Intelligent Skin Cancer Diagnosis Using Improved Particle Swarm Optimization and Deep Learning Models

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 an intelligent decision support system for skin cancer detection. Since generating an effective lesion representation is a vital step to ensure the success of lesion classification, the discriminative power of different types of features is exploited. Specifically, we combine clinically important asymmetry, border irregularity, colour and dermoscopic structure features with texture features extracted using Grey Level Run Length Matrix, Local Binary Patterns, and Histogram of Oriented Gradients operators for lesion representation. Then, we propose two enhanced Particle Swarm Optimization (PSO) models for feature optimization. The first model employs adaptive acceleration coefficients, multiple remote leaders, in-depth sub-dimension feature search and re-initialization mechanisms to overcome stagnation. The second model uses random acceleration coefficients, instead of adaptive ones, based on non-linear circle, sine and helix functions, respectively, to increase diversification and intensification. Ensemble

classifiers are also constructed with each base model trained using each optimized feature subset. A deep convolutional neural network is devised whose hyper-parameters are fine-tuned using the proposed PSO models. Extensive experimental studies using dermoscopic skin lesion data, medical data from the UCI machine learning repository, and ALL-IDB2 image data are conducted to evaluate the model efficiency systematically. The results from empirical evaluations and statistical tests indicate the superiority of the proposed models over other advanced PSO variants and classical search methods pertaining to discriminative feature selection and optimal hyper-parameter identification for deep learning networks in lesion classification as well as other disease diagnosis.

Keywords: Skin Cancer Detection, Feature Selection, Hyper-parameter Tuning, Evolutionary Algorithm, and Deep and Ensemble Classifier.

1. INTRODUCTION

Melanoma is the deadliest form of skin cancer. Accurate and early detection could significantly increase the survival rate [1-3]. Owing to the visual similarity between non-melanoma and melanoma dermoscopic images, it is a challenging task in attaining precise diagnosis of different lesion cases. Since generating an effective lesion representation is a vital step to ensure the success of lesion classification, we explore dermoscopic feature extraction and evolutionary algorithm-based lesion feature selection in this research.

Specifically, to formulate an effective lesion representation, we employ several feature descriptors to obtain not only clinically important asymmetry, border irregularity, colour and dermoscopic structure (ABCD) features, but also textual features using Grey Level Run Length Matrix (GLRLM), Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG) operators. In this way, the discriminating power of different types of features is explored. Then, two variants of the Particle Swarm Optimization (PSO) model are proposed for feature selection. The objective is to identify the most significant characteristics pertaining to different lesions. The first proposed PSO algorithm employs dynamic adaptive acceleration coefficients, multiple remote swarm leaders, sub-dimension regional lesion feature enhancement, and re-initialization mechanisms to overcome

stagnation and accelerate convergence. The second proposed PSO model uses random acceleration coefficients, instead of adaptive ones, based on non-linear circle, sine and helix functions to increase diversification and intensification.

After identifying the most significant features by the proposed PSO models, two ensemble models are built with three K-Nearest Neighbour (KNN) models and three Support Vector Machines (SVMs) as the base classifiers. Ensemble models are known to yield better performances as compared with single classifiers. The base classifiers of each ensemble model are generated using the identified three sets of optimal features, i.e. ABCD+GLRLM, LBP, and HOG feature subsets. A weighted majority voting is adopted to draw a final conclusion on benign and malignant lesion classification. Moreover, a deep convolutional neural network (CNN) is employed for melanoma classification. Optimal hyper-parameter identification of the deep CNN is performed using the proposed PSO models, owing to the importance of these hyper-parameters in affecting the classification performance. Figure 1 depicts the overall system architecture. The research contributions are summarized as follows.

1. A set of shape, colour and textural features (ABCD+GLRLM) is first extracted, owing to their significance in clinical diagnosis. The LBP operator with a (8, 1) circular neighbourhood and the HOG operator with an overlap of half the block size are used to extract refined textural information for lesion representation.
2. Since generating an effective lesion representation is vital, two modified PSO models are proposed for discriminative feature selection. The first modified PSO model not only combines lesion features for performing global search, but also separates lesion features into specific areas for in-depth local search. Firstly, three remote swarm leaders with competitive fitness scores but low correlation in positions are identified. The overall population is then randomly split into three subswarms with each leader leading the search of each subswarm, respectively. Dynamic descending and ascending acceleration coefficients are generated by using partial circle, sine and helix waveforms to lead the search in the three subswarms, respectively. Such adaptive coefficients are able to increase global exploration in early iterations and converge towards the global best solution towards the end of iterations. Sub-dimension-based

search is also conducted to obtain more refined regional discriminative information for lesion representation. A re-initialization strategy is used to assign random positions for the weakest solutions in the swarm, in order to diversify the search.

3. Instead of using adaptive coefficients, the second proposed PSO model employs random acceleration coefficients based on the full waveforms generated by non-linear circle, sine, and helix functions. Such random coefficients not only increase diversification and enable a wider exploration of the search space, but also increase intensification to fine-tune the regions of both local and global optimal solutions. Specifically, the second model simulates the hovering flight behaviours of hummingbirds around attraction (e.g. flowers), and is able to explore a wider search space, therefore is more capable of finding global optima.
4. Besides ensemble classifiers, a deep CNN model with adaptive parameter tuning is employed to further assess efficiency of the proposed PSO models. The hyper-parameters, such as the learning rate and regularization strength, have important effects on the CNN results. As such, these parameters are adaptively adjusted and fine-tuned for each lesion recognition task using the proposed PSO models. Evaluated using dermoscopic skin lesion, UCI and ALL-IDB2 image data sets, the results from empirical evaluations and statistical tests indicate the superiority of the proposed PSO models over 11 advanced PSO variants and 10 classical search methods for discriminative feature selection and optimal hyper-parameter identification.

The paper is organized as follows. Section 2 presents the related studies on benign and malignant lesion classification and Evolutionary Computation (EC) techniques for feature selection. Details of ABCD+GLRLM, LBP, and HOG feature extraction, the proposed PSO-based feature selection, and ensemble and deep networks for lesion classification are explained in Section 3. The evaluation results and discussion are presented in Section 4. A summary of this research and the directions for further research are given in Section 5.

