segmentation, while a subset of 400 (200 benign and 200 melanoma) images with reasonable Jaccard scores for lesion segmentation is selected for evaluation of the two proposed systems in lesion classification.

Different PSO and FA variants, i.e. Enhanced Leader PSO (ELPSO) [60], Autonomous Group PSO (AGPSO) [61], Dynamic Neighbourhood Learning PSO (DNLPSO) [62], Genetic PSO (GPSO) [63], GMPSO [21], FA with neighbourhood attraction (NaFA) [64], CFA1 [22], CFA2 [65], a modified FA (MFA) [66], FA with a variable step size (VSSFA) [67], as well as classical search methods, i.e. PSO, FA and Moth-Flame Optimization (MFO) [51], are implemented for comparison. We inherit the parameter settings of these classical methods and PSO and FA variants from their original studies, as shown in Table 2. The setting of our model is identified via trial and error to achieve the best trade-off between model performance and computational efficiency.

Owing to the variations and complexities of the problem domains, search dimensions, and model convergence speed, we have adopted different experimental settings for the following four optimization tasks, i.e. (1) basic and complex benchmark functions, (2) centroid enhancement for image segmentation, (3) deep architecture generation and (4) discriminative feature selection for ensemble classification. In addition, to ensure a fair comparison, we have employed the same number of Maximum Fitness Evaluations (MaxFEs) as the termination criterion for all the methods in each optimization task. In other words, all the algorithms terminate when the number of MaxFEs is reached under each experimental setting.

Table 2 Parameter settings of different methods

Method	Parameter setting
PSO [19]	maximum velocity=0.6, inertia weight=0.5, acceleration constants c_1= c_2= 1.5
FA [20]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.5
MFO [51]	Use adaptive parameter settings
ELPSO [60]	c_1=c_2=2, standard deviation of Gaussian mutation=1, scale parameter of Cauchy mutation=2, scale factor of DE-based mutation=1.2, inertia weight=0.9−(0.9−0.4)×(k−1)/(MaxGeneration−1), where k and MaxGeneration represent the current and maximum iteration numbers, respectively.
AGPSO [61]	maximum velocity=0.6, adaptive decreasing c_1 and increasing c_2 over generations, inertia weight=0.9−(0.9−0.4)×(k)/(MaxGeneration), where k and MaxGeneration represent the current and maximum iteration numbers, respectively.
DNLPSO [62]	c_1=c_2=1.49445, refreshing gap=3, regrouping period=5, inertia weight=0.9−(0.9−0.4)×(k−1)/(MaxGeneration−1), where k and MaxGeneration represent the current and maximum iteration numbers, respectively.
GPSO [63]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1 = 2.6, c_2 = 1.5
GMPSO [21]	maximum velocity=0.6, inertia weight=0.5, acceleration constants c_1 = c_2 = 1.5, standard deviation of Gaussian distribution=1, scaling factor of Cauchy distribution=2, crossover probability=0.6, mutation probability=0.05
NaFA [64]	attractiveness=$(1-0.2)\times e^{-\gamma r_{ij}^2}+0.2$, absorption coefficient=1.0, Levy's index=1.5, and randomization parameter $\alpha(t+1)=\left(\frac{1}{9000}\right)^{\frac{1}{t}}\times\alpha(t)$ where t is the current iteration.
CFA1 [22]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=Logistic map
CFA2 [65]	attractiveness=Gauss map, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.5
MFA [66]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.5
VSSFA [67]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.4/(1+exp(0.015*(k-MaxGeneration)/3)), where k and MaxGeneration represent the current and maximum iteration numbers, respectively.
HLPSO	maximum velocity=0.6, inertia weight=0.65, acceleration constants c_1=c_2=2.5, attractiveness=$(1-0.2)\times e^{-\gamma r_{ij}^2}+0.2$, absorption coefficient=1.0, Levy's index=1.5, and randomization parameter $\alpha'=\alpha'\times\left(\frac{10^{-4}}{0.9}\right)^{\frac{1}{\text{MaxGeneration}}}$

7.1 Evaluation Using Classical Benchmark Functions and CEC 2014 Test Suite

Before performing image segmentation, deep architecture generation, and feature selection for ensemble lesion classification, we first employ a set of 11 classic benchmark functions and the well-known CEC 2014 test suite [68] for evaluation, in order to ascertain the efficiency of the proposed HLPSO model.

Firstly, the selected 11 classical benchmark functions are defined in Table 3, which represent challenging artificial landscapes with varied difficulties. They include unimodal mathematical landscapes with single global minima such as F2, and F5-F10, as well as multimodal benchmark functions with multiple local minima such as F1, F3, F4, and F11. The following experimental settings are employed for model evaluation, i.e. the number of MaxFEs=25,000, dimension=30 and trials=30. All the methods terminate when reaching the number of MaxFEs. Table 4 illustrates the detailed empirical and Wilcoxon rank sum test results [33], where mean, min, max, std and RS denote the mean, minimum, maximum, standard deviation, and rank sum test results, respectively. The symbols of '+', '=' and '-' are used to indicate whether HLPSO is significantly better than, the same, or worse than other baseline methods based on the Wilcoxon rank sum test. Similar symbols are also used in subsequent tables for performance analysis.

Table 3 Unimodal and multimodal benchmark functions

	Name	Formula	Range
F1	Ackley	$f(x) = -a\,exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d}x_i^2}\right) - exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)\right) + a + exp(1)$ $a = 20$, $b = 0.2$ and $c = 2\pi$	[-15, 30]
F2	Dixon-Price	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{d} i\,(2x_i^2 - x_{i-1})^2$	[-10, 10]
F3	Griewank	$f(x) = \sum_{i=1}^{d}\frac{x_i^2}{4000} - \prod_{i=1}^{d} cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	[-600, 600]
F4	Rastrigin	$f(x) = 10d + \sum_{i=1}^{d}[x_i^2 - 10\cos(2\pi x_i)]$	[-5.12, 5.12]
F5	Rotated Hyper-Ellipsoid	$f(x) = \sum_{i=1}^{d}\sum_{j=1}^{i} x_j^2$	[-65, 65]
F6	Rosenbrock	$f(x) = \sum_{i=1}^{d-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	[-5,10]
F7	Sphere	$f(x) = \sum_{i=1}^{d} x_i^2$	[-5.12, 5.12]
F8	Sum of Different Powers	$f(x) = \sum_{i=1}^{d}\lvert x_i\rvert^{i+1}$	[-1, 1]
F9	Zakharov	$f(x) = \sum_{i=1}^{d} x_i^2 + \left(\sum_{i=1}^{d} 0.5ix_i\right)^2 + \left(\sum_{i=1}^{d} 0.5ix_i\right)^4$	[-5, 10]
F10	Sum Squares	$f(x) = \sum_{i=1}^{d} ix_i^2$	[-5.12, 5.12]
F11	Powell	$f(x) = \sum_{i=1}^{d/4}[(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4]$	[-4, 5]

Table 4 Evaluation and statistical test results for basic benchmark functions with dimension=30

		HLPSO	AGPSO	GPSO	GMPSO	DNLPSO	ELPSO	MFA	NaFA	CFA1	CFA2	VSSFA	PSO	FA	MFO
Ackley	**MEAN**	**9.18E-06**	2.32E+00	1.77E+01	1.38E+00	3.06E+00	1.47E+01	2.02E+01	8.71E-03	1.58E+01	1.47E+01	1.04E+01	6.21E+00	4.09E-02	8.85E+00
	MIN	**3.02E-06**	1.34E+00	1.67E+01	5.41E-04	3.33E-02	1.26E+01	2.02E+01	6.82E-03	1.48E+01	1.38E+01	9.53E+00	3.30E+00	2.02E-02	1.56E-01
	MAX	**2.18E-05**	3.29E+00	1.84E+01	3.40E+00	9.56E+00	1.61E+01	2.02E+01	1.15E-02	1.66E+01	1.54E+01	1.11E+01	9.59E+00	8.16E-02	1.50E+01
	STD	5.44E-06	4.58E-01	4.53E-01	8.37E-01	2.42E+00	8.88E-01	**3.79E-15**	1.05E-03	4.27E-01	4.34E-01	3.15E-01	1.48E+00	1.46E-02	5.14E+00

	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Dixon	**MEAN**	**1.80E-01**	1.51E+01	5.11E+05	7.13E-01	2.83E+01	1.60E+05	1.62E+06	3.00E+00	1.53E+05	1.34E+05	1.07E+04	1.94E+00	1.93E+00	3.51E+04
	MIN	**1.13E-04**	6.67E-01	2.24E+05	6.67E-01	6.67E-01	6.33E+04	1.62E+06	6.70E-01	9.16E+04	9.07E+04	4.59E+03	6.79E-01	6.89E-01	1.93E+00
	MAX	2.49E+00	3.21E+02	8.56E+05	**1.61E+00**	3.66E+02	2.50E+05	1.62E+06	1.61E+01	1.93E+05	1.75E+05	1.54E+04	4.90E+00	1.00E+01	6.27E+05
	STD	5.25E-01	5.97E+01	1.50E+05	1.77E-01	7.24E+01	4.03E+04	**2.12E-10**	3.93E+00	2.97E+04	1.91E+04	2.84E+03	1.51E+00	1.99E+00	1.21E+05
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Griewank	**MEAN**	5.91E-03	2.70E-02	2.85E+02	1.67E-02	4.22E+00	1.35E+02	6.08E+02	**3.31E-03**	1.71E+02	1.52E+02	4.37E+01	2.33E-01	5.60E-03	2.18E+01
	MIN	**5.68E-12**	1.65E-04	1.21E+02	8.40E-09	1.37E-03	7.71E+01	6.08E+02	1.89E-03	1.23E+02	1.18E+02	3.18E+01	2.76E-02	3.64E-03	6.04E-01
	MAX	2.95E-02	1.35E-01	4.01E+02	6.82E-02	6.14E+01	1.76E+02	6.08E+02	**5.98E-03**	2.12E+02	1.82E+02	5.24E+01	1.21E+00	7.80E-03	9.10E+01
	STD	8.15E-03	2.96E-02	5.59E+01	2.11E-02	1.23E+01	2.49E+01	**2.36E-13**	1.00E-03	2.08E+01	1.62E+01	4.19E+00	2.74E-01	1.11E-03	3.87E+01
	RS	n/a	+	+	+	+	+	+	=	+	+	+	+	=	+
Rastrigin	**MEAN**	**1.69E+01**	5.55E+01	3.55E+02	4.05E+01	9.77E+01	2.71E+02	4.29E+02	2.87E+01	2.76E+02	2.67E+02	2.12E+02	5.53E+01	2.40E+01	1.48E+02
	MIN	**3.98E+00**	2.21E+01	3.14E+02	1.50E+01	2.86E+01	2.33E+02	4.29E+02	1.69E+01	2.40E+02	2.41E+02	1.94E+02	2.40E+01	1.22E+01	7.17E+01
	MAX	**2.89E+01**	1.10E+02	3.85E+02	6.78E+01	1.79E+02	3.08E+02	4.29E+02	4.58E+01	3.11E+02	2.87E+02	2.28E+02	1.13E+02	4.09E+01	2.23E+02
	STD	5.95E+00	1.87E+01	1.86E+01	1.22E+01	4.03E+01	1.86E+01	**1.31E-13**	6.53E+00	1.42E+01	1.11E+01	9.95E+00	1.70E+01	6.58E+00	3.98E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Rothyp	**MEAN**	**9.26E-09**	6.66E-03	2.01E+05	2.78E-06	2.66E+02	9.68E+04	4.38E+05	5.50E+01	1.07E+05	1.00E+05	2.98E+04	5.68E+02	9.04E+00	1.51E+04
	MIN	**7.67E-10**	1.00E-04	1.29E+05	2.34E-08	9.43E-04	6.20E+04	4.38E+05	1.85E-01	8.62E+04	7.23E+04	2.15E+04	7.21E-03	2.20E-01	4.37E+00
	MAX	**3.19E-08**	5.61E-02	2.81E+05	2.65E-05	3.21E+03	1.20E+05	4.38E+05	4.10E+02	1.30E+05	1.15E+05	3.82E+04	1.69E+04	8.74E+01	6.34E+04
	STD	8.06E-09	1.28E-02	3.80E+04	5.13E-06	6.35E+02	1.53E+04	**2.09E-10**	9.00E+01	1.13E+04	1.03E+04	4.24E+03	3.08E+03	1.85E+01	1.76E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Rosenbrock	**MEAN**	**2.35E+01**	1.56E+02	3.70E+05	4.41E+01	1.19E+02	1.22E+05	2.84E+06	6.65E+01	1.19E+05	7.63E+04	7.92E+03	7.60E+01	5.02E+01	6.62E+04
	MIN	**2.68E-02**	5.91E+00	1.28E+05	1.35E+01	2.41E+01	2.03E+04	2.84E+06	2.59E+01	3.61E+04	3.99E+04	3.49E+03	2.79E+00	2.73E+01	3.15E+01
	MAX	1.37E+02	2.53E+03	6.34E+05	**9.65E+01**	4.20E+02	1.97E+05	2.84E+06	2.10E+02	1.70E+05	1.10E+05	1.10E+04	3.24E+02	1.28E+02	2.01E+05
	STD	3.33E+01	4.49E+02	1.21E+05	2.72E+01	9.50E+01	3.95E+04	**5.74E-10**	5.58E+01	2.87E+04	1.63E+04	1.84E+03	7.59E+01	3.69E+01	5.81E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Sphere	**MEAN**	**3.58E-10**	1.56E-06	8.31E+01	2.23E-09	5.65E-01	4.20E+01	1.77E+02	6.88E-06	5.00E+01	4.29E+01	1.23E+01	2.46E-02	1.41E-03	3.50E+00
	MIN	**2.80E-11**	1.29E-08	4.09E+01	4.70E-11	4.68E-10	2.43E+01	1.77E+02	4.37E-06	3.82E+01	3.05E+01	9.32E+00	7.56E-03	2.19E-04	2.36E-04
	MAX	**2.09E-09**	2.10E-05	1.08E+02	2.77E-08	1.13E+01	5.17E+01	1.77E+02	1.23E-05	6.13E+01	5.39E+01	1.49E+01	8.37E-02	2.80E-03	2.62E+01
	STD	4.66E-10	3.95E-06	1.46E+01	5.04E-09	2.11E+00	7.24E+00	**2.24E-14**	2.01E-06	6.25E+00	6.31E+00	1.44E+00	1.56E-02	7.17E-04	9.06E+00
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Sumpow	**MEAN**	**3.36E-24**	7.05E-17	1.37E-01	1.19E-13	1.71E-08	1.29E-02	5.82E-01	1.09E-07	9.38E-03	6.99E-03	2.09E-04	1.73E-05	2.21E-07	2.51E-11
	MIN	3.53E-28	1.06E-22	2.43E-02	6.59E-15	**1.13E-30**	2.64E-04	5.82E-01	7.10E-09	1.21E-03	6.90E-04	3.36E-05	1.83E-06	3.01E-08	2.83E-15
	MAX	**3.80E-23**	2.04E-15	4.85E-01	3.75E-13	1.83E-07	3.66E-02	5.82E-01	4.41E-07	2.39E-02	1.43E-02	4.45E-04	1.09E-04	6.74E-07	1.54E-10
	STD	**8.61E-24**	3.73E-16	1.22E-01	1.09E-13	4.37E-08	9.36E-03	1.34E-16	1.10E-07	4.77E-03	2.90E-03	1.01E-04	2.09E-05	1.61E-07	4.28E-11
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Zakharov	**MEAN**	**1.61E+01**	7.15E+01	4.72E+02	4.10E+01	1.29E+02	3.39E+02	8.89E+02	3.56E+01	3.76E+02	3.28E+02	2.41E+02	1.35E+02	2.50E+01	1.99E+02
	MIN	**7.03E+00**	2.60E+01	4.35E+02	2.12E+01	4.84E+01	2.92E+02	8.89E+02	1.49E+01	3.50E+02	2.88E+02	2.07E+02	8.38E+01	1.43E+01	1.08E+02
	MAX	**2.98E+01**	1.26E+02	5.04E+02	8.95E+01	2.58E+02	3.88E+02	8.89E+02	6.37E+01	3.99E+02	3.59E+02	2.65E+02	2.63E+02	5.22E+01	2.95E+02
	STD	6.44E+00	2.55E+01	1.89E+01	1.43E+01	6.00E+01	2.16E+01	**3.33E-13**	1.27E+01	1.55E+01	1.83E+01	1.28E+01	4.10E+01	7.51E+00	4.84E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Sumsqu	**MEAN**	**5.08E-09**	8.74E-01	1.09E+03	7.56E-08	1.72E+01	5.52E+02	2.74E+03	2.16E-01	6.53E+02	6.19E+02	1.82E+02	2.49E-01	5.18E-01	7.18E+01
	MIN	**8.63E-11**	1.31E-06	7.75E+02	1.19E-10	4.15E-05	2.50E+02	2.74E+03	2.18E-03	4.46E+02	3.92E+02	1.34E+02	1.34E-02	4.57E-02	1.17E-02
	MAX	**2.57E-08**	2.62E+01	1.56E+03	1.13E-06	2.86E+02	7.71E+02	2.74E+03	1.60E+00	8.89E+02	7.41E+02	2.22E+02	5.16E+00	1.51E+00	2.90E+02
	STD	6.37E-09	4.79E+00	1.89E+02	2.20E-07	6.13E+01	1.13E+02	**5.47E-13**	3.38E-01	9.75E+01	7.65E+01	1.98E+01	9.35E-01	4.51E-01	7.62E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
Powell	**MEAN**	**2.03E-03**	1.75E+01	4.22E+03	3.79E-03	3.66E+01	1.50E+03	8.55E+03	2.50E+00	1.90E+03	1.43E+03	3.04E+02	8.81E-02	4.88E+00	5.17E+02
	MIN	**4.95E-04**	3.69E-03	1.59E+03	8.22E-04	1.25E-03	4.84E+02	8.55E+03	6.84E-02	1.03E+03	5.04E+02	2.04E+02	1.36E-02	1.43E+00	1.35E+00
	MAX	**5.10E-03**	1.04E+02	5.74E+03	1.40E-02	9.17E+02	3.19E+03	8.55E+03	6.66E+00	2.37E+03	1.89E+03	4.03E+02	8.87E-01	1.61E+01	3.42E+03
	STD	1.16E-03	3.64E+01	9.80E+02	2.78E-03	1.66E+02	6.75E+02	**1.01E-12**	1.71E+00	3.23E+02	3.07E+02	4.63E+01	1.59E-01	3.50E+00	7.95E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+