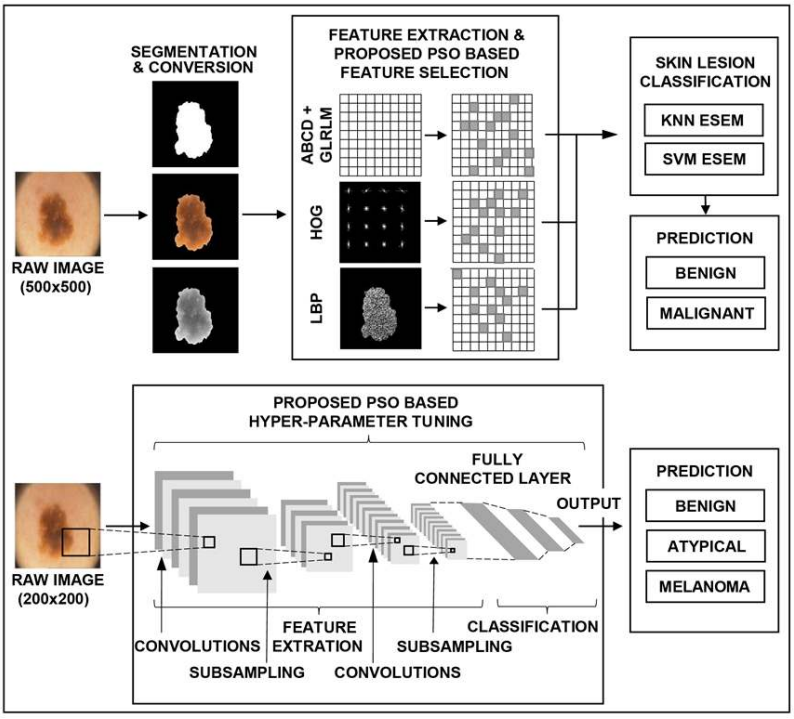

Figure 1. The system architecture

2. RELATED WORK

In this section, we introduce diverse evolutionary algorithm-based feature selection methods as well as related research on automated skin cancer detection.

2.1 Feature Selection

Owing to the powerful global search capabilities, evolutionary algorithms such as PSO are widely used for feature selection problems. Nayak et al. [4] employed a modified PSO (MPSO) model to optimize the input weights and hidden biases of an Extreme Learning Machine (ELM) for pathological brain detection from MR images. Both time-varying acceleration coefficients and an inertia weight factor were employed in MPSO to enable the search process to focus on global exploration in early iterations and converge towards the global optima in final iterations. Issa et al. [5] proposed a hybrid optimization method, known as ASCA-PSO, which integrated an enhanced sine cosine optimization algorithm (SCA) with the PSO model. The search process was split into two layers, with the top and bottom layers dedicated to the search strategies of PSO and SCA, respectively. The bottom layer used SCA to conduct subswarm-based search. The identified subswarm

leaders were stored in the upper layer, which employed the PSO mechanism to explore the search space. Evaluated with well-known benchmark functions, their model showed more chances of finding global optima by combining intensification and diversification processes in each iteration. Patwal et al. [6] developed another modified PSO model, namely time varying acceleration coefficient PSO with mutation strategies (TVAC-PSOMS). Besides employing adaptive acceleration coefficients, Cauchy, Gaussian and opposition based mutations were used to further enhance the local best solutions respectively. TVAC-PSOMS was used to solve hydrothermal generation scheduling problems, and it outperformed other modified search methods. Xue et al. [7] proposed a modified PSO model for feature selection in classification. The swarm initialization was conducted using both small and relatively large feature subsets. Classification performance was also used as the highest priority for updating both personal and global best solutions. Tested with 20 data sets, their PSO model, with a reduced feature subset, outperformed the original PSO model with a single objective function combining the two objectives of maximizing performance and minimizing the number of selected features. Lu et al. [8] proposed six PSO models for feature selection in text categorization, including PSO with (1) fixed inertia weight, (2) liner decreasing inertia weight, (3) fixed constriction factor, (4) adaptive constriction factor, (5) synchronous inertia weight and constriction factor, and (6) asynchronous inertia weight and constriction factor, respectively. Their experiments indicated the superiority of the PSO model with asynchronous inertia weight and constriction factor among all the proposed methods in terms of classification performance and convergence speed. Chuang et al. [9] proposed a chaotic binary PSO (BPSO) model for feature selection. It employed two chaotic maps, i.e. Logistic and Tent maps, as the inertia weight in BPSO, respectively. Evaluated with 10 UCI data sets, BPSO with the Tent map showed better feature selection performances in comparison with those of BPSO with the Logistic map. A hybrid PSO model, known as GMPSO [10], was proposed for static and dynamic feature selection in bodily expression regression by integrating PSO with a Genetic Algorithm (GA) and Gaussian/Cauchy/Levy distributions. A modified bare-bones PSO model, denoted as BBPSOV [11], incorporating new position updating mechanisms, was used to conduct shape, colour, and texture feature selection in leukaemia diagnosis. PSO incorporated with a micro-GA concept was proposed for horizontal and vertical LBP-based texture feature selection for facial

expression recognition [12], whereas a threshold-based PSO model (ThBPSO) was developed in [13] for pose and illumination invariant facial feature selection in face recognition. Three enhanced Whale Optimization Algorithms (WOA) were proposed in [14] for feature selection, which incorporated WOA with tournament, roulette wheel selection and crossover and mutation operators, respectively.

2.2 Skin Cancer Detection

In our previous study [1], we developed a decision support system for dermoscopic skin cancer diagnosis. A total of 11 shape, 15 colour, and 3888 Generalised Co-Occurrence Matrix (GCM) texture features were extracted. A GA was employed for feature selection while a SVM classifier was used for benign/malignant lesion classification. Owing to the severe imbalanced shape, colour, and texture features as well as low discriminative capabilities of GCM features, in this research, we employ more effective texture descriptors, which include GLRLM, LBP, and HOG, for texture generation. More discriminative shape (13) and colour (87) features are also extracted. Since the classical GA is likely to be trapped in local optima, two PSO variants are proposed for feature selection. The empirical results indicate their superior discriminative capabilities for feature selection. We also employ ensemble classifiers and deep CNN models for lesion classification, in order to further enhance performance. Moreover, Adjed et al. [2] employed the fusion of structural features and textual features for melanoma classification. Wavelet and curvelet transforms and diverse LBP variants were used for structural and textural feature extraction, respectively. Tested with the PH2 data set using a SVM model and a random sampling cross-validation strategy, their study achieved an accuracy rate of 86.07%. Sáez et al. [3] conducted thickness classification for melanoma lesions. Properties and features associated with dermoscopic images and tumour depth were explored. A total of 81 morphological, colour, and textural features were derived for lesion representation. Both 2-class (thin and thick) and 3-class (thin, intermediate, and thick) categorization schemes were proposed. Logistic regression using Initial variables and Product Units (LIPU) obtained the highest accuracy for binary classification tasks. For the 3-class thickness classification, a set of ordinal classification methods was used, which performed more stably in comparison with LIPU. Yu et al. [15] proposed deep CNNs for melanoma classification. First, a fully convolutional residual network (FCRN) was used

for skin lesion segmentation. Further enhancement was realized by combining the FCRN with a multi-scale contextual information integration mechanism. A deep CNN model with more than 50 layers was then used to extract discriminative features from the segmented lesion regions for melanoma classification. A residual learning mechanism was also employed for training both networks to avoid overfitting. Tested with the ISBI 2016 Challenge data set, their model achieved the first and second places for classification and segmentation tasks, respectively. Barata et al. [16] proposed global and local methods for melanoma classification, and indicated the efficiency of colour features over texture features for melanoma identification.

Codella et al. [17] proposed ensembles of deep learning models for melanoma detection. The lesion segmentation was conducted using a fully convolutional network, which has a similar structure to that of U-Net. Besides sparse coding and hand-coded feature representations (e.g. a multi-scale variant of colour LBP), deep residual networks, CNNs, and fully convolutional U-Net architecture were also employed for feature extraction. A set of SVM classifiers was subsequently trained with each SVM dedicated to each extracted feature vector. The final classification results were obtained by averaging the outputs of all SVMs. Kruk et al. [18] proposed a skin cancer detection system, which consisted of image filtering, comprehensive ABCD feature extraction, feature selection, and non-melanoma/melanoma classification. A set of powerful descriptors was employed for feature extraction including Kolmogorov–Smirnov statistical distance, maximum sub-region principle, percolation theory, and fractal texture analysis. Moreover, Fisher discriminant measure, correlation feature selection, and fast correlation-based filter were employed for discriminative feature selection. SVM and random forest (RF) were applied to melanoma classification. Their empirical results indicated that Fisher discriminant measure integrated with the SVM model obtained the best performance. Sánchez-Monedero et al. [19] conducted not only melanoma classification but also the identification of different stages of lesions. A set of 100 morphological, colour, textural and pigment network features was extracted for each lesion image. A five-class classification was performed to identify benign and four other melanoma cases with different stages. A partial order assumption was considered in their methodologies owing to the presence of an order relationship among different melanoma stages and the absence

of such a relationship in the benign cases. Over-sampling techniques were also employed for the generation of synthetic images to deal with imbalanced data problems. Bozorgtabar et al. [20] proposed deep networks guided by local unsupervised learning for lesion segmentation, while Li and Shen [21] developed a comprehensive deep architecture consisting of the Lesion Indexing Network (LIN) and the Lesion Feature Network (LFN) for skin lesion classification.

3. THE PROPOSED SYSTEM

In this research, we propose intelligent melanoma classification consisting of three key stages. Firstly, different feature descriptors are used to extract not only clinically significant ABCD characteristics, but also both high-level and low-level textural features using GLRLM, LBP, and HOG descriptors, respectively. Secondly, we propose two PSO models for discriminative feature selection. The first PSO model employs adaptive decreasing and increasing coefficients and in-depth local and global search for feature selection. The second model is equipped with the capability of exploring an extended search space by using random coefficients to attain global optima. Finally, two ensemble classifiers are generated for lesion classification using the identified three sets of optimal features, i.e. (1) ABCD+GLRLM, (2) LBP, and (3) HOG feature subsets. A deep CNN model is also implemented. Its optimal hyper-parameters are identified using the proposed PSO models. We introduce each key stage comprehensively, as follows.

3.1 Feature Extraction

First of all, the lesion images used in this research are re-sized to 500×500. Several pre-processing steps are conducted including noise filtering, image segmentation, and grayscale conversion for noise removal and contrast enhancement. We extract ABCD morphological and colour properties, high-level GLRLM features, as well as low-level LBP and HOG histograms for representing skin lesions.

We extract ABCD characteristics and GLRLM textural features owing to their clinical significance in medical skin cancer diagnosis. The following morphological features are extracted, i.e. asymmetry, compactness, border irregularity, perimeter, solidity, extent, etc. The extracted colour features include colour variance, entropy, skewness, correlation, etc,

while the obtained textural features contain GLRLM in 4 orientations with 11 different emphases. Overall, a set of 146 ABCD and high-level GLRLM textural features (13 shape, 87 colour and 46 texture features) is obtained.

Since LBP and HOG operators are popular techniques to represent texture deformations, we employ both operators for feature extraction. Owing to computational simplicity and robustness to illumination changes, LBP has been successfully used for a wide range of image processing problems. According to [12], the LBP operator with a neighbourhood of 8 and a radius of 1 is able to extract 90% discriminative uniform patterns among all the patterns retrieved, which play key roles in texture classification. Therefore, the LBP operator with a neighbourhood of 8 and a radius of 1 is employed in this research. We use a cell size of 191×191 in our experiment. A total of 236 features are extracted using the LBP operator. Since different texture descriptors are able to extract distinguishing textural deformations, the HOG operator is also used in this research. It employs a cell size of 110×110 with a block size of 4×4. In order to ensure sufficient contrast normalization, the HOG operator with an overlap of half the block size is employed. A total of 324 features are extracted.

The above extracted skin lesion representations using high-level textural, morphological and colour features, and low-level LBP and HOG features have high dimensionality. Two PSO-based feature selection models are proposed to remove redundant attributes and to extract the most significant characteristics from the above three sets of features, respectively, for benign and melanoma classification.

3.2 The Proposed PSO-based Feature Selection

Proposed in [22], PSO is one of the popular swarm intelligence algorithms, and it shows impressive search capabilities in solving diverse optimization problems. It employs the personal and global best solutions to guide the search process. Equations (1)-(2) define the velocity and position updating operations in PSO.

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_1 * (p_{id} - x_{id}^{t}) + c_2 * r_2 * (p_{gd} - x_{id}^{t}) \quad (1)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (2)$$

where x_{id}^{t+1} and x_{id}^{t} represent the positions of particle i in the d-th dimension in t+1-th and t-th iterations respectively. v_{id}^{t+1} and v_{id}^{t} denote the velocities in the t+1-th and t-th iterations, respectively. w is the inertia weight to embed iteration influence of the previous velocity. c_1 and c_2 are acceleration coefficients with r_1 and r_2 as random vectors. Moreover, p_{id} and p_{gd} represent the personal best solution of particle i and the global best solution in the d-th dimension, respectively. Since the search process of PSO is led by a single swarm leader, it is more likely to be trapped in local optima.

In this research, we propose two modified PSO models to mitigate the premature convergence problem of the original PSO for lesion feature optimization. Instead of using fixed acceleration coefficients, dynamic coefficients are proposed in both PSO models. Specifically, dynamic ascending and descending acceleration coefficients, in-depth local search and re-initialization mechanisms are used in the first PSO model. A wider exploration using both positive and negative random acceleration coefficients oriented by non-linear functions is implemented in the second PSO model. We discuss the proposed PSO algorithms in detail, as follows.