As shown in Table 4, HLPSO shows great efficiency in solving these unimodal and multimodal benchmark functions, and outperforms the advanced PSO and FA variants and classical search methods in most of the test cases significantly. The exception is for Griewank, where HLPSO shows similar result distributions to those of NaFA and FA over a set of 30 runs.

Secondly, the CEC 2014 test suite has been employed to further ascertain the model efficiency. This test suite represents various complex minimization problems, and contains a total of 30 test functions including 3 rotated unimodal, 13 shifted and rotated multimodal, 6 hybrid and 8 composition functions with the search range of [-100, 100]. We have employed the following experimental settings for evaluation, i.e. the number of MaxFEs=25,000, dimension=30 and trials=30. All the methods terminate when the number of MaxFEs is reached. Table 5 tabulates the detailed empirical and statistical test results.

Table 5 Evaluation and statistical test results for CEC 2014 benchmark functions with dimension=30

		HLPSO	AGPSO	GPSO	GMPSO	DNLPSO	ELPSO	MFA	NaFA	CFA1	CFA2	VSSFA	PSO	FA	MFO
F1	**MEAN**	**5.27E+02**	6.21E+06	6.17E+08	3.75E+06	4.63E+06	1.30E+09	1.18E+07	1.84E+07	4.76E+08	2.54E+07	2.09E+07	6.17E+08	8.38E+06	7.65E+07
	MIN	**4.88E+02**	6.36E+05	2.92E+08	2.59E+05	3.71E+05	8.28E+08	4.88E+06	5.76E+06	1.51E+08	4.67E+06	3.01E+06	3.15E+08	3.75E+06	7.79E+06
	MAX	**1.73E+03**	1.99E+07	8.28E+08	2.33E+07	1.55E+07	1.63E+09	2.64E+07	3.17E+07	6.75E+08	1.49E+08	8.40E+07	1.04E+09	2.80E+07	4.05E+08
	STD	**4.01E+00**	5.38E+06	1.18E+08	4.14E+06	3.43E+06	2.37E+08	5.33E+06	7.04E+06	1.14E+08	2.76E+07	1.89E+07	2.01E+08	4.34E+06	8.14E+07
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F2	**MEAN**	**2.84E+02**	5.04E+03	5.22E+10	2.66E+04	4.11E+07	6.25E+10	5.90E+06	9.56E+04	2.80E+10	7.10E+06	7.31E+08	5.04E+10	2.42E+04	3.95E+09
	MIN	**2.01E+02**	4.58E+02	3.30E+10	4.89E+02	2.82E+02	4.97E+10	6.53E+05	6.60E+04	2.14E+10	6.28E+05	7.33E+03	2.38E+10	1.14E+04	1.08E+06
	MAX	**3.92E+02**	1.95E+04	6.95E+10	5.12E+05	6.07E+08	7.14E+10	2.47E+07	1.47E+05	3.27E+10	2.26E+07	1.08E+10	7.19E+10	4.37E+04	2.00E+10
	STD	**3.99E-02**	4.58E+03	9.60E+09	9.19E+04	1.33E+08	4.91E+09	5.40E+06	2.02E+04	2.99E+09	5.43E+06	2.05E+09	1.15E+10	8.39E+03	6.17E+09
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F3	**MEAN**	**3.99E+02**	1.14E+04	2.03E+05	1.12E+04	2.61E+04	1.24E+05	7.98E+04	5.78E+04	8.19E+04	3.32E+04	1.35E+05	2.41E+05	3.12E+04	7.37E+04
	MIN	**3.57E+02**	6.25E+02	1.47E+05	4.59E+02	1.84E+03	6.72E+04	1.26E+04	3.46E+04	3.46E+04	1.43E+04	2.49E+04	1.17E+05	1.49E+04	1.40E+04
	MAX	**4.21E+02**	6.36E+04	2.74E+05	6.33E+04	7.68E+04	1.80E+05	1.68E+05	8.81E+04	1.18E+05	5.58E+04	8.23E+05	3.55E+05	5.26E+04	1.64E+05
	STD	**4.94E-02**	1.27E+04	2.94E+04	1.39E+04	2.06E+04	2.23E+04	4.03E+04	1.14E+04	2.05E+04	1.15E+04	1.49E+05	6.22E+04	9.46E+03	4.64E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F4	**MEAN**	**4.06E+02**	4.63E+02	7.59E+03	4.49E+02	5.43E+02	9.86E+03	5.51E+02	4.51E+02	3.48E+03	6.09E+02	6.92E+02	9.58E+03	4.28E+02	8.20E+02
	MIN	4.00E+02	4.17E+02	3.85E+03	**4.00E+02**	4.23E+02	6.27E+03	4.60E+02	4.41E+02	1.99E+03	4.60E+02	4.46E+02	5.52E+03	4.25E+02	4.32E+02
	MAX	**4.10E+02**	5.51E+02	1.29E+04	4.92E+02	2.17E+03	1.15E+04	6.70E+02	5.04E+02	4.71E+03	9.03E+02	1.99E+03	1.29E+04	4.29E+02	2.80E+03
	STD	2.27E+00	3.89E+01	2.25E+03	2.90E+01	3.20E+02	1.31E+03	5.15E+01	1.51E+01	7.47E+02	8.60E+01	3.32E+02	2.13E+03	**8.83E-01**	4.99E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F5	**MEAN**	**5.15E+02**	5.20E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.20E+02	5.21E+02	5.20E+02	5.21E+02	5.21E+02	5.19E+02	5.20E+02
	MIN	**5.00E+02**	5.20E+02	5.21E+02	5.20E+02	5.20E+02	5.21E+02	5.21E+02	5.20E+02	5.21E+02	5.20E+02	5.20E+02	5.21E+02	5.00E+02	5.20E+02

	MAX	5.21E+02	5.20E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	5.21E+02	**5.20E+02**	5.21E+02
	STD	8.81E+00	1.17E-01	**4.79E-02**	1.63E-01	3.24E-01	5.45E-02	7.20E-02	1.53E-01	5.72E-02	1.19E-01	2.07E-01	6.69E-02	3.63E+00	1.80E-01
	RS	n/a	+	+	+	+	+	+	+	+	+	=	+	+	+
F6	**MEAN**	6.01E+02	6.19E+02	6.40E+02	6.14E+02	6.14E+02	6.68E+02	6.34E+02	6.03E+02	6.38E+02	6.41E+02	6.37E+02	6.39E+02	**6.01E+02**	6.21E+02
	MIN	**6.00E+02**	6.14E+02	6.34E+02	6.06E+02	6.06E+02	6.62E+02	6.26E+02	6.01E+02	6.36E+02	6.32E+02	6.23E+02	6.35E+02	6.01E+02	6.15E+02
	MAX	6.04E+02	6.25E+02	6.43E+02	6.22E+02	6.34E+02	6.72E+02	6.46E+02	6.06E+02	6.41E+02	6.50E+02	6.61E+02	6.44E+02	**6.02E+02**	6.30E+02
	STD	1.03E+00	2.74E+00	2.13E+00	3.37E+00	6.05E+00	2.16E+00	5.53E+00	1.48E+00	1.35E+00	4.67E+00	9.45E+00	2.39E+00	**7.28E-01**	3.67E+00
	RS	n/a	+	+	+	+	+	+	+	=	+	+	+	-	+
F7	**MEAN**	**7.00E+02**	7.00E+02	1.13E+03	7.00E+02	7.01E+02	1.29E+03	7.01E+02	7.00E+02	9.39E+02	7.01E+02	7.19E+02	1.17E+03	7.00E+02	7.32E+02
	MIN	7.00E+02	7.00E+02	8.63E+02	7.00E+02	**7.00E+02**	1.11E+03	7.01E+02	7.00E+02	8.73E+02	7.00E+02	7.00E+02	8.50E+02	7.00E+02	7.01E+02
	MAX	**7.00E+02**	7.00E+02	1.24E+03	7.00E+02	7.25E+02	1.40E+03	7.01E+02	7.00E+02	1.01E+03	7.01E+02	8.46E+02	1.51E+03	7.00E+02	9.37E+02
	STD	7.79E-02	5.40E-02	8.12E+01	**1.72E-02**	4.70E+00	6.68E+01	8.13E-02	4.28E-02	3.41E+01	2.03E-01	3.63E+01	1.45E+02	2.93E-02	5.75E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F8	**MEAN**	**8.06E+02**	8.56E+02	1.15E+03	8.26E+02	8.68E+02	1.31E+03	8.87E+02	8.45E+02	1.07E+03	9.50E+02	9.48E+02	1.15E+03	8.26E+02	9.47E+02
	MIN	**8.02E+02**	8.27E+02	1.10E+03	8.10E+02	8.20E+02	1.25E+03	8.46E+02	8.32E+02	1.03E+03	8.99E+02	8.49E+02	1.08E+03	8.16E+02	8.64E+02
	MAX	**8.15E+02**	8.87E+02	1.20E+03	8.58E+02	9.46E+02	1.34E+03	9.55E+02	8.72E+02	1.11E+03	1.03E+03	1.11E+03	1.22E+03	8.67E+02	1.03E+03
	STD	**2.88E+00**	1.50E+01	2.53E+01	1.20E+01	3.25E+01	2.12E+01	2.75E+01	1.09E+01	1.81E+01	3.34E+01	6.23E+01	3.45E+01	9.98E+00	3.88E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F9	**MEAN**	**9.09E+02**	1.00E+03	1.32E+03	9.75E+02	9.65E+02	1.51E+03	1.05E+03	9.49E+02	1.22E+03	1.09E+03	1.06E+03	1.31E+03	9.24E+02	1.11E+03
	MIN	**9.03E+02**	9.57E+02	1.24E+03	9.38E+02	9.29E+02	1.46E+03	9.75E+02	9.27E+02	1.16E+03	9.97E+02	9.58E+02	1.22E+03	9.15E+02	1.06E+03
	MAX	**9.26E+02**	1.05E+03	1.41E+03	1.03E+03	1.02E+03	1.56E+03	1.13E+03	9.86E+02	1.25E+03	1.16E+03	1.26E+03	1.41E+03	9.41E+02	1.18E+03
	STD	**4.42E+00**	2.37E+01	3.97E+01	2.07E+01	2.44E+01	2.29E+01	3.99E+01	1.19E+01	2.18E+01	3.73E+01	7.02E+01	4.22E+01	6.35E+00	3.77E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F10	**MEAN**	**1.05E+03**	2.79E+03	9.67E+03	2.91E+03	4.71E+03	1.51E+04	5.12E+03	4.58E+03	8.63E+03	5.46E+03	7.78E+03	9.83E+03	2.93E+03	3.59E+03
	MIN	**1.00E+03**	1.68E+03	8.70E+03	1.56E+03	2.21E+03	1.41E+04	2.62E+03	2.74E+03	7.36E+03	3.92E+03	3.97E+03	9.00E+03	1.44E+03	2.60E+03
	MAX	**1.46E+03**	4.44E+03	1.05E+04	4.17E+03	9.01E+03	1.57E+04	7.44E+03	6.59E+03	9.19E+03	7.67E+03	1.52E+04	1.05E+04	4.24E+03	5.45E+03
	STD	**9.61E+01**	5.51E+02	4.85E+02	6.65E+02	2.04E+03	3.97E+02	1.26E+03	9.15E+02	3.22E+02	9.06E+02	3.07E+03	4.06E+02	6.73E+02	6.37E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F11	**MEAN**	**1.45E+03**	4.55E+03	8.80E+03	4.35E+03	6.35E+03	1.51E+04	7.99E+03	5.22E+03	8.59E+03	8.02E+03	1.16E+04	8.55E+03	3.01E+03	5.12E+03
	MIN	**1.10E+03**	2.93E+03	7.63E+03	3.31E+03	2.81E+03	1.41E+04	5.09E+03	3.23E+03	7.70E+03	6.15E+03	5.59E+03	7.77E+03	1.68E+03	3.68E+03
	MAX	**1.87E+03**	7.01E+03	9.46E+03	5.81E+03	8.92E+03	1.58E+04	9.29E+03	6.96E+03	8.98E+03	1.05E+04	1.57E+04	9.20E+03	4.46E+03	6.36E+03
	STD	**2.26E+02**	7.33E+02	3.72E+02	5.91E+02	1.84E+03	4.23E+02	1.08E+03	8.71E+02	3.35E+02	1.12E+03	3.27E+03	3.91E+02	6.82E+02	6.72E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F12	**MEAN**	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	**1.20E+03**	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03
	MIN	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	**1.20E+03**	1.20E+03
	MAX	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	**1.20E+03**	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03	1.20E+03
	STD	4.10E-01	3.28E-01	4.37E-01	6.42E-01	1.24E+00	4.20E-01	1.01E+00	**2.85E-02**	3.33E-01	4.28E-01	1.51E+00	3.49E-01	1.35E-01	1.67E-01
	RS	n/a	+	+	+	+	+	+	-	+	+	+	+	-	+
F13	**MEAN**	**1.30E+03**	1.30E+03	1.31E+03	1.30E+03	1.30E+03	1.31E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.31E+03	1.30E+03	1.30E+03
	MIN	**1.30E+03**	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03
	MAX	1.30E+03	1.30E+03	1.31E+03	1.30E+03	1.30E+03	1.31E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.30E+03	1.31E+03	**1.30E+03**	1.30E+03
	STD	6.84E-02	1.38E-01	6.01E-01	1.18E-01	1.35E-01	1.99E-01	9.19E-02	9.08E-02	2.96E-01	1.33E-01	4.51E-01	8.44E-01	**4.45E-02**	1.18E+00
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F14	**MEAN**	**1.40E+03**	1.40E+03	1.54E+03	1.40E+03	1.40E+03	1.57E+03	1.40E+03	1.40E+03	1.48E+03	1.40E+03	1.40E+03	1.55E+03	1.40E+03	1.42E+03
	MIN	**1.40E+03**	1.40E+03	1.49E+03	1.40E+03	1.40E+03	1.53E+03	1.40E+03	1.40E+03	1.46E+03	1.40E+03	1.40E+03	1.48E+03	1.40E+03	1.40E+03
	MAX	**1.40E+03**	1.40E+03	1.60E+03	1.40E+03	1.41E+03	1.60E+03	1.40E+03	1.40E+03	1.50E+03	1.40E+03	1.45E+03	1.64E+03	1.40E+03	1.48E+03
	STD	8.67E-02	2.09E-01	2.75E+01	1.33E-01	2.52E+00	1.61E+01	1.66E-01	1.03E-01	1.03E+01	2.84E-01	1.01E+01	3.75E+01	**8.49E-02**	2.32E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F15	**MEAN**	**1.50E+03**	1.52E+03	1.42E+06	1.51E+03	1.51E+03	6.78E+05	1.55E+03	1.51E+03	1.21E+05	1.66E+03	1.56E+03	2.35E+06	1.50E+03	1.10E+05
	MIN	**1.50E+03**	1.50E+03	1.17E+05	1.51E+03	1.50E+03	2.00E+05	1.52E+03	1.51E+03	3.26E+04	1.54E+03	1.51E+03	7.62E+04	1.50E+03	1.51E+03
	MAX	**1.50E+03**	1.54E+03	6.41E+06	1.53E+03	1.54E+03	1.19E+06	1.62E+03	1.51E+03	2.57E+05	3.93E+03	1.67E+03	1.03E+07	1.51E+03	2.43E+06
	STD	**5.40E-01**	9.08E+00	1.41E+06	6.04E+00	8.81E+00	2.09E+05	2.00E+01	1.62E+00	5.08E+04	4.31E+02	4.42E+01	2.05E+06	1.03E+00	4.41E+05
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F16	**MEAN**	**1.60E+03**	1.61E+03	1.61E+03	1.61E+03	1.61E+03	1.62E+03	1.62E+03	1.62E+03	1.61E+03	1.62E+03	1.62E+03	1.61E+03	1.61E+03	1.61E+03
	MIN	**1.60E+03**	1.61E+03	1.61E+03	1.61E+03	1.61E+03	1.62E+03	1.62E+03	1.62E+03	1.61E+03	1.62E+03	1.62E+03	1.61E+03	1.61E+03	1.61E+03
	MAX	**1.60E+03**	1.61E+03	1.61E+03	1.61E+03	1.61E+03	1.62E+03	1.62E+03	1.62E+03	1.61E+03	1.62E+03	1.62E+03	1.61E+03	1.61E+03	1.61E+03
	STD	4.96E-01	6.83E-01	**1.65E-01**	5.87E-01	7.65E-01	3.35E-01	7.13E-01	1.11E+00	1.93E-01	8.95E-01	1.14E+00	1.72E-01	8.10E-01	5.48E-01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F17	**MEAN**	**3.72E+03**	6.09E+05	1.87E+07	7.53E+05	4.29E+05	9.84E+07	2.12E+06	1.41E+06	2.01E+07	2.05E+06	1.79E+06	2.28E+07	5.48E+05	3.10E+06
	MIN	**1.95E+03**	6.93E+04	8.02E+06	2.39E+04	3.01E+04	4.77E+07	4.07E+05	1.94E+05	1.01E+07	2.88E+05	2.59E+05	5.31E+06	4.47E+04	1.20E+05
	MAX	**9.72E+03**	2.38E+06	3.88E+07	2.59E+06	1.49E+06	1.69E+08	6.03E+06	4.40E+06	3.89E+07	6.00E+06	5.84E+06	6.54E+07	1.45E+06	1.54E+07
	STD	**1.86E+03**	5.07E+05	7.82E+06	6.02E+05	3.70E+05	2.97E+07	1.35E+06	9.35E+05	7.85E+06	1.58E+06	1.43E+06	1.18E+07	4.16E+05	4.05E+06
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F18	**MEAN**	**2.33E+03**	5.87E+03	1.34E+09	4.28E+03	1.08E+04	4.85E+09	3.61E+03	4.77E+03	1.22E+09	5.43E+03	2.11E+07	1.67E+09	3.27E+03	1.64E+04
	MIN	**1.84E+03**	2.02E+03	2.23E+08	1.86E+03	1.98E+03	3.56E+09	1.98E+03	2.40E+03	5.42E+08	2.14E+03	2.35E+03	6.43E+08	2.09E+03	2.19E+03
	MAX	**3.89E+03**	1.65E+04	2.15E+09	1.44E+04	1.40E+05	6.25E+09	1.27E+04	1.29E+04	1.93E+09	2.08E+04	3.05E+08	2.72E+09	5.99E+03	2.86E+04
	STD	**6.30E+02**	4.41E+03	4.13E+08	3.71E+03	2.52E+04	7.07E+08	2.18E+03	2.71E+03	3.67E+08	4.36E+03	7.57E+07	5.49E+08	1.36E+03	1.12E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F19	**MEAN**	**1.90E+03**	1.92E+03	2.17E+03	1.91E+03	1.92E+03	2.37E+03	1.95E+03	1.93E+03	2.06E+03	1.96E+03	1.94E+03	2.27E+03	1.92E+03	1.94E+03
	MIN	**1.90E+03**	1.91E+03	2.09E+03	1.91E+03	1.91E+03	2.20E+03	1.92E+03	1.92E+03	2.02E+03	1.92E+03	1.92E+03	2.14E+03	1.91E+03	1.91E+03
	MAX	**1.90E+03**	2.00E+03	2.31E+03	1.92E+03	1.99E+03	2.50E+03	2.00E+03	1.93E+03	2.13E+03	2.09E+03	2.00E+03	2.52E+03	1.92E+03	2.01E+03
	STD	**5.81E-01**	2.19E+01	5.90E+01	3.08E+00	1.82E+01	7.86E+01	2.97E+01	2.49E+00	2.89E+01	4.11E+01	1.86E+01	8.73E+01	1.72E+00	3.57E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F20	**MEAN**	**2.06E+03**	1.10E+04	3.03E+05	1.82E+04	1.62E+04	1.97E+05	4.27E+04	2.26E+04	5.45E+04	2.94E+04	3.50E+04	4.83E+05	8.35E+03	4.36E+04
	MIN	**2.01E+03**	3.12E+03	3.78E+04	6.81E+03	2.79E+03	6.36E+04	1.42E+04	5.38E+03	1.56E+04	1.06E+04	5.57E+03	1.05E+05	3.52E+03	1.72E+04
	MAX	**2.14E+03**	2.84E+04	8.45E+05	4.14E+04	5.84E+04	5.06E+05	1.09E+05	7.34E+04	1.49E+05	5.20E+04	1.18E+05	1.06E+06	2.46E+04	8.45E+04
	STD	**2.86E+01**	6.08E+03	1.87E+05	9.67E+03	1.41E+04	1.03E+05	2.22E+04	1.32E+04	2.81E+04	1.20E+04	2.86E+04	2.87E+05	4.38E+03	1.53E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F21	**MEAN**	**2.79E+03**	2.60E+05	5.21E+06	3.88E+05	2.62E+05	4.58E+07	1.20E+06	1.11E+06	7.84E+06	8.45E+05	1.09E+06	5.14E+06	1.49E+05	1.63E+06
	MIN	**2.17E+03**	8.13E+03	1.61E+06	8.71E+03	2.24E+04	2.29E+07	1.47E+05	8.70E+04	2.46E+06	8.81E+04	1.54E+05	1.81E+06	1.02E+04	6.74E+04
	MAX	**6.80E+03**	9.98E+05	1.56E+07	2.41E+06	1.92E+06	7.49E+07	3.53E+06	4.51E+06	1.28E+07	2.39E+06	2.50E+06	1.16E+07	5.36E+05	1.42E+07
	STD	**8.89E+02**	2.12E+05	3.16E+06	5.06E+05	3.74E+05	1.24E+07	8.66E+05	1.05E+06	2.77E+06	5.79E+05	7.98E+05	2.60E+06	1.46E+05	2.62E+06
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F22	**MEAN**	**2.22E+03**	2.58E+03	3.73E+03	2.63E+03	2.55E+03	5.42E+03	3.20E+03	2.78E+03	3.43E+03	3.29E+03	3.26E+03	3.87E+03	2.33E+03	2.89E+03
	MIN	**2.20E+03**	2.26E+03	3.33E+03	2.23E+03	2.24E+03	4.88E+03	2.41E+03	2.26E+03	2.79E+03	2.72E+03	2.38E+03	3.43E+03	2.23E+03	2.26E+03
	MAX	**2.24E+03**	3.17E+03	4.31E+03	3.08E+03	3.45E+03	6.08E+03	4.22E+03	3.25E+03	3.79E+03	4.04E+03	4.43E+03	4.27E+03	2.71E+03	3.38E+03
	STD	**1.02E+01**	2.21E+02	2.40E+02	2.25E+02	2.57E+02	3.10E+02	3.47E+02	2.23E+02	2.12E+02	3.17E+02	5.55E+02	2.21E+02	1.33E+02	2.67E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F23	**MEAN**	**2.50E+03**	2.52E+03	2.97E+03	2.50E+03	2.50E+03	3.01E+03	2.50E+03	2.51E+03	2.77E+03	2.54E+03	2.54E+03	3.07E+03	2.50E+03	2.60E+03
	MIN	**2.49E+03**	2.50E+03	2.76E+03	2.50E+03	2.50E+03	2.80E+03	2.50E+03	2.50E+03	2.69E+03	2.50E+03	2.50E+03	2.83E+03	2.50E+03	2.55E+03
	MAX	**2.50E+03**	2.58E+03	3.15E+03	2.50E+03	2.55E+03	3.12E+03	2.50E+03	2.54E+03	2.86E+03	2.65E+03	2.63E+03	3.28E+03	2.52E+03	2.68E+03
	STD	5.23E+00	2.53E+01	9.18E+01	**6.08E-02**	9.59E+00	6.74E+01	1.03E+00	1.07E+01	3.99E+01	3.90E+01	3.44E+01	1.20E+02	4.79E+00	3.22E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F24	**MEAN**	**2.56E+03**	2.63E+03	2.80E+03	2.62E+03	2.63E+03	2.88E+03	2.67E+03	2.65E+03	2.73E+03	2.68E+03	2.68E+03	2.85E+03	2.61E+03	2.66E+03
	MIN	**2.54E+03**	2.61E+03	2.75E+03	2.61E+03	2.61E+03	2.86E+03	2.66E+03	2.65E+03	2.70E+03	2.66E+03	2.66E+03	2.79E+03	2.61E+03	2.61E+03
	MAX	**2.59E+03**	2.64E+03	2.84E+03	2.65E+03	2.65E+03	2.91E+03	2.70E+03	2.66E+03	2.75E+03	2.72E+03	2.77E+03	2.92E+03	2.62E+03	2.73E+03
	STD	1.29E+01	7.67E+00	2.42E+01	8.35E+00	1.03E+01	1.51E+01	1.02E+01	**2.19E+00**	1.10E+01	1.52E+01	2.58E+01	3.63E+01	2.59E+00	2.98E+01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F25	**MEAN**	**2.69E+03**	2.71E+03	2.75E+03	2.71E+03	2.70E+03	2.87E+03	2.72E+03	2.71E+03	2.76E+03	2.73E+03	2.72E+03	2.75E+03	2.70E+03	2.71E+03
	MIN	**2.67E+03**	2.71E+03	2.72E+03	2.70E+03	2.70E+03	2.81E+03	2.70E+03	2.70E+03	2.74E+03	2.72E+03	2.70E+03	2.73E+03	2.70E+03	2.71E+03
	MAX	**2.70E+03**	2.72E+03	2.79E+03	2.72E+03	2.71E+03	2.91E+03	2.73E+03	2.71E+03	2.77E+03	2.76E+03	2.77E+03	2.79E+03	2.71E+03	2.73E+03