3.2.1 The First Proposed PSO Model

The first proposed PSO model with adaptive coefficients is denoted as ACPSO. It employs adaptive cognitive and social components, sub-dimension based search, and re-initialization mechanisms to diversify the search. Algorithm 1 shows the pseudo-codes of ACPSO.

Algorithm 1: Pseudo-Codes of Proposed ACPSO
1. **Start**
2. Initialize a particle swarm randomly;
3. Evaluate each particle using the fitness/objective function $f(x)$;
4. Rank the swarm based on fitness values and identify g_{best}.
5. **While** (Stopping criterion is not satisfied) {
6. Identify the second and third swarm leaders with comparable

	fitness scores but remote in positions to g_{best};
7.	Randomly divide the swarm into three subswarms with each subswarm led by each leader;
8.	**While** (!Stagnation)
9.	{ //use each strategy to lead each subswarm-based search
10.	Update position using Eqn. (1)-(3) in subswarm 1;
11.	Update position using Eqn. (1), (2) and (4) in subswarm 2;
12.	Update position using Eqn. (1), (2) and (5)-(7) in subswarm 3;
13.	Evaluate each particle in each subswarm and update its personal best and the subswarm leaders;
14.	Compare three subswarm leaders and store the best subswarm leader;
15.	**}Until** (stagnation detected)
16.	Combine the three subswarms and update g_{best};
17.	Conduct sub-dimension based search in Algorithm 2 and return a new best solution, g'_{best};
18.	**If** (g'_{best} is fitter than g_{best})
19.	{ Replace the worst particle in the swarm with g_{best};
20.	Update g_{best} with g'_{best}; }
21.	**Else If**
22.	{ Replace the worst particle in the swarm with g'_{best};}
23.	**EndIf**
24.	Replace the last second and third worst solutions in the swarm with two newly generated random particles;
25.	}**Until (Stagnate several times and no improvement found)**
26.	Output g_{best};
27.	**End**

In Algorithm 1, after initializing the swarm and identifying the swarm leader, g_{best}, the second and third swarm leaders which show comparable fitness scores but a low correlation in position to g_{best} are identified. Then, the overall swarm is randomly divided into three subswarms, with each leader leading the search in each subswarm. Three sets of non-linear ascending or descending functions are used to generate dynamic acceleration coefficients in three subswarms to lead adaptive cognitive and social search components. Specifically, c_1 and c_2 are generated as partial descending and ascending contours of circle, sine, and helix functions, respectively, to increase local and global exploration. After a number of iterations, three subswarm leaders are produced, with the best subswarm leader used to update g_{best}. The three subswarms are subsequently merged, and a sub-dimension based search defined in Algorithm 2 is conducted using the overall swarm to further enhance g_{best}. Specifically, we divide each particle into three sub-

dimensions. These sub-dimensions represent the search areas of top, middle, and bottom sections of each lesion image. The sub-dimension based search performs local exploitation of the lesion sub-regions (i.e. feature subsections) to generate a new global best solution, g'_{best}. It updates g_{best} if it is fitter than g_{best}, otherwise it replaces the worst particle in the swarm. A re-initialization strategy is also used to replace the last second and third worst solutions in the swarm with two newly generated random particles. The above search process iterates until the stopping criteria are fulfilled.

3.2.1.1 Descending and ascending acceleration coefficients

We introduce the three proposed non-linear functions for adaptive coefficient generation. Equation (3) defines a circle function for descending and ascending adaptive coefficient generation in subswarm 1. Specifically, c_1 and c_2 are generated as partial descending and ascending contours of circles, respectively.

$$y_1 = \sqrt{radius^2 - k^2} \qquad k \in [0, 2.5] \tag{3}$$

In Equation (3), $radius$ denotes the radius of a circle with a value of 2.5, i.e. the maximum value for both coefficients, while $k \in [0, 2.5]$ is the input value for circle generation. In order to produce a descending coefficient $c_1 \in (0, 2.5]$, the upper right quarter of the circle generated by Equation (3) is divided into *m* portions evenly, as indicated in Figure 2, where *m* is the maximum iteration number. Therefore, each portion is used for the generation of c_1 in each iteration. Specifically, we start the traverse of all the *m* portions from 90 to 0 degree, and randomly select one value from the circle contour in each portion and assign it as c_1. This process ensures the generation of a series of descending c_1. On the contrary, by traversing through all the *m* portions from 0 to 90 degree, a series of ascending c_2 is generated, where $c_2 \in (0, 2.5]$.

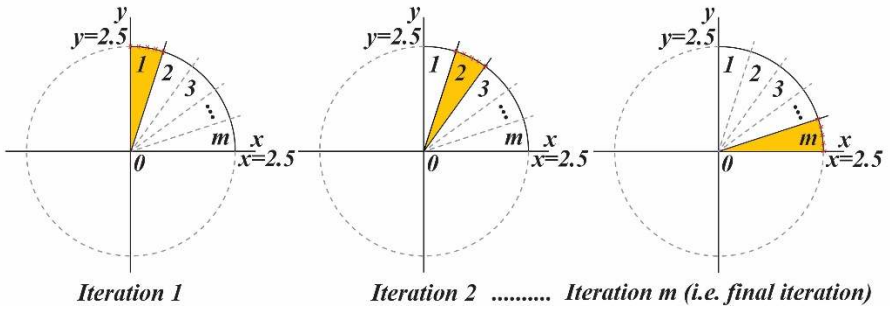

Figure 2. Dynamic coefficient generation using the circle function

Therefore, c_1 is decreased while c_2 is increased over the generation. The impact of the personal best and global best experiences to guide the search process is, therefore, dynamically adjusted. When $c_1 > c_2$, the subswarm-based search focuses more on local exploitation, whereas when $c_2 > c_1$, the search process emphasizes more on global exploration.

Similarly, Equation (4) defines a sine function for dynamic coefficient generation in subswarm 2. Specifically, c_1 and c_2 are generated as partial descending and ascending contours of a sine function, respectively.

$$y_2 = 2.5 * \sin(l) \quad l \in [\pi/2, \pi] \tag{4}$$

In Equation (4), l is the input value. Figure 3 illustrates the generated sine waveform defined in Equation (4). Similar to the abovementioned circle-oriented coefficient generation, we split the upper right contour of the sine waveform into *m* portions. One value from each portion is randomly selected for the generation of each coefficient. A series of descending c_1 is produced by traversing through all the *m* portions from 90 to 0 degree, and vice versa for the generation of a series of ascending c_2. Both coefficients are dynamically adjusted through iterations. When $c_1 > c_2$, the search process is more strongly guided by the personal best experiences, whereas when $c_2 > c_1$, the global best experiences show more impact on the subswarm-based search.