	STD	8.30E+00	3.71E+00	1.66E+01	4.64E+00	3.52E+00	2.49E+01	8.05E+00	2.83E+00	8.71E+00	1.04E+01	1.71E+01	1.50E+01	**1.82E+00**	7.05E+00
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F26	**MEAN**	**2.70E+03**	2.72E+03	2.71E+03	2.76E+03	2.71E+03	2.71E+03	2.79E+03	2.80E+03	2.71E+03	2.75E+03	2.80E+03	2.73E+03	2.73E+03	2.71E+03
	MIN	**2.70E+03**	2.71E+03	2.71E+03	2.70E+03	2.70E+03	2.71E+03	2.71E+03	2.80E+03	2.71E+03	2.71E+03	2.71E+03	2.71E+03	2.71E+03	2.71E+03
	MAX	**2.70E+03**	2.80E+03	2.71E+03	2.80E+03	2.80E+03	2.71E+03	3.00E+03	2.80E+03	2.71E+03	2.81E+03	3.02E+03	3.00E+03	2.80E+03	2.71E+03
	STD	**3.56E-02**	3.59E+01	9.92E-01	4.79E+01	1.74E+01	4.74E-01	5.29E+01	8.77E-01	3.52E-01	4.76E+01	1.21E+02	5.95E+01	4.11E+01	5.63E-01
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F27	**MEAN**	**3.11E+03**	3.38E+03	3.75E+03	3.35E+03	3.28E+03	4.93E+03	4.07E+03	3.43E+03	3.88E+03	4.26E+03	4.08E+03	3.85E+03	3.25E+03	3.56E+03
	MIN	**3.10E+03**	3.12E+03	3.32E+03	3.11E+03	3.11E+03	4.87E+03	3.18E+03	3.33E+03	3.66E+03	3.96E+03	3.65E+03	3.34E+03	3.16E+03	3.13E+03
	MAX	**3.13E+03**	3.76E+03	3.97E+03	3.72E+03	3.84E+03	5.00E+03	4.30E+03	3.55E+03	4.05E+03	4.52E+03	4.66E+03	4.07E+03	3.42E+03	3.84E+03
	STD	**6.15E+00**	2.50E+02	1.86E+02	2.05E+02	1.74E+02	3.32E+01	2.15E+02	5.60E+01	9.47E+01	1.11E+02	2.86E+02	1.69E+02	6.04E+01	2.06E+02
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
F28	**MEAN**	**3.20E+03**	4.33E+03	4.34E+03	3.26E+03	3.21E+03	8.63E+03	3.26E+03	4.03E+03	4.92E+03	5.81E+03	3.30E+03	4.52E+03	3.67E+03	3.65E+03
	MIN	**3.14E+03**	3.82E+03	3.70E+03	3.15E+03	3.14E+03	7.30E+03	3.17E+03	3.87E+03	4.50E+03	4.45E+03	3.14E+03	3.85E+03	3.52E+03	3.54E+03
	MAX	**3.43E+03**	5.11E+03	5.56E+03	3.48E+03	3.48E+03	9.80E+03	3.64E+03	4.91E+03	5.29E+03	7.68E+03	4.04E+03	5.68E+03	4.69E+03	3.84E+03
	STD	**7.83E+01**	3.60E+02	4.19E+02	8.42E+01	8.03E+01	6.80E+02	8.89E+01	1.79E+02	2.33E+02	7.68E+02	1.84E+02	4.62E+02	2.58E+02	7.96E+01
	RS	n/a	+	+	+	+	+	+	+	+	=	+	+	+	+
F29	**MEAN**	3.29E+03	1.62E+06	3.99E+06	**3.13E+03**	3.15E+03	1.30E+08	3.16E+03	3.29E+04	1.53E+07	2.10E+07	3.92E+03	5.84E+06	1.31E+04	3.38E+05
	MIN	3.12E+03	4.40E+03	1.90E+05	3.11E+03	**3.11E+03**	7.15E+07	3.12E+03	1.19E+04	8.63E+06	4.62E+03	3.13E+03	1.75E+06	5.63E+03	4.63E+03
	MAX	3.92E+03	1.47E+07	1.09E+07	**3.23E+03**	3.34E+03	1.81E+08	3.24E+03	5.91E+04	2.61E+07	6.81E+07	1.40E+04	1.36E+07	2.13E+04	3.43E+06
	STD	1.50E+02	3.52E+06	2.71E+06	**3.22E+01**	6.33E+01	2.56E+07	3.52E+01	9.88E+03	3.50E+06	2.61E+07	2.29E+03	2.64E+06	3.91E+03	8.75E+05
	RS	n/a	+	+	-	-	+	-	+	+	+	+	+	+	+
F30	**MEAN**	**3.40E+03**	5.88E+03	7.63E+04	4.15E+03	4.24E+03	2.35E+06	4.67E+03	2.34E+04	4.11E+05	1.29E+04	5.43E+03	1.59E+05	1.20E+04	1.02E+04
	MIN	**3.27E+03**	4.28E+03	2.49E+04	3.31E+03	3.68E+03	1.12E+06	3.90E+03	8.70E+03	1.11E+05	5.61E+03	4.11E+03	4.38E+04	5.48E+03	5.84E+03
	MAX	**3.67E+03**	1.21E+04	2.30E+05	4.98E+03	4.99E+03	3.28E+06	5.31E+03	4.84E+04	7.16E+05	4.40E+04	2.46E+04	8.29E+05	2.66E+04	6.32E+04
	STD	**1.03E+02**	1.87E+03	4.56E+04	3.86E+02	2.95E+02	5.94E+05	3.78E+02	9.40E+03	1.47E+05	1.00E+04	3.64E+03	1.61E+05	5.21E+03	1.03E+04
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+

As shown in Table 5, the results indicate that HLPSO achieves the best performances for 27 out of 30 functions, i.e. F1-F5, F7-F11, F13-F28, and F30, in comparison with those of all the baseline methods. GMPSO obtains the most optimal results for F29, while FA and NaFA perform the best for F6 and F12, respectively. To further indicate the significance of the proposed HLPSO model, the Wilcoxon rank sum test is conducted. The statistical results indicate that our model performs statistically better than the baseline methods in most of the test cases. The exceptions are for F5, F6 and F28, where the proposed HLPSO model shows similar result distributions to those of VSSFA, CFA1, and CFA2, respectively. In addition, HLPSO is statistically worse than GMPSO, DNLPSO, and MFA for F29, FA and NaFA for F12, and FA for F6, respectively. Overall, HLPSO shows great superiority over other methods in solving diverse complex rotated, shifted, hybrid and composition benchmark functions.

7.2 Evaluation of Skin Lesion Segmentation

We subsequently evaluate the HLPSO-based KM clustering model for skin lesion segmentation. The classical search methods, other PSO and FA variant models, as well as KM and FCM clustering algorithms are employed for performance comparison. The following experimental setting is employed, i.e. image size=200×200, number of cluster=2, and trials=30. All the search methods employ the same number of MaxFEs, i.e. population (20) × iteration (20), as the stopping criterion. Each search method is integrated with KM and the new cost function based on the within- and between-cluster variance for foreground background pixel classification. We employ 484 and 550 images from the combined and the ISIC 2017 data sets, respectively, for segmentation evaluation. The Jaccard scores and the Wilcoxon rank sum test results (denoted as 'RS') are calculated for performance comparison. Table 6 tabulates the segmentation results for both the combined and the ISIC data sets.

Table 6 Jaccard scores and statistical test results for skin lesion segmentation

		HLPSO	**KM**	**FCM**	**AGPSO**	**GPSO**	**GMPSO**	**DNLPSO**	**ELPSO**	**MFA**	**NaFA**	**CFA1**	**CFA2**	**VSSFA**	**PSO**	**FA**	**MFO**
Combined	**MEAN**	**0.727**	0.672	0.6502	0.6945	0.7133	0.6946	0.7072	0.7027	0.7109	0.693	0.7123	0.7069	0.6901	0.6957	0.6906	0.7139
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ISIC	**MEAN**	**0.7315**	0.685	0.6931	0.715	0.7132	0.7006	0.7122	0.6951	0.7107	0.6957	0.7049	0.7075	0.7138	0.6983	0.7051	0.6966
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 7 Comparison with related studies for the ISIC data set in lesion segmentation

Studies	Methodology	Segmentation results (Jaccard score)
Yuan and Lo [5]	Fully convolutional-deconvolutional networks	0.765
Li and Shen [6]	two fully convolutional residual networks+a lesion index calculation unit	0.711
Bi et al. [15]	MResNet-Seg (Multi-scale)	0.760
Ronneberger et al. [69]	U-net	0.651
Shelhamer et al. [70]	FCN-8s	0.696
Alvarez and Iglesias [7]	KM+ensemble of regressions	0.679
Wen [71]	II-FCN	0.697
This research	**HLPSO+KM+within- and between-cluster variance**	**0.7315**

As indicated in Table 6, the HLPSO-based clustering model shows superior search capabilities in improving cluster centroids for pixel classification. In comparison with conventional KM and FCM models, it is insensitive to noise and centroid initialization, and achieves improved segmentation performance. It also outperforms all other baseline search methods statistically for foreground background pixel classification. The convergence curves for the combined and the ISIC data sets over a set of 30 runs are depicted in Figures 6-7.

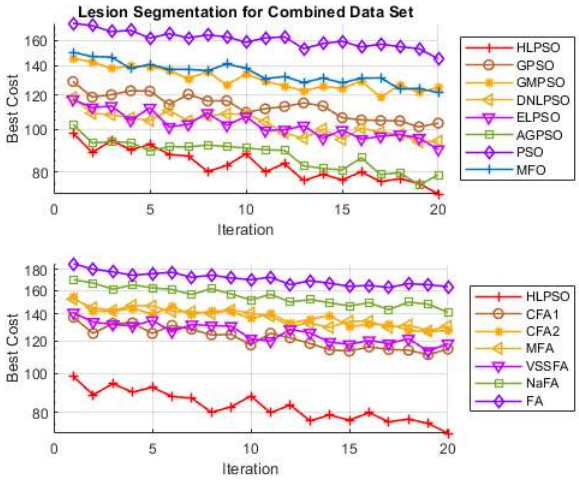

Figure 6 Convergence curves over a set of 30 runs for the combined data set for image segmentation (where x and y axes denote the iteration number and the within- and between-cluster variance measure, respectively.)

As shown in Figure 6, for the combined data set, most of the PSO variants operate with faster convergence rates than those of the FA methods over a set of 30 runs. The HLPSO-based clustering model shows the fastest convergence rate comparing with other clustering models incorporated with classical and advanced search methods. CFA1 and VSSFA have faster convergence speeds in comparison with those of other FA methods while AGPSO, ELPSO and DNLPSO converge faster than other PSO variants. Owing to the inclusion of diverse search strategies to mitigate premature convergence, all the PSO variants possess faster convergence rates in comparison with that of the original PSO algorithm, while the proposal of diverse strategies to overcome stagnation in the FA variants ascertains their search efficiency as compared with the original FA model. MFO also operates with a faster convergence speed in comparison with those of the original FA and PSO algorithms.