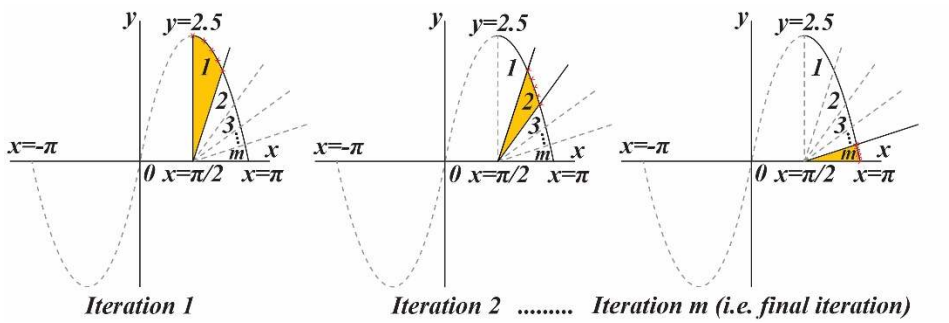

Figure 3. Dynamic coefficient generation using the sine function

Moreover, Equations (5)-(7) define the proposed helix function for dynamic coefficient generation in subswarm 3.

$$x = 0.0065 \times t \times \cos(2 \times t) \quad x \in [-2.522, 2.532] \tag{5}$$

$$y_3 = 0.0065 \times t \times \sin(2 \times t) \quad y_3 \in [0, 2.517] \tag{6}$$

$$z = t \tag{7}$$

where $t \in [0, 389.5575]$.

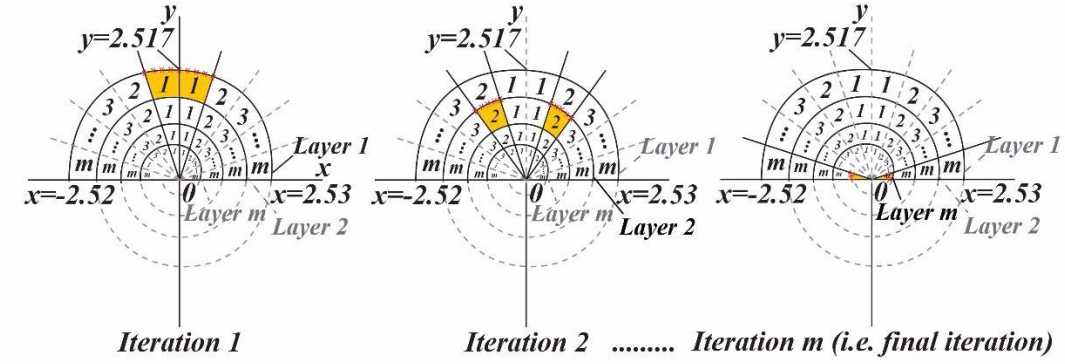

Figure 4. Dynamic coefficient generation using the helix function

Figure 4 shows the generated helix waveform in two dimensions, although in our implementation, we generate the helix waveform in three dimensions. The top half of the contours of the helix waveform is used for adaptive coefficient generation. The descending and ascending contours of c_1 and c_2 are produced as follows. We first generate *m* layers of the waveforms using the helix function defined in Equations (5)-(7). These *m* layers are dedicated to the generation of each coefficient in each iteration, as shown in Figure 4. We then divide both the upper right and upper left waveforms in all the layers into *m* portions, respectively. By starting from 90 degree, we label each portion from both sides successively as 1, 2, …, and *m* in each layer.

Instead of generating descending and ascending coefficients using a single waveform as illustrated in the sine and circle-based functions, we employ all the *m* layers to generate the dynamic adaptive coefficients for subswarm 3. In layer 1 (i.e. the largest outer layer),

we only use the two contours with portion index 1 from both right and left sides for the generation of c_1(see Iteration 1 in Figure 4). A random value is selected from the above two contours in layer 1, which is assigned as $c_1$in iteration 1. Then, we select a random value from the two contours with portion index 2 from both right and left sides in layer 2 (see Iteration 2 in Figure 4) and assign it to $c_1$in iteration 2. This process continues until a random value is selected from the two contours with portion index *m* from right and left sides in layer *m* (i.e. the smallest inner layer) for the last iteration (see Iteration *m* in Figure 4). The above process ensures the generation of a series of descending c_1 as indicated in Figure 4. A reverse process of the above operation ensures the generation of a series of ascending c_2, where the operation starts from portion index *m* in layer *m* and increases to finish at portion index 1 in layer 1.

When $c_1 > c_2$, the subswarm-based search leans more towards local exploitation, whereas when $c_2 > c_1$, the search process is more dominated by the global best experiences. The above process employs multiple waveforms with different radiuses for adaptive coefficient generation in comparison with the previous circle and sine-based processes where single waveforms are adopted. Overall, the above diverse dynamic strategies are used for the generation of adaptive acceleration coefficients to equip the proposed model with more flexibility in solving a wide range of optimization problems.

3.2.1.2 The sub-dimension based search

After obtaining a global best solution g_{best} using the above subswarm based search with adaptive acceleration coefficients, a sub-dimension based search is conducted to further enhance g_{best}. The pseudo-codes of the sub-dimension based search operation are provided in Algorithm 2.

Algorithm 2: Pseudo-Codes of Sub-dimension based Search	
1.	**Start**
2.	Divide each particle in the population into 3 subsections;
3.	// Conduct the search using Equations (1)-(2) in each subsection;
4.	**For** subsection 1 in each individual in the swarm **do**
5.	{Apply operations of Equations (1)-(2);}
6.	Obtain the best solution g_{sub1_best} for subsection 1 based search;
7.	**For** subsection 2 in each individual in the swarm **do**

8.	{Apply operations of Equations (1)-(2);}
9.	Obtain the best solution g_{sub2_best} for subsection 2 based search;
10.	**For** subsection 3 in each individual in the swarm **do**
11.	{Apply operations of Equations (1)-(2);}
12.	Obtain the best solution g_{sub3_best} for subsection 3 based search;
13.	Select the best leader, g'_{best}, among the above three best solutions g_{sub1_best}, g_{sub2_best}, and g_{sub3_best};
14.	Return g'_{best};
15.	**End**

As indicated in Algorithm 2, each particle is divided into three sub-dimensions, which are also denoted as three subsections. For the swarm particles, the sub-dimension based search mechanism implements the original PSO operation purely for each selected subsection, with the remaining subsections fixed. As an example, with respect to the feature selection task, for the extracted HOG features, a total of 324 dimensions are available to represent the texture information of a lesion image, which are further divided into three subsections, i.e. [1, 108], [109, 216], and [217, 324]. These three subsections are denoted as subsections 1, 2, and 3. For the sub-dimension based search for subsection 1, the PSO operation defined in Equations (1)-(2) is applied to sub-dimensions [1, 108] with the remaining two subsections comprising dimensions [109, 216] and [217, 324] fixed. For the sub-dimension based search for subsection 2, the search operation of the PSO model is deployed to sub-dimensions [109, 216], with two subsections comprising dimensions [1, 108] and [217, 324] intact. Similarly, for the sub-dimension based search with respect to subsection 3, sub-dimensions [217, 324] are updated using the PSO operation, with two subsections comprising dimensions [1, 108] and [109, 216] un-altered. In this way, the sub-dimension based search mechanism conducts local exploitation of each particle to implement fine-tuning.