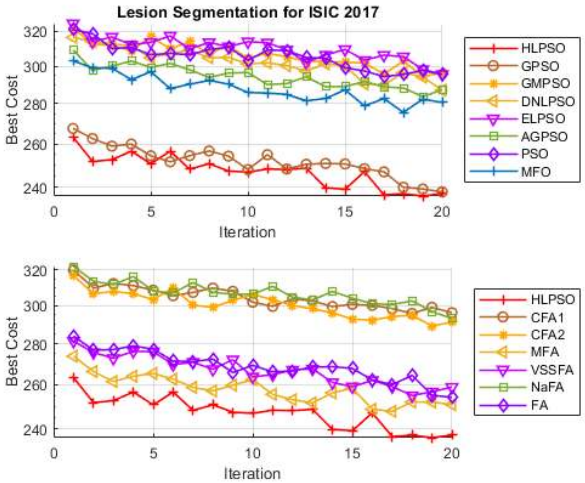

Figure 7 Convergence curves over a set of 30 runs for the ISIC data set for image segmentation (where x and y axes denote the iteration number and the within- and between-cluster variance measure, respectively.)

Moreover, as illustrated in Figure 7, for the ISIC data set, owing to the embedding of diverse random walk strategies and a number of global search mechanisms, the proposed HLPSO model shows superior search efficiency and obtains a faster convergence speed than those of all the baseline methods. Equipped with a

genetic crossover operator for the enhancement of swarm diversity, the GPSO model has a faster convergence speed over a set of 30 runs in comparison with those of other PSO variants whereas ELPSO operates with the slowest convergence rate. Guided by both neighbouring and global best experiences, the search process of MFA possesses better convergence capabilities in comparison with those of other FA methods with NaFA converging the slowest. Also, FA has a faster convergence rate, followed by those of MFO and PSO.

Figure 8 depicts several example segmentation results of the proposed HLPSO model. As illustrated in Table 7, in comparison with related methods in the literature for the ISIC 2017 data set, the proposed clustering model is among top performers for lesion segmentation. It also does not require any training, and has a better trade-off between computational efficiency and model performance, in comparison with those of other existing supervised learning methods [5, 6, 15].

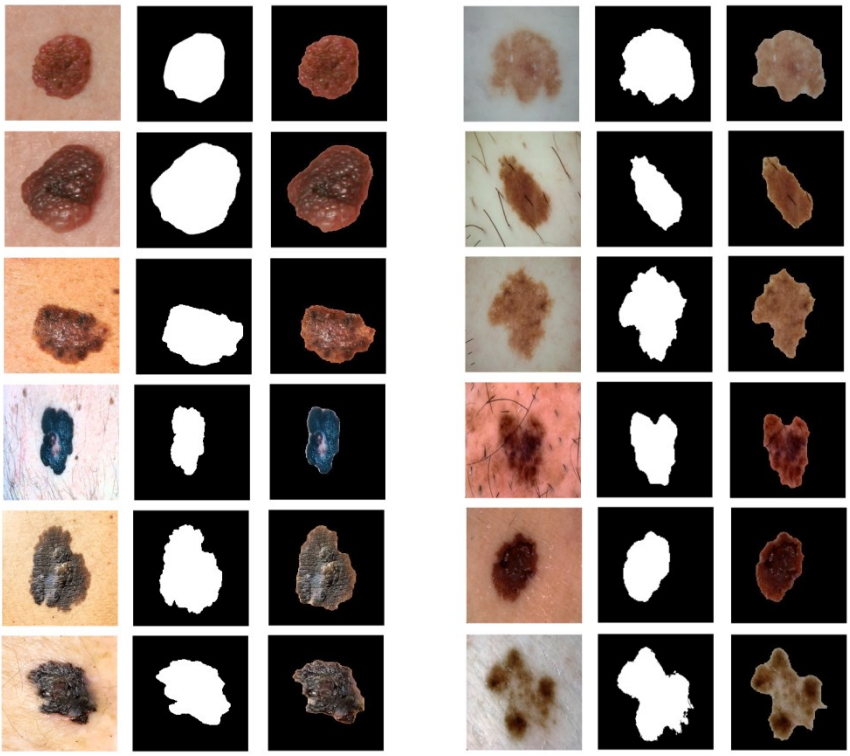

Figure 8 Example segmentation results using the HLPSO-based KM clustering model (Left: the mixed data set, Right: the ISIC data set. In each set, from left to right, 1. the original image, 2. GT mask, 3. segmented lesion output using the proposed clustering model)

7.3 Evaluation of Evolving Deep CNNs for Lesion Classification

Lesion images filtered using both the ground-truth (GT) masks and the segmentation outputs using the enhanced KM clustering are used for lesion classification. As such, the input images contain coloured lesion regions in the foreground with a black background. In short, only lesion information is used for melanoma classification in deep learning models. All the segmented lesion images for the mixed data set are used for lesion classification while for the ISIC data set, we select 400 segmented images with reasonable Jaccard scores for lesion classification task. The following experimental setting is used, i.e. image size=32×32, and dimension=3. To ensure a fair comparison, all the methods employ the same number of MaxFEs, i.e. population (30) × iteration (10), as the stopping criterion. The adaptive CNN models devised by HLPSO and other search methods are subsequently used for melanoma classification. A default CNN network, as shown in Figure 4 (where network depth=1, learning rate=0.01 (MATLAB default setting) and the regularization factor=0.0001 (MATLAB default setting)), is also employed for performance comparison. We employ 60:20:20 as the training, validation and test set ratio for each data set.

7.3.1 Evaluation using the combined data set

The combined data set is first used for evaluating deep architecture generation. The mean results of 30 individual runs for each method are used as the main criterion for performance comparison. Table 8 shows the detailed results for the combined data set. The input lesion images filtered by the GT and segmented masks are

denoted as 'GT' and 'Seg' cases, respectively. As indicated in Table 8, for both cases, the adaptive CNN models devised by HLPSO outperform those of other search methods and the default CNN network significantly in most cases. The lesion images filtered by the GT masks possess accurate lesion segmentation information for lesion classification while the input images filtered by the segmentation outputs using the extended KM clustering also carry significant discriminative information. They both lead to reasonable performances, with a higher accuracy rate obtained for GT filtered input images.

Table 8 also tabulates the statistical Wilcoxon rank sum test results (denoted as 'RS') to indicate the effectiveness of the proposed model. With the GT filtered images as the inputs, the deep learning models generated by HLPSO show great superiority over those yielded by other search methods and the default network, except for those of GPSO, CFA2 and VSSFA, which have the same result distributions to those of HLPSO. With the input images filtered by the segmentation outputs, our adaptive CNNs also achieve improved performance as compared with those of others. The exceptions are for GMPSO, VSSFA, PSO, and FA based deep learning models, which achieve similar performances to those of the HLPSO model.

Table 8 Classification and Wilcoxon rank sum test results for optimized CNNs using the combined data set

		HLPSO	AGPSO	GPSO	GMPSO	DNLPSO	ELPSO	MFA	NaFA	CFA1	CFA2	VSSFA	PSO	FA	MFO	default
GT	**MEAN**	**0.9137**	0.8481	0.8871	0.7889	0.7980	0.6631	0.7569	0.8142	0.5876	0.8152	0.8788	0.8385	0.8159	0.8086	0.7495
	RS	n/a	+	=	+	+	+	+	+	+	=	=	+	+	+	+
Seg	**MEAN**	**0.8317**	0.7263	0.7865	0.7888	0.7046	0.6919	0.7185	0.7543	0.6120	0.7446	0.7768	0.7784	0.7818	0.7502	0.7446
	RS	n/a	+	+	=	+	+	+	+	+	+	=	=	=	+	+

Figure 9 illustrates the convergence curves over a series of 30 runs for the mixed data set with GT filtered images for training and validation. HLPSO shows the fastest convergence rate in comparison with those of all the baseline models. All PSO models have comparatively faster convergence speeds in comparison with those of all the FA variants. The convergence speed of MFO is also faster than those of the original FA and its variant models. As mentioned earlier, GPSO employs a genetic crossover operator to increase swarm diversity. It converges faster than all other PSO variants. Besides GPSO, AGPSO with adaptive search parameters and the original PSO algorithm also show comparatively faster convergence speeds. Moreover, equipped with chaotic attractiveness coefficients, CFA1 and CFA2 possess better convergence rates than those of other FA variants. Furthermore the search process of PSO illustrates more efficient convergence capabilities in comparison with those of MFO and FA for deep architecture generation for this combined data set. The convergence curves for the mixed data set with the images filtered by the segmented masks show similar cases for all the algorithms.

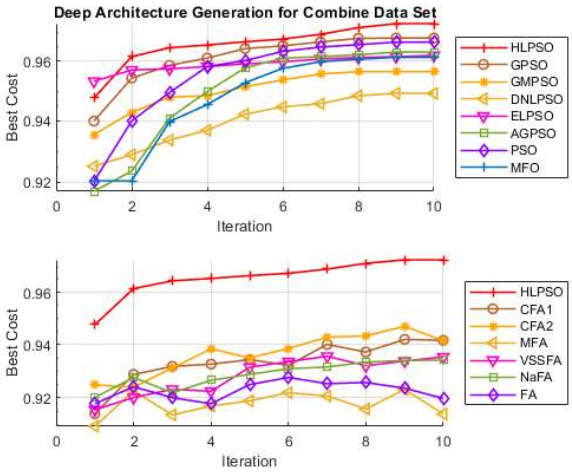

Figure 9 Convergence curves over a set of 30 runs for the combined data set for deep architecture generation (where x and y axes denote the iteration number and classification performance, respectively.)

Moreover, Table 9 shows the identified average learning rate, weight decay, and block depth for each method over a series of 30 runs for the mixed data set with GT segmented images as the inputs. The empirical results indicate that HLPSO, GPSO, and CFA2 generate comparatively smaller deep architectures owing to the identified smaller block depths for each convolutional block, as well as comparatively smaller learning rates and moderate weight decays. These configurations result in networks with sufficient capability of overcoming local optima. MFA also constructs comparatively smaller deep architectures but with comparatively larger learning

rates and smaller weight decays. The empirical results indicate that the identified shallow networks with large learning rates and small penalties to large weights have constrained capabilities in feature learning, and are more likely to converge prematurely. On the other hand, ELPSO and CFA1 yield the largest networks with the largest block depths for each convolutional block, as well as comparatively larger learning rates. The experimental results indicate that such networks with large learning steps are more likely to converge to sub-optimal solutions with a more expensive training cost. More in-depth analysis on deep architecture generation is provided in Section 7.3.2.

Table 9 The identified average optimal hyper-parameters over a series of 30 runs for the combined data set with GT filtered images as inputs

		Accuracy	Learning rate	Weigh decay	Block depth
HLPSO	mean	0.9137	2.47E-04	2.58E-02	2.2333
	std.	2.67E-02	8.03E-04	1.75E-02	8.17E-01
AGPSO	mean	0.8481	4.78E-03	3.10E-02	2.3
	std.	1.12E-01	1.33E-02	1.72E-02	8.37E-01
GPSO	mean	0.8871	1.00E-04	2.97E-02	2.0667
	std.	3.26E-02	4.14E-20	2.02E-02	9.07E-01
GMPSO	mean	0.7889	2.12E-03	1.87E-02	2.4
	std.	9.05E-02	6.34E-03	1.61E-02	8.55E-01
DNLPSO	mean	0.798	1.16E-02	3.52E-02	2.6333
	std.	1.51E-01	2.04E-02	1.97E-02	4.90E-01
ELPSO	mean	0.6631	2.62E-02	2.28E-02	2.6667
	std.	1.94E-01	1.59E-02	1.37E-02	5.47E-01
MFA	mean	0.7569	1.59E-02	1.75E-02	2.2
	std.	1.91E-01	1.30E-02	1.19E-02	4.07E-01
NaFA	mean	0.8142	4.53E-03	2.32E-02	2.5667
	std.	7.84E-02	5.09E-03	1.32E-02	5.04E-01
CFA1	mean	0.5876	3.20E-02	3.50E-02	2.9333
	std.	2.07E-01	5.47E-03	1.04E-02	2.54E-01
CFA2	mean	0.8152	1.01E-03	2.37E-02	2.2
	std.	6.74E-02	2.74E-03	1.11E-02	5.51E-01
VSSFA	mean	0.8788	2.72E-03	2.36E-02	2.4333
	std.	8.00E-02	6.15E-03	1.17E-02	6.26E-01
PSO	mean	0.8385	2.29E-03	2.17E-02	2.4
	std.	1.41E-01	8.78E-03	1.35E-02	8.14E-01
FA	mean	0.8159	7.29E-03	2.79E-02	2.5333
	std.	1.23E-01	7.69E-03	1.25E-02	5.07E-01
MFO	mean	0.8086	1.90E-04	2.14E-02	2.3667
	std.	4.69E-02	4.74E-04	1.77E-02	7.65E-01

7.3.2 Evaluation using the ISIC data set

The above experimental setting is also used for evaluation of the ISIC data set. The average confusion matrix performance over 30 individual runs for each method is also used for performance comparison. Table 10 presents the classification and statistical test results (denoted as 'RS') for the ISIC data set. Again the lesion images filtered by both the GT masks and segmentation outputs are used for evaluating the optimized CNN models. As indicated in Table 10, the deep learning models devised by HLPSO achieve the best performances, and outperform all other networks statistically. In other words, when using the lesion images filtered by both the GT and the segmented masks, our adaptive CNN models outperform all other optimized models and the default network statistically.

Table 10 Classification and Wilcoxon rank sum test results for optimized CNNs using the ISIC data set

		HLPSO	AGPSO	GPSO	GMPSO	DNLPSO	ELPSO	MFA	NaFA	CFA1	CFA2	VSSFA	PSO	FA	MFO	default
GT	**MEAN**	**0.7376**	0.6314	0.6710	0.6525	0.5985	0.5744	0.5876	0.6629	0.5093	0.6722	0.6314	0.6571	0.6597	0.6476	0.6138
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Seg	**MEAN**	**0.7005**	0.6519	0.6204	0.6103	0.6326	0.5880	0.6356	0.6549	0.5177	0.6225	0.6541	0.6301	0.6352	0.6441	0.6150
	RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Besides the above experiments, we have also incorporated a fine-tuning procedure into the optimization process of our model to further enhance its performance. Based on Fielding and Zhang [72], after the identification of the best performing deep neural network using the proposed HLPSO model, we combine the training and validation sets of each test data set into a larger training set and conduct fine-tuning of the optimized model using this larger data set for a number of epochs, in order to reduce any effect of overfitting. As indicated in Table 11, the results obtained after fine-tuning for both data sets show significant improvements.