The above three sub-dimension based search procedures return three best solutions, i.e. g_{sub1_best}, g_{sub2_best}, and g_{sub3_best}, pertaining to the corresponding feature subsections, respectively. We further identify the best leader g'_{best} from g_{sub1_best}, g_{sub2_best}, and g_{sub3_best}. This new best leader g'_{best} is used to substitute the previous global best solution, g_{best}, if it has a better fitness score. Otherwise, it replaces the worst particle in the swarm. The devised sub-dimension based search mechanism is able to overcome premature convergence to further enhance g_{best}. The performance analysis of this sub-

dimension based search operation is provided in Section 4.4.1.

3.2.2 The Second Proposed PSO Model

In the proposed strategies in Section 3.2.1.1, partial circle, sine, and helix waveforms are used for the generation of descending and ascending acceleration coefficients, i.e. the upper right quarters of the circle and sine contours and top half of the helix waveform. In this section, we propose another alternative PSO algorithm by using these three entire waveforms for the generation of both coefficients to increase search diversity. In addition, instead of generating descending and ascending coefficients, we produce random coefficients in this new PSO model by simulating hovering flight behaviours of hummingbirds around attraction. This alternative strategy for dynamic coefficient generation provides multiple positive and negative value choices in each iteration for each coefficient generation, therefore having more search diversity and more possibilities in attaining global optima. The sub-dimension based search as those in ACPSO is also subsequently conducted to further enhance g_{best}. The pseudo-codes of this alternative PSO model with random coefficients, denoted as RCPSO, are provided in Algorithm 3.

As indicated in Algorithm 3, dynamic random coefficients generated by the entire circle, sine and helix waveforms are used to guide the three subswarm-based searches, respectively. In comparison with Algorithm 1, both positive and negative random coefficients are used to enable exploration of a wider search space, therefore increasing the chances of finding global optima.

Algorithm 3: Pseudo-Codes of Proposed RCPSO
1. **Start**
2. Initialize a particle swarm randomly;
3. Evaluate each particle using the fitness/objective function $f(x)$;
4. Sort the swarm based on fitness values and identify g_{best};
5. **While** (Stopping criterion is not satisfied) {
6. Identify the second and third swarm leaders with comparable fitness scores but remote in positions to g_{best};
7. Randomly divide the swarm into three subswarms with each subswarm led by each leader;
8. **While** (!Stagnation)
9. { //use each strategy to lead each subswarm-based search

10.	**For** each particle in subswarm 1 **do**
11.	{ Generate j (e.g. 60) values using Eqn. (8);
12.	Randomly select 10 values from the generated 60 values to assign both coefficients, respectively;
13.	Generate 10 offspring particles using Eqn. (1)-(2) with the above randomly assigned coefficients;
14.	Select the best offspring among the 10 new solutions to replace the current particle x_i if it is fitter than x_i ;
15.	Update personal best of x_i and the subswarm leader;
16.	}
17.	Repeat lines 10-16 for subswarm 2 but generate j (e.g. 60) values using Eqn. (9);
18.	Repeat lines 10-16 for subswarm 3 but generate j (e.g. 60) values using Eqn. (10)-(12);
19.	Compare three subswarm leaders and store the best subswarm leader;
20.	**}Until** (stagnation detected)
21.	Combine the three subswarms and update g_{best};
22.	Conduct sub-dimension based search as shown in lines 17-23 in ACPSO to further enhance g_{best};
23.	}**Until (Stagnate several times and no improvement found)**
24.	Output g_{best};
25.	**End**

3.2.2.1 Dynamic random acceleration coefficients

As mentioned earlier, we employ both positive and negative random coefficients in RCPSO to increase exploration. Equation (8) illustrates the circle-based dynamic coefficient generation as the alternative velocity updating strategy for subswarm 1. Instead of using descending and ascending coefficients, both positive and negative random values from the circle contour are used for coefficient generation.

$$y_1' = \sqrt{radius'^2 - k'^2} \qquad k' \in [-2.5, 2.5] \tag{8}$$

In each iteration, Equation (8) is used to produce j random values for each particle. These j random values are generated in a way to have a reasonable coverage of the full circle contour. For instance, any newly generated random value needs to be different enough to any previously generated ones, otherwise it is abandoned. Based on trial and error, we employ j=60 randomly generated values, instead of other numbers (e.g. 100), as the pool for coefficient generation, since these values show sufficient variations and have a reasonable coverage of the entire circle waveform. Then, in each iteration, we randomly

select 10 out of 60 randomly generated values, and assign each of the 10 selected values to both coefficients, respectively. These 10 values, therefore, show sufficient differences to one another, and allow reasonable exploitation of the entire circle waveform to ensure the adoption of search parameters with appropriate diversity and variation. In this way, the search operation is accorded with enhanced capabilities in global exploration, as compared with using 10 randomly generated coefficients directly. The same method is also applied to random coefficient generation using the sine and helix waveforms.

As mentioned above, we assign each of the above selected 10 values among the 60 randomly generated ones to both coefficients, respectively. The fitness scores of all 10 newly generated solutions are subsequently calculated. The best offspring among these newly generated solutions is used to replace the current particle if it is fitter than the current particle. Otherwise, the current particle is reserved and passed on to the next generation.

Similarly, Equation (9) defines the alternative sine-based dynamic parameter generation for velocity updating in subswarm 2. Random positive and negative values are generated using the sine waveform for both coefficients.

$$y_2' = 2.5 * \sin(l') \quad l' \in [-\pi, \pi] \tag{9}$$

In each iteration, the sine function defined in Equation (9) is used to produce 60 random positive and negative values for each particle. Similar to the circle-based random coefficient generation, these 60 values show sufficient variations, and have a reasonable coverage of the entire sine waveform. The proposed PSO model then randomly selects 10 values from these 60 randomly generated ones to assign each coefficient respectively in each iteration. A set of 10 new solutions is therefore generated. The best offspring among these newly generated solutions is used to replace the current particle if it is fitter. Otherwise, we retain the current particle for the next generation.