Table 11 Evaluation results of HLPSO with and without fine-tuning

	HLPSO (before fine-tuning)	HLPSO (after fine-tuning)
Combined data set (GT)	0.9137	**0.9537**
Combined data set (Segmented)	0.8317	**0.8885**
ISIC data set (GT)	0.7376	**0.8427**
ISIC data set (Segmented)	0.7005	**0.7883**

Figure 10 illustrates the convergence curves over a series of 30 runs for the ISIC data set with the images filtered by the GT masks for training and validation. HLPSO shows the fastest convergence speed in comparison with those of all the baseline methods. The original and modified PSO algorithms converge faster in comparison with the original FA and its variant models. The convergence speed of MFO is between those of PSO and FA variants. Moreover, GMPSO, which incorporates subswarm strategies, GA operators and probability distributions, has the fastest convergence speed among the PSO variants. Besides GMPSO, GPSO and DNLPSO show better convergence capabilities than those of other PSO models. Also owing to the employment of optimal solutions from a dynamic neighbourhood, the convergence speed of NaFA is the fastest among the FA variants. MFA and FA also show subsequent more efficient convergence rates than those of other FA models. The convergence curves of the ISIC data set with the images filtered by the segmented masks show similar cases for all the algorithms.

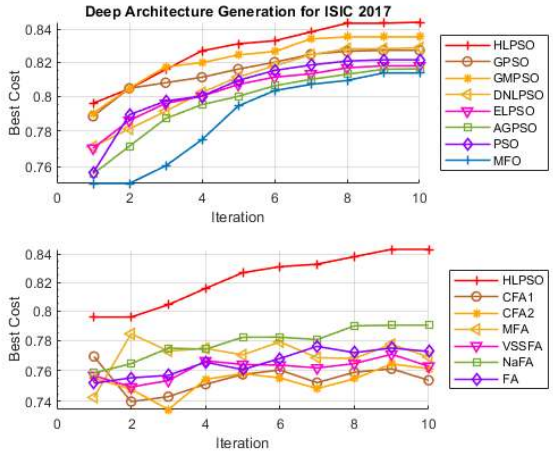

Figure 10 Convergence curves over a set of 30 runs for the ISIC data set for deep architecture generation (where x and y axes denote the iteration number and classification performance, respectively.)

Further analysis on the identified hyper-parameters for the ISIC data sets is provided. The optimal hyper-parameters identified over a set of 30 runs for the ISIC data set with GT filtered images are listed in Table 12.

Table 12 The identified average optimal hyper-parameters over a series of 30 runs for the ISIC data set with GT filtered images as inputs

		Accuracy	Learning rate	Weight decay	Block depth
HLPSO	mean	0.7376	4.30E-03	2.39E-02	2.3667
	std.	5.71E-02	2.11E-03	1.61E-02	4.90E-01
AGPSO	mean	0.6314	1.09E-02	2.93E-02	2.5333
	std.	1.22E-01	1.27E-02	1.53E-02	5.07E-01
GPSO	mean	0.671	5.28E-03	3.00E-02	2.6333
	std.	9.83E-02	8.66E-03	1.45E-02	4.90E-01
GMPSO	mean	0.6525	4.94E-03	2.58E-02	2.7
	std.	1.15E-01	8.43E-03	1.46E-02	4.66E-01
DNLPSO	mean	0.5985	1.28E-02	3.06E-02	2.67
	std.	1.19E-01	1.37E-02	2.21E-02	4.79E-01
ELPSO	mean	0.5744	2.05E-02	2.38E-02	2.5333
	std.	1.10E-01	1.52E-02	1.65E-02	5.07E-01
MFA	mean	0.5876	1.09E-02	1.45E-02	2.0667
	std.	1.00E-01	1.05E-02	1.29E-02	2.54E-01
NaFA	mean	0.6629	4.77E-03	2.56E-02	2.7333
	std.	1.04E-01	3.04E-03	1.26E-02	4.50E-01
CFA1	mean	0.5093	3.50E-02	4.20E-02	2.9
	std.	3.76E-02	9.72E-03	1.44E-02	3.16E-01

CFA2	mean	0.6722	7.39E-03	2.66E-02	2.4595
	std.	1.18E-01	1.05E-02	1.01E-02	5.45E-01
VSSFA	mean	0.6314	7.84E-03	2.29E-02	2.5
	std.	9.64E-02	1.06E-02	1.06E-02	5.09E-01
PSO	mean	0.6571	8.11E-03	2.40E-02	2.6333
	std.	1.11E-01	9.56E-03	1.59E-02	4.90E-01
FA	mean	0.6597	7.56E-03	2.34E-02	2.4667
	std.	1.15E-01	6.69E-03	1.12E-02	5.07E-01
MFO	mean	0.6476	8.08E-03	2.15E-02	2.6667
	std.	1.03E-01	1.10E-02	1.54E-02	4.71E-01

The HLPSO-based deep architectures achieve the best performance in comparison with the deep networks generated by all other FA and PSO variants and classical methods. As illustrated in Table 12, it identifies comparatively smaller convolutional block architectures with an average network depth of 2.3667 for each convolutional block, in comparison with those of nearly all the baseline methods. The following two sets of identified hyper-parameters lead to the best performances of the proposed HLPSO-based CNN models with 60-20-20 for training, validation, and test, i.e. learning rate=0.0014, weight decay=0.05, depth=2 (for each convolutional block), and learning rate=0.004, weight decay=0.0472, depth=2 (for each convolutional block). These settings yield the accuracy rates of 0.8486 and 0.8443, respectively, before fine-tuning.

As indicated in Table 12, MFA, HLPSO, CFA2, FA and VSSFA identify deep networks with comparatively smaller architectures whereas GMPSO, NaFA, PSO, MFO, DNLPSO, GPSO, and CFA1 yield deeper networks. ELPSO and AGPSO construct middle-sized networks. Most of the models produce similar weight decay settings, resulting in similar penalties to large weights in the networks. The obtained learning rates vary in different models.

Specifically, the MFA model constructs the smallest deep architectures with large learning rates and small weight decays in most of the test cases. However, the identified shallow deep networks have constrained capabilities in feature learning, and the large learning rates are more likely to result in the CNN models being trapped in local optima, therefore, leading to less optimal performance. HLPSO constructs the second smallest CNN architectures as well as comparatively smaller learning rates and reasonable settings for weight decays. In comparison with all the baseline methods, its identified optimal topologies show sufficient capabilities in feature learning. On the other hand, the identified small learning rates are more capable of avoiding local optima by applying suitable learning steps to guide the model training process more efficiently. The identified weight decays also achieve reasonable trade-off between the application of large penalties to large weights and possibilities in allowing the weights to grow freely. Therefore, HLPSO-based deep networks achieve the best performances for a successive of 30 runs in comparison with all other baseline methods. CFA2 identifies the third smallest network topologies. Its search process identifies comparatively larger learning rates and weight decays as compared with those of HLPSO, which lead to comparatively slightly larger learning steps in the training process and more intense penalties to large weights, therefore, resulting in comparatively lower average accuracy scores, as shown in Table 12. Other methods, e.g. GMPSO, NaFA, PSO, MFO, DNLPSO, GPSO, and CFA1, tend to generate comparatively larger networks with higher learning rates in comparison with those of HLPSO. While most of these larger networks produce sufficiently good performances, they are still less competitive than those of HLPSO. Besides that, such deeper networks may also lead to high computational costs in the subsequent training process. CFA1 generates the deepest networks with the largest learning rates and weight decays. These characteristics result in less efficient models, with comparatively worse classification performance.

Overall, large learning rates with extremely deeper (e.g. CFA1 and DNLPSO) or shallower networks (e.g. MFA) are more likely to converge to sub-optimal solutions and result in less competitive performances. On the other hand, reasonable network sizes with comparatively smaller learning hyper-parameters tend to achieve compelling performances for both data sets. Moreover, owing to the proposed diverse search strategies, HLPSO is able to overcome stagnation and attain global optimal solutions effectively pertaining to deep architecture generation. It outperforms all the baseline models significantly with optimal sizes of the deep architectures and more effective learning parameters. The empirical and statistical test results further ascertain the efficiency of the proposed HLPSO model.

7.3.3 Computational cost for deep architecture design

We further analyse the computational cost for deep architecture design. Since all the models employ the same maximum number of function evaluations as the stopping criterion for deep architecture generation and the fitness evaluation is the most time consuming component that contributes to most of the computational cost, all the methods have the same computational cost for CNN model design principally, i.e. population × the

maximum number of iterations. However, each model has different internal search strategies. The computational cost varies slightly. Table 13 illustrates the mean computational cost of each method over 30 runs for deep architecture generation for both the combined and ISIC data sets with images filtered by the GT masks as inputs.

Table 13 The mean computational cost (in seconds) of each method over a set of 30 runs for deep architecture generation

Seconds	HLPSO	AGPSO	GPSO	GMPSO	DNLPSO	ELPSO	MFA	NaFA	CFA1	CFA2	VSSFA	PSO	FA	MFO
Combined	586.89	591.73	397.54	700.87	624.01	595.52	588.42	553.73	537.44	432.93	572.19	547.30	498.42	502.00
ISIC	451.92	513.0	435.04	518.08	484.97	493.63	320.18	373.15	384.20	354.67	382.29	386.26	383.84	362.66

HLPSO and certain PSO and FA variants embed diverse search strategies, and they are likely to generate slightly larger computational costs than those of classical search methods. For the ISIC data set, HLPSO shows similar or lower computational costs as compared with those of GPSO, AGPSO, GMPSO, DNLPSO, and ELPSO. For the mixed data set, HLPSO shows comparatively lower computational costs as compared with those of AGPSO, GMPSO, DNLPSO, ELPSO, and MFA, respectively.

7.4 Evaluation of Ensemble Classifiers for Lesion Classification

We evaluate HLPSO-based feature selection and ensemble lesion classification system in this section. We extract shape, colour and texture features from both lesion and skin sub-images filtered by the GT and the generated masks. Optimized features are identified for lesion classification. In short, both lesion and skin features are used for lesion classification.

The following experimental settings are applied, i.e. image size=500×500, and dimension=147(shape+colour) or 1062(texture). We employ 80:20 as the training and test ratio for each data set. To ensure a fair comparison, all the methods employ the same number of MaxFEs, i.e. population (50) × iteration (500), as the termination criterion. Each ensemble model comprises two base classifiers, which are dedicated to the optimized shape and colour as well as texture features, respectively. Sub-images filtered by both the GT and segmented masks are used as the inputs for ensemble lesion classification.

A total of 30 individual runs are used for evaluating each search method. The average GM scores over 30 runs are used as the main criterion for performance comparison. Table 14 illustrates the average GM performance and Wilcoxon rank sum test results (denoted as 'RS') for the combined data set. Using GT filtered input images, HLPSO outperforms all other methods for both KNN- and SVM-based ensembles, with MFA achieving the best performance for KNN-based ensemble with hold-out validation. When the SVM-based ensemble classifier is used, HLPSO-based feature selection shows statistically significant improvements over other feature selection methods. Integrated with the KNN-based ensemble model for both 10-fold and hold-out validations, our results are statistically better than those of other methods in most cases. The exceptions are for DNLPSO, GMPSO, and MFA, which have the same result distributions to those of HLPSO. NaFA and CFA1 also show similar results to those of our model when combined with the KNN-based ensemble classifier in the hold-out validation. Using the input images filtered by our segmentation outputs for both training and test, feature selection using HLPSO achieves significant superiority over all other methods statistically for both ensemble classifiers, except for PSO-based feature selection, which has a similar GM performance to those of HLPSO integrated with SVM-based ensemble in hold-out validation. Figure 11 shows the convergence curves of all the methods over 30 runs with GT segmented images as the inputs in the training stage. HLPSO operates with the fastest convergence rate as compared with those of all the baseline methods for discriminative shape+colour and texture feature selection during the training process. Besides HLPSO, for both feature selection tasks, equipped with dynamic neighbourhood learning, adaptive search coefficients and swarm diversity enhancement respectively, DNLPSO, AGPSO and GMPSO show faster convergence rates than those of other PSO variants. With the employment of chaotic and adaptive search parameters respectively, CFA1, CFA2 and VSSFA operate with the fastest convergence rates among all the FA variants. MFO also shows a better convergence speed than those of PSO and FA for both feature selection tasks.