Moreover, the helix-based dynamic coefficient generation is used in subswarm 3 for velocity updating, as defined in Equations (10)-(12). Random coefficients using the helix

waveform are generated, instead of descending and ascending ones.

$$x' = 0.0065 \times t' \times \cos(2 \times t') \quad x' \in [-2.522, 2.532] \tag{10}$$

$$y_3' = 0.0065 \times t' \times \sin(2 \times t') \quad y_3' \in [-2.527, 2.517] \tag{11}$$

$$z' = t' \tag{12}$$

where $t' \in [0, 389.5575]$. In each iteration, Equation (11) is used to produce 60 random positive and negative values for each particle. Again, these 60 random values show adequate variations, and have a sufficient coverage of the entire helix contour. In addition, 10 are randomly selected from these 60 values and assigned to both coefficients respectively in each iteration. Then, 10 new offspring solutions are generated. The best solution among these newly generated particles is used to replace the current particle if it is fitter.

The above acceleration coefficient generation mechanism is also motivated by the clonal mutation processes and certain characteristics of the clonal selection theory where the antibodies with the highest affinities are cloned proportionally to their antigenic affinities and the attributes of these clones are subsequently mutated [23-27].

Moreover, in comparison with ACPSO, RCPSO employs random coefficients selected from the entire circle/sine/helix waveforms to guide the search process. Both positive and negative coefficients are used to diversify the search by providing hovering behaviours around both personal and global best solutions (i.e. attraction). With wider value choices for assigning both coefficients in each iteration, this enhanced PSO model increases search diversity, and has more discriminative capabilities for feature selection in lesion classification.

Recommended by related studies [10, 11], the following fitness function is employed to evaluate each particle, p.

$$f(p) = w_{GM} \times GM_p + w_f \times (number\_feature_p)^{-1} \tag{13}$$

where w_{GM} and w_f denote the weights for Geometric Mean (GM) and the number of selected features, respectively, with $w_f = 1 - w_{GM}$. In this research, GM is used as the performance indicator, owing to its effectiveness in evaluating imbalanced data problems. In order to ensure a higher priority for classification performance over feature selection, a higher weight (e.g. 0.9) is assigned to w_{GM} in comparison with that (e.g. 0.1) of w_f.

As indicated in Equation (13), the proposed fitness function for feature selection is a maximization problem. This fitness function is selected since it minimizes the size of the selected feature subset while maximizing or maintaining the classification performance.

The decision variables (i.e. the particles) are evaluated as follows. The search starts with a swarm of randomly initialized particles. In other words, each particle is initialized with a random position in the search space with each element of the particle representing the position in each dimension. In order to achieve subtle movement and avoid premature convergence, we employ a continuous value for each element during the search process. For the fitness evaluation of each particle, we convert the continuous value in each dimension into a binary value, i.e. 1 or 0, to indicate the selection of a particular feature or otherwise. In this way, each particle is used to represent a selected feature subset. This selected feature subset is subsequently used for model training. The trained model is then used to conduct lesion classification using 10-fold cross-validation for fitness score generation. The final set of features recommended by the global best solution is regarded as the most optimal feature subset, which is used to assess the model performance using the test data set.

Moreover, in the fitness function shown in Equation (13), based on recommendations of related studies [10-12], we use a higher score (i.e. 0.9) as the weight associated with the classification performance as compared with that (i.e. 0.1) associated with the number of selected features, i.e., we assign a higher priority for classification performance over feature selection. This setting is in agreement with theoretical findings. When a higher weight is assigned to the classification performance, the search process is focused on increasing the classification capability of the model at the early stage. When the performance has reached a certain level of saturation, the search starts to focus on

removing the redundant features to further improve the model fitness. On the contrary, when a higher weight (e.g. 0.9) is given to the number of selected features with a lower weight (e.g. 0.1) assigned to the classification performance, the model tends to reduce the number of features at the beginning of the iteration process without considering the importance of the features in terms of classification capability. Therefore, the current weight setting (0.9:0.1) for the classification performance and the number of selected features achieves a better trade-off to ensure that classification capability is considered as the top priority before removing redundant features, in an attempt to achieve the best feature selection outcome.

3.3 Skin Lesion Classification

After identifying the most discriminative features using both proposed PSO algorithms, adaptive ensemble classifiers are used for benign and malignant lesion classifications. A deep learning model with adaptive hyper-parameter tuning is also proposed.

Table 1 The topology of the generated deep CNN model

Layers	Filter size	No. of filters	Stride	Padding
conv1	3	64	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
conv2	3	128	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
conv3	3	256	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
conv4	3	512	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
conv5	3	512	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
conv6	3	512	1	1
reluLayer	-	-	-	-
maxPooling	2	-	2	0
fc1	-	4096	-	-
reluLayer	-	-	-	-
fc2	-	3	-	-

We first introduce the ensemble models for lesion classification. The motivations of using

ensemble models are as follows. Ensemble classifiers are able to yield more reliable results in comparison with those of single classifiers [28]. The diversity of the base classifiers has a great impact on the performance of an ensemble model. Therefore, we employ three base classifiers of the same model, but trained with different lesion features to increase diversity. Two types of base classifiers, i.e. KNN and SVM, have been employed for ensemble model generation. In short, two ensemble models with KNN and SVM as the base classifiers, respectively, are constructed for prediction.

KNN and SVM are useful classification models that have been applied to a wide range of application domains [11, 12, 28]. KNN is a nonparametric classification method with an attractive property of computational simplicity, i.e., it does not require any parameter tuning process. On the other hand, SVM with a linear kernel has shown great robustness and capacities in solving diverse classification problems. The default parameter setting of the soft-margin constant, *Co*, has been employed for each SVM base classifier, in order to evaluate efficiency of the feature subsets identified by different search methods. The GM measure is adopted as the performance indicator, as it is frequently used for evaluation of imbalanced data sets. The GM measure is used for calculating both the fitness scores during the training process and the final classification results at the test stage, respectively.

In this research, three different types of features, i.e. ABCD+GLRLM, LBP and HOG characteristics, have been extracted for each lesion image. Subsequently, feature selection is performed on each of these raw feature vectors. Three optimized feature subsets are then used for the construction of three base models, respectively. Each base model is dedicated to each feature type. A weighted majority voting mechanism is used to combine the results from each base classifier, in order to obtain the final classification outcome for each test image. We present the detailed evaluation results in Section 4.