Table 14 Ensemble classification and statistical test results for the combined data set

		HLPSO	AGPSO	DNLPSO	GPSO	ELPSO	GMPSO	CFA1	MFA	VSSFA	NaFA	CFA2	PSO	FA	MFO
GT	**ESEM KNN**	**10-Fold**	**0.9543**	0.9387	0.9460	0.9396	0.9411	0.9449	0.9436	0.9511	0.9401	0.9422	0.9422	0.9394	0.9306	0.9395
		RS	n/a	+	=	+	+	=	+	=	+	+	+	+	+	+
		Hold-out	0.9452	0.9322	0.9407	0.9347	0.9350	0.9389	0.9384	**0.9475**	0.9342	0.9382	0.9361	0.9334	0.9283	0.9347
		RS	n/a	+	=	+	+	=	=	=	+	=	+	+	+	+
	ESEM SVM	**10-Fold**	**0.9788**	0.9575	0.9571	0.9573	0.9519	0.9573	0.9557	0.9534	0.9620	0.9612	0.9577	0.9554	0.9533	0.9567
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+

SEG		**Hold-out**	**0.9736**	0.9584	0.9583	0.9590	0.9534	0.9590	0.9577	0.9544	0.9642	0.9634	0.9630	0.9579	0.9565	0.9567
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
	ESEM KNN	**10-Fold**	**0.9033**	0.8783	0.8720	0.8794	0.8801	0.8796	0.8791	0.8847	0.8732	0.8763	0.8782	0.8745	0.8725	0.8777
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.9070**	0.8814	0.8761	0.8826	0.8824	0.8817	0.8800	0.8880	0.8749	0.8769	0.8803	0.8745	0.8767	0.8804
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
	ESEM SVM	**10-Fold**	**0.9063**	0.8832	0.8822	0.8895	0.8862	0.8880	0.8859	0.8824	0.8777	0.8864	0.8889	0.8928	0.8802	0.8847
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.9005**	0.8875	0.8865	0.8936	0.8911	0.8921	0.8889	0.8844	0.8837	0.8903	0.8931	0.8962	0.8840	0.8894
		RS	n/a	+	+	+	+	+	+	+	+	+	+	=	+	+

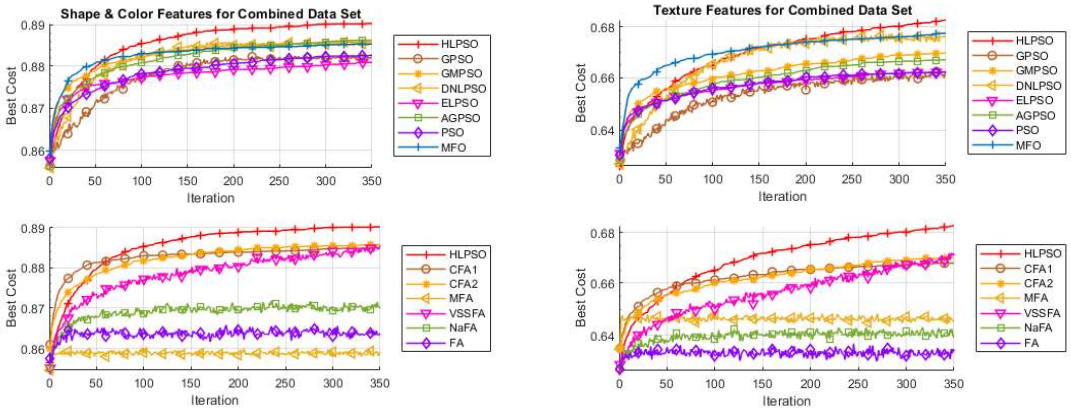

Figure 11 Convergence curves over a set of 30 runs for the combined data set for shape+colour (left) and texture (right) feature selection (where x and y axes denote the iteration number and classification performance, respectively.)

Table 15 illustrates the average GM results and statistical rank sum test results (denoted as 'RS') for the ISIC subset. Using GT filtered images as the inputs, HLPSO-based feature selection shows statistically significant improvements over nearly all other methods. The exceptions are for AGPSO, GMPSO, CFA1, VSSFA, and FA, which show similar result distributions to that of HLPSO when the KNN-based ensemble model in hold-out validation is used. Using images filtered by the segmented masks for training and test, our model outperforms nearly all other methods statistically, except for GMPSO which achieves a similar performance to that of HLPSO combined with KNN-based ensemble in hold-out validation. As illustrated in Figure 12, over a set of 30 runs, HLPSO converges faster than all other baseline models for both shape+colour and texture feature optimization in the training stage with images filtered by the GT masks for training and validation. Besides HLPSO, for both feature selection tasks, MFO illustrates a faster convergence speed than those of nearly all the PSO and FA variants. Moreover DNLPSO, GMPSO and AGPSO show faster convergence rates than those of other PSO variants, while CFA2, VSSFA and CFA1 illustrate the best convergence capabilities than those of all other FA variants.

Table 15 Ensemble classification and statistical test results for the ISIC data set

			HLPSO	**AGPSO**	**DNLPSO**	**GPSO**	**ELPSO**	**GMPSO**	**CFA1**	**MFA**	**VSSFA**	**NaFA**	**CFA2**	**PSO**	**FA**	**MFO**
GT	**ESEM KNN**	**10-Fold**	**0.8805**	0.8394	0.8315	0.8362	0.8247	0.8451	0.8439	0.8126	0.8564	0.8453	0.8336	0.8378	0.8407	0.8467
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.8162**	0.8043	0.7988	0.7942	0.7865	0.8027	0.8017	0.7725	0.8156	0.8006	0.7964	0.7948	0.8000	0.8007
		RS	n/a	=	+	+	+	=	=	+	=	+	+	+	=	+
	ESEM SVM	**10-Fold**	**0.8825**	0.7682	0.7722	0.7797	0.7709	0.7755	0.7631	0.7783	0.7951	0.7802	0.7700	0.7764	0.7940	0.7875
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.8398**	0.7741	0.7780	0.7797	0.7748	0.7749	0.7680	0.7837	0.7993	0.7815	0.7822	0.7827	0.7898	0.7902
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
SEG	**ESEM KNN**	**10-Fold**	**0.8236**	0.7916	0.7930	0.7864	0.7883	0.7985	0.7841	0.7619	0.7728	0.7865	0.7970	0.7934	0.7910	0.7869
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.8132**	0.7868	0.7927	0.7856	0.7891	0.7999	0.7900	0.7686	0.7756	0.7878	0.7943	0.7902	0.7897	0.7841
		RS	n/a	+	+	+	+	=	+	+	+	+	+	+	+	+
	ESEM SVM	**10-Fold**	**0.8355**	0.7767	0.7782	0.7613	0.7832	0.7870	0.7717	0.7689	0.7292	0.7525	0.7687	0.7789	0.7482	0.7521
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+
		Hold-out	**0.8250**	0.7797	0.7740	0.7605	0.7852	0.7878	0.7688	0.7561	0.7362	0.7622	0.7691	0.7832	0.7505	0.7601
		RS	n/a	+	+	+	+	+	+	+	+	+	+	+	+	+

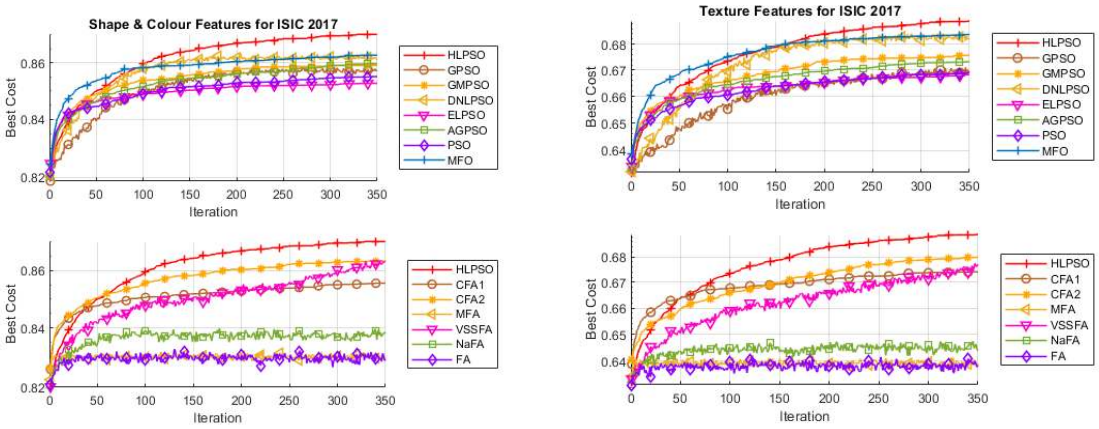

Figure 12 Convergence curves over a set of 30 runs for the ISIC data set for shape+colour (left) and texture (right) feature selection (where x and y axes denote the iteration number and classification performance, respectively.)

Table 16 depicts the comparison with related lesion classification methods in the literature for the ISIC data set. Since each study employed different sets of (archived, original and augmented) training and test images and different strategies for performance evaluation, Table 16 serves as an approximate performance indicator. Most existing methods used the original images filtered with GT masks for training. With the GT filtered images as the inputs, ensemble lesion classification with HLPSO-based feature selection shows superior performance, and is among the top performers for the ISIC data set. Moreover, with a smaller size (only 32×32) of images as inputs, the evolving deep neural networks also achieve a superior performance after fine-tuning.

Table 16 Comparison with related results in the literature for the ISIC data set for lesion classification

Studies	Methodology	Recognition rate
Li and Shen [6]	two fully convolutional residual networks+a lesion index calculation unit	0.823
Bi at al. [15]	deep ResNets	0.896
Mirunalini et al. [73]	Inception-v3 (to extract feature vector) + neural networks	0.655
Chang [74]	Inception-v3 (full image) + Inception-v3 (segmented region)	0.705
Zhang et al. [75]	CNN in Caffe	0.658
This research	**HLPSO-based evolving deep CNNs (with 32x32 input)**	**0.8427 (hold-out)**
This research	**Kirsch+SLBP+HLPSO-based feature selection+ensemble classification**	**0.8825 (10-fold) 0.8398 (hold-out)**

The proposed HLPSO model has both local and global strategies to enhance its search process and achieves superior performances as compared with those of FA-based as well as PSO-based variants. On one hand, NaFA [64] employs brighter fireflies from a pre-defined constrained neighbourhood to lead the search process. CFA1 [22] and CFA2 [65] adopt Logistic and Gauss maps as the attractiveness coefficients respectively. An adaptive random search parameter is implemented in the VSSFA [67] model while MFA [66] uses the Tent map for population initialization and both neighbouring and global best solutions for position updating. Since the search processes of these FA variants are mainly based on the original FA operation, if there are no brighter fireflies in the neighbourhood, they are very likely to be trapped in local optima.

On the other hand, ELPSO [60] employs a five-step strategy to further enhance the global best solution including Gaussian, Cauchy, opposition- (for each dimension as well as for the whole particle) and DE-based mutations. GMPSO [21] employs the GA and probabilities distributions to diversify the subpopulation and enhance the subswarm leader, respectively. Adaptive search parameters, i.e. increasing and decreasing acceleration coefficients, are implemented in AGPSO [61], whereas GPSO [63] employs a genetic crossover operator to diversify the particle swarm. DNLPSO [62] employs the historical best memories in a dynamically updated neighbourhood to lead the search process. Again, since the original PSO operation led by a single leader largely dominates the search process of the primary swarms of the above PSO variants, such search processes tend to stagnate prematurely.

In comparison with the above PSO and FA variants, diverse local and global search strategies are proposed for the HLPSO model to overcome the local optima traps. Two FA variants (led by a randomly selected promising neighbouring solution and the average of all brighter neighbouring fireflies, respectively), the PSO operation and

the spiral action are used as the four global search mechanisms to increase exploration, while crossover, mutation and probability distributions are used to increase local exploitation. In particular, the above four global search actions cooperate with each other to overcome stagnation. As an example, when the PSO operation led by a single leader stagnates, the FA and spiral search actions based on multiple neighbouring and global promising solutions are able to drive the search process out of the local optima traps. When FA-based strategies are trapped in local optima due to the lack of brighter fireflies present in the neighbourhood, the PSO and spiral search operations led by the global promising solutions are used to mitigate premature convergence. In short, the above local and global search strategies work in a collaborative manner to attain global optimality and account for the superior performance of the proposed model in comparison with those of other search methods.

8. CONCLUSIONS

In this research, we have proposed a new PSO variant, i.e., HLPSO, for segmentation, discriminative feature selection, and evolving adaptive deep CNN model generation for lesion classification. The key contributions of this research include the following. (1) The proposed HLPSO model incorporated with KM clustering and a new cost function based on intra- and inter-cluster variance is used to overcome centroid initialization sensitivity and local optima traps of the KM clustering model for foreground background pixel classification in lesion segmentation. (2) Deep architecture design and optimal hyper-parameter selection are the main bottleneck for deploying deep neural networks to new application domains. These tasks require profound domain knowledge, and are mainly conducted by manual trial-and-error procedures. To overcome such limitations, HLPSO is used to automate this process and identify optimal topologies and hyper-parameters of CNNs automatically for skin lesion classification. The empirical results indicate that the adaptive CNN models with optimized topologies and hyper-parameters devised by HLPSO outperform those of other state-of-the-art FA and PSO variants and classical search methods and the default network for lesion classification statistically. (3) Moreover, ensemble lesion classification with HLPSO-based feature selection has achieved great superiority over other existing lesion classification methods in the literature. The proposed diverse search strategies account for the superiority of discriminative feature selection over other classical and advanced search methods. (4) To further indicate the model efficiency, we have employed classical benchmark functions as well as the complex CEC 2014 test suite for evaluation. The experimental results and statistical tests further indicate the efficiency of HLPSO in solving diverse complex rotated, shifted, hybrid and composition benchmark functions.

For future work, other clustering algorithms can be used for integration with the proposed HLPSO model and the intra- and inter-cluster variance measures to enhance lesion segmentation. We will also conduct evolving deep architecture generation for other complex computer vison tasks [76, 77], such as medical image segmentation, object tracking, salient object detection, and visual question generation. The proposed ensemble, deep and clustering models could also be incorporated to identify any new, unseen lesion case on an incremental basis.

ACKNOWLEDGEMENT

This research is supported by the European Union (EU) sponsored (Erasmus Mundus) cLINK (Centre of excellence for Learning, Innovation, Networking and Knowledge) project (EU Grant No. 2645).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] D. Farid, L. Zhang, A.M. Hossain, C.M., Rahman, R. Strachan, G. Sexton, K. Dahal. (2013). An Adaptive Ensemble Classifier for Mining Concept-Drifting Data Streams. *Expert Systems with Applications*, 40 (15) 5895-5906.

[2] L. Zhang, K. Mistry, M. Jiang, S.C. Neoh, A. Hossain. (2015). Adaptive facial point detection and emotion recognition for a humanoid robot. *Computer Vision and Image Understanding*, 140 (2015) 93-114.

[3] S.C. Neoh, W. Srisukkham, L. Zhang, S. Todryk, B. Greystoke, C.P. Lim, A. Hossain, N. Aslam. (2015). An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images, *Scientific Reports*. 5 (14938) 1-14. Nature Publishing Group.

[4] A. Mekhmoukh, K. Mokrani. (2015). Improved Fuzzy C-Means based Particle Swarm Optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation. *Computer Methods and Programs in Biomedicine*. 122 (2) 266-281.

[5] Y. Yuan, Y. Lo. (2018). Improving Dermoscopic Image Segmentation with Enhanced Convolutional-Deconvolutional Networks. *IEEE Journal of Biomedical and Health Informatics*. 1-8. (In Press).

[6] Y. Li, L. Shen. (2018). Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. *Sensors*. 18 (2), 556. 2018.

[7] D. Alvarez, M. Iglesias. (2017). k-Means Clustering and Ensemble of Regressions: An Algorithm for the ISIC 2017 Skin Lesion Segmentation Challenge. CoRR abs/1702.07333 (2017).

[8] A.S. Ashour, Y. Guo, E. Kucukkulahli, P. Erdogmus, K. Polat. (2018). A hybrid dermoscopy images segmentation approach based on neutrosophic clustering and histogram estimation. *Applied Soft Computing*. 69 (2018) 426–434.

[9] J.L. García-Arroyo, B. Garcia-Zapirain. (2017). Segmentation of skin lesions based on fuzzy classification of pixels and histogram thresholding. CoRR abs/1703.03888.

[10] Y. Guo, A.S. Ashour, F. Smarandache. (2018). A Novel Skin Lesion Detection Approach Using Neutrosophic Clustering and Adaptive Region Growing in Dermoscopy Images. *Symmetry*. 10 (4) 119.

[11] B.S. Lin, K. Michael, S. Kalra, H.R. Tizhoosh. (2017). Skin lesion segmentation: U-Nets versus clustering. In *Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI)*. USA.

[12] T.Y. Tan, L. Zhang, S.C. Neoh, C.P. Lim. (2018). Intelligent Skin Cancer Detection Using Enhanced Particle Swarm Optimization. *Knowledge-Based Systems*. (In Press).

[13] A.R. Lopez, X. Giro-i-Nieto, J. Burdick, O. Marques. (2017). Skin lesion classification from dermoscopic images using deep learning techniques. In *Proceedings of 13th IASTED International Conference on Biomedical Engineering (BioMed)*, 2017.

[14] T. Devries, D. Ramachandram. (2017). Skin Lesion Classification Using Deep Multi-scale Convolutional Neural Networks. CoRR abs/1703.01402.

[15] L. Bi, K. Kim, E. Ann, D. Feng. (2017). Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197.

[16] G. Schaefer, B. Krawczyk, M.E. Celebi, H. Iyatomi. (2014). An ensemble classification approach for melanoma diagnosis, *Memetic Computing*, 6 (4) 233-240.

[17] J. Kawahara, A. BenTaieb, G. Hamarneh. (2016). Deep Features to Classify Skin Lesions. In *Proceedings of IEEE 13th International Symposium on Biomedical Imaging (ISBI)*.

[18] X.S. Yang. (2008). *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, UK.

[19] J. Kennedy, R. Eberhart. (1995). Particle Swarm Optimization, In *Proceedings of IEEE Int. Conf. Neural Networks*, vol. 4, 1942–1948.

[20] X.S. Yang. (2010). Firefly algorithm, levy Flights and global optimization, *Research and Development in Intelligent Systems*. 26 (2010) 209–218.

[21] Y. Zhang, L. Zhang, S.C. Neoh, K. Mistry, A. Hossain. (2015). Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles, *Expert Systems with Applications*. 42 (22) 8678–8697.

[22] A. Kazem, E. Sharifi, F.K. Hussain, M. Saberic, O.K. Hussain. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting, *Applied Soft Computing*. 13 (2) 947–958.

[23] A. Bouyer, A. Hatamlou. (2018). An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms. *Applied Soft Computing*, 67 (2018) 172–182.

[24] D. Pandit, L. Zhang, S. Chattopadhyay, C.P. Lim, C. Liu. (2018). A Scattering and Repulsive Swarm Intelligence Algorithm for Solving Global Optimization Problems. *Knowledge-Based Systems*. (In Press).

[25] K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim, B. Fielding. (2017). A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition, *IEEE Transactions on Cybernetics*. PP (99) 1-14.

[26] L. Zhang, M. Jiang, D. Farid, A. Hossain. (2013). Intelligent Facial Emotion Recognition and Semantic-based Topic Detection for a Humanoid Robot, *Expert Systems with Applications*. 40 (13) 5160-5168.

[27] S.C. Neoh, L. Zhang, K. Mistry, M.A. Hossain, C.P. Lim, N. Aslam, P. Kinghorn. (2015). Intelligent Facial Emotion Recognition Using a Layered Encoding Cascade Optimization Model, *Applied Soft Computing*. 34 (2015) 72–93.

[28] L. Zhang, K. Mistry, S.C. Neoh, C.P. Lim. (2016). Intelligent facial emotion recognition using moth-firefly optimization. *Knowledge-Based Systems*, 111 (2016) 248-267.

[29] D. Pandit, L. Zhang, C. Liu, S. Chattopadhyay, N. Aslam, C.P. Lim. (2017). A lightweight QRS detector for single lead ECG signals using a max-min difference algorithm. *Computer Methods and Programs in Biomedicine*, 144 (2017) 61-75.

[30] L. Zhang, K. Mistry, C.P. Lim, S.C. Neoh. (2018). Feature Selection Using Firefly Optimization for Classification and Regression Models. *Decision Support Systems*. 106 (2018) 64–85.

[31] W. Srisukkham, L. Zhang, S.C. Neoh, S. Todryk, C.P. Lim. (2017). Intelligent Leukaemia Diagnosis with Bare-Bones PSO based Feature Optimization. *Applied Soft Computing*, 56 (2017) 405-419.

[32] T.Y. Tan, L. Zhang, M. Jiang. (2016). An intelligent decision support system for skin cancer detection from dermoscopic images. In *Proceedings of 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. 2016. 2194-2199.

[33] L. Zhang, W. Srisukkham, S.C. Neoh, C.P. Lim, D. Pandit. (2018). Classifier ensemble reduction using a modified firefly algorithm: An empirical evaluation. *Expert Systems with Applications*. 93 (2018) 395–422.

[34] K. Tamura, K. Yasuda. (2011). Primary study of spiral dynamics inspired optimization. *IEEJ Transactions on Electrical and Electronic Engineering*. 6 (S1) 98-100.

[35] K. Tamura, K. Yasuda. (2011). Spiral Dynamics Inspired Optimization. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 15 (8) 1116-1122.

[36] K. Thangavel, J. Bagyamani, R. Rathipriya. (2012). Novel hybrid PSO-SA model for biclustering of expression data. *Procedia Engineering*. 30 (2012) 1048-1055.

[37] A. Hatamlou. (2014). Heart: a novel optimization algorithm for cluster analysis. *Progress in Artificial Intelligence*. 2 (2-3) 167-173.

[38] S. Farzi, S. Kianian. (2018). A novel clustering algorithm for attributed graphs based on K-medoid algorithm. *Journal of Experimental & Theoretical Artificial Intelligence*. 30 (6) 795-809.

[39] C. Yilmaz, H.T. Kahraman, S. Söyler. (2018). Passive Mine Detection and Classification Method Based on Hybrid Model. *IEEE Access*. 6, 47870-47888.

[40] P. Nerurkar, A. Shirke, M. Chandane, S. Bhirud. (2018). A novel heuristic for evolutionary clustering. *Procedia Computer Science*. 125 (2018) 780-789.

[41] N. Nananukul. (2013). Clustering model and algorithm for production inventory and distribution problem. *Applied Mathematical Modelling*. 37 (24) 9846-9857.

[42] D. Xu, Y. Tian. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*. 2 (2) 165-193.

[43] J.M. Moyano, E.L. Gibaja, K.J. Cios, S. Ventura. (2019). An evolutionary approach to build ensembles of multi-label classifiers. *Information Fusion*, 50 (2019) 168-180.

[44] G.T. Ribeiro, V.C. Mariani, L. dos Santos Coelho. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Engineering Applications of Artificial Intelligence*, 82 (2019) 272-281.

[45] P. Thomas, H.B. El Haouzi, M.C. Suhner, A. Thomas, E. Zimmermann, M. Noyel. (2018). Using a classifier ensemble for proactive quality monitoring and control: The impact of the choice of classifiers types, selection criterion, and fusion process. *Computers in Industry*, 99 (2018) 193-204.

[46] D. Guo, Y. Jin, J. Ding, T. Chai. (2018). Heterogeneous ensemble-based infill criterion for evolutionary multiobjective optimization of expensive problems. *IEEE Transactions on Cybernetics*, (99), 1-14.

[47] R.C.T. De Souza, L. dos Santos Coelho, C.A. De Macedo, J. Pierezan. (2018). A V-Shaped Binary Crow Search Algorithm for Feature Selection. In *Proceedings of 2018 IEEE Congress on Evolutionary Computation (CEC)*, 1-8. IEEE.

[48] H. Song, A.K. Qin, F.D. Salim. (2018). Evolutionary Multi-objective Ensemble Learning for Multivariate Electricity Consumption Prediction. In *Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN)*, 1-8. IEEE.

[49] S.F. Bushehri, M.S. Zarchi. (2019). An expert model for self-care problems classification using probabilistic neural network and feature selection approach. *Applied Soft Computing*, p.105545.

[50] B. Xue, M. Zhang, W.N. Browne, X. Yao. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20 (4) 606-626.

[51] S. Mirjalili. (2015). Moth-Flame optimization algorithm: A novel nature-inspired heuristic paradigm, *Knowledge-Based Systems*. 89 (2015) 228–249.

[52] J. Holland. (1975). *Adaptation in Natural and Artificial Systems*, Cambridge, MA, USA: MIT Press.

[53] J. Han, M. Kamber, J. Pei. (2011). *Data mining: concepts and techniques.* 3rd edn, Morgan Kaufmann.

[54] B.C. Kuo, D.A. Landgrebe. (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*. 42 (5) 1096-1105.

[55] R. Kirsch. (1971). Computer determination of the constituent structure of biological images. *Computers and Biomedical Research*. 4: 315–328.

[56] G. Kylberg, I.M. Sintorn. (2013). Evaluation of noise robustness for local binary pattern descriptors in texture classification. *EURASIP Journal on Image and Video Processing*. 17 (2013).

[57] L. Ballerini, R.B. Fisher, R.B. Aldridge, J. Rees. (2013). A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions, in: M.E. Celebi, G. Schaefer (Eds.), *Color Medical Image Analysis*, Lecture Notes in Computational Vision and Biomechanics, vol 6, 63–86.

[58] T. Mendonca, P.M. Ferreira, J.S. Marques, A.R. Marcal, J. Rozeira. (2013). PH2 – a dermoscopic image database for research and benchmarking, In *Proceedings of EMBC 2013 – 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Japan, pp. 5437–5440.

[59] N. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern. (2017). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). arXiv: 1710.05006.

[60] A.R. Jordehi. (2015). Enhanced leader PSO (ELPSO): a new PSO variant for solving global optimisation problems, *Applied Soft Computing*. 26 (2015) 401–417.

[61] S. Mirjalili, A. Lewis, A.S. Sadiq. (2014). Autonomous particles groups for particle swarm optimization, *Arabian Journal for Science and Engineering*. 39 (6) 4683–4697.

[62] M. Nasir, S. Das, D. Maity, S. Sengupta, U. Halder, P.N. Suganthan. (2012). A dynamic neighborhood learning based particle swarm optimizer for global numerical optimization, *Information Sciences*. 209 (2012) 16–36.

[63] Q. Chen, Y. Chen, W. Jiang. (2016). Genetic particle swarm optimization-based feature selection for very-high-resolution remotely sensed imagery object change detection, *Sensors*. 16 (8) (2016) 1204.

[64] H. Wang, W. Wang, X. Zhou, H. Sun, J. Zhao, X. Yu, Z. Cui. (2017). Firefly algorithm with neighborhood attraction. *Information Sciences*, 382–383 (2017) 374–387.

[65] A.H. Gandomi, X.S. Yang, S. Talatahari, A.H. Alavi. (2013). Firefly algorithm with chaos, *Communications in Nonlinear Science and Numerical Simulation*, 18, 89–98.

[66] L. He, S. Huang. (2017). Modified firefly algorithm based multilevel thresholding for colour image segmentation, *Neurocomputing*, 240 (2017) 152-174.

[67] S.H. Yu, S.L. Zhu, Y. Ma, D.M. Mao. (2015). A variable step size firefly algorithm for numerical optimization, *Applied Mathematics and Computation*. 263, 214–220.

[68] J. J. Liang, B-Y. Qu, P.N. Suganthan. (2013). Problem Definitions and Evaluation Criteria for the CEC 2014 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization, *Technical Report 201311*, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China and *Technical Report*, Nanyang Technological University, Singapore.

[69] O. Ronneberger, P. Fischer, T. Brox. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer Assisted Interventions*, Munich, Germany, 5–9 October 2015, pp. 234–241.

[70] E. Shelhamer, J. Long, T. Darrell. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015. pp. 3431–3440.

[71] H. Wen. (2017). II-FCN for skin lesion analysis towards melanoma detection. arXiv:1702.08699.

[72] B. Fielding, L. Zhang. (2018). Evolving Image Classification Architectures with Enhanced Particle Swarm Optimisation. *IEEE Access*. Vol 6. 68560–68575.

[73] P. Mirunalini, C. Aravindan, V. Gokul, S.M. Jaisakthi. (2017). Deep Learning for Skin Lesion Classification. CoRR abs/1703.04364.

[74] H. Chang. (2017). Skin cancer reorganization and classification with deep neural network. CoRR abs/1703.00534.

[75] W. Zhang, L. Gao, R. Liu. (2017). Using Deep Learning Method for Classification: A Proposed Algorithm for the ISIC 2017 Skin Lesion Classification Challenge. CoRR abs/1703.02182.

[76] P. Kinghorn, L. Zhang, L. Shao. (2018). A region-based image caption generator with refined descriptions. *Neurocomputing*. 272 (2018) 416-424.

[77] T.Y. Tan, L. Zhang, C.P. Lim, B. Fielding, Y. Yu, E. Anderson. (2019). Evolving Ensemble Models for Image Segmentation Using Enhanced Particle Swarm Optimization. *IEEE Access*. 7, 34004-34019.