Secondly, a deep CNN model with 8 trainable layers (convolutional and fully connected layers) is employed for the classification of different lesion cases. Adaptive fine-tuning of the hyper-parameters of the above CNN model is conducted, such as the initial learning rate and L2Regularization (factor for L2 regularizer, i.e. weight decay), using the two proposed PSO models, owing to the importance of these hyper-parameters to the

classification performance of deep CNNs. The former parameter is used to set the initial learning rate of the deep network, while the latter is used in the loss function to reduce overfitting. Their settings could affect the network performance significantly. As an example, a small initial learning rate may result in a longer training time, while a large learning rate may lead to a suboptimal result. The regularization coefficient also plays an important role in influencing the network weight decay in the loss function. Both factors require the search of the corresponding strength space to find an optimal value [29-32]. Therefore, we integrate the two proposed PSO models with the CNN architecture to automatically identify the two optimized hyper-parameters. Classification performance using the GM measure is employed as the fitness evaluation of the optimal hyper-parameter selection in deep networks. The deep CNN model with the optimized hyper-parameters identified by the proposed PSO models is also compared against those of other PSO variants and the network with default parameter settings.

The details of the convolutional, pooling, and fully connected layers of the employed deep CNN model are provided in Table 1. The following training option is used for the deep CNN. The maximum number of epochs is 10, and the mini-batch size for each training iteration is 8. Moreover, the PH2 data set with three classes is employed for the evaluation of the proposed deep network. The network is trained from scratch for each set of the recommended hyper-parameters. The evaluation details are provided in Section 4.

4. EVALUATION

The proposed research is evaluated using two skin lesion data sets, i.e. Edinburgh Research and Innovation (Dermofit) [33] and the Dermatology Service of Hospital Pedro Hispano (PH2) [34] data sets. The Dermofit image library consists of a total of 1,300 skin lesion images with ten skin lesion types such as Actinic Keratosis and Malignant Melanoma. The PH2 database has a total of 200 lesion images with 80, 80, and 40 for common nevi, atypical nevi, and melanomas cases, respectively. The first experiment focuses on 2-class benign and malignant lesion classification, and employs a mixed data set with dermoscopic images of melanocytic lesions extracted from both Dermofit and PH2 databases. This mixed data set contains a total of 270 benign and 214 malignant images. Specifically, a set of 190 and 80 benign images is extracted from Dermofit and

PH2, respectively, while a set of 174 and 40 malignant images is extracted from Dermofit and PH2 data sets, respectively. The second experiment performs 3-class (i.e. benign, atypical and melanoma) lesion classification using the PH2 data set. For both the mixed and PH2 data sets, we use the 80:20 ratio for the training and test, respectively. KNN and SVM based ensemble classifiers and a deep CNN model are used for skin lesion classification.

Moreover, we implement 10 classical search methods and 11 state-of-the-art PSO variants for performance comparison, i.e. PSO, Bat Algorithm (BA), Cuckoo Search (CS), Dragonfly Algorithm (DA), Harmony Search (HS), Artificial Bee Colony (ABC), Flower Pollination Algorithm (FPA), Moth-Flame Optimization (MFO), bare-bones PSO (BBPSO), Cultural Algorithm (CA), ThBPSO [13], Genetic PSO (GPSO) [35], MPSO [4], GMPSO [10], F-BPSO [36], FS-BPSO [36], Enhanced Leader PSO (ELPSO) [37], PSO with multiple subpopulations (MFOPSO) [38], Autonomous Particles Groups for PSO (AGPSO) [39], Dynamic Neighbourhood Learning PSO (DNLPSO) [40], and BBPSOV [11].

4.1 Evaluation of the Mixed Data Set Using Ensemble Models

First of all, we present the evaluation results using two ensemble classification models. Note that the proposed PSO-based feature selection is a wrapper method since we evaluate the quality (i.e. fitness) of each feature subset recommended by each particle at the training stage by involving a classification model. In other words, the proposed feature selection process using ACPSO or RCPSO involves interaction with classifiers to capture feature dependencies. Other feature selection processes incorporated with other search methods in our study are also wrapper-based methods where feature selection is guided by the fitness evaluation in correspondence with the classification model [41].

The following experimental setting is employed for each method, i.e. image size=500×500, population=50, dimension=146 (ABCD+GLRLM)/236 (LBP)/324 (HOG), iterations=500, and runs=30. In order to ensure a fair comparison, the maximum number of fitness evaluations, i.e. population (50) × iterations (500), is used for each method. The experimental settings such as the population size (50) and the maximum number of

iterations (500) are determined by trial-and-error with the intention to achieve a reasonable trade-off between performance and computational cost. The experimental settings in our study are also in agreement with the recommended settings reported in other related studies for discriminative feature selection [10, 11, 12, 28].

Specifically, classical methods such as PSO employ a maximum iteration number of 500 as the stopping criterion whereas the proposed two PSO variants use fewer maximum iteration numbers (e.g. 125 for ACPSO and 40 for RCPSO) owing to additional fitness evaluations incurred in the proposed models for in-depth local lesion feature search. Moreover, two well-known filter methods are also used for comparison of feature selection, i.e. Minimum Redundancy and Maximum Relevance (mRMR) [42] and ReliefF [43]. Feature ranking is performed by both filter methods. Based on respective feature ranking, mRMR or ReliefF selects 100 ABCD+GLRLM, 150 LBP and 225 HOG features, respectively, for ensemble model generation. Each selected feature size is also similar to the average number of features obtained by all the wrapper-based methods for each feature type over 30 runs. A series of 30 runs has been conducted for mRMR, ReliefF and each search method. The detailed evaluation results for the ensemble models are provided in Table 2. We employ the mean GM performance over 30 runs as the main criterion for comparison.

Tables 2-4 indicate the empirical and statistical Wilcoxon rank sum test results [11, 37] of each method for the mixed data set. The proposed ACPSO and RCPSO models achieve the highest mean GM scores for both ensemble models in comparison with those of other methods. Especially, RCPSO achieves the best GM performances of 99.66% and 99.54% using the SVM ensemble model for 10-fold and hold-out validations, respectively. The statistical results indicate that both proposed models show statistically significant superiority over other methods for nearly all test cases. The exception is for ThBPSO which yields similar results to those of ACPSO in combination with the SVM-based ensemble classifier for 10-fold validation. Moreover, owing to the fact that mRMR and ReliefF mainly focus on feature ranking and do not take feature interaction with the classifier into account [41], they achieve lower mean GM scores in comparison with those of all wrapper-based methods.

Moreover, as shown in Figure 5, RCPSO selects the smallest feature subsets for each of the original feature vectors in comparison with those of other methods, while ACPSO shows a promising trade-off between computational efficiency and performance by selecting medium sizes of features. A similar observation is also obtained for the PH2 data set.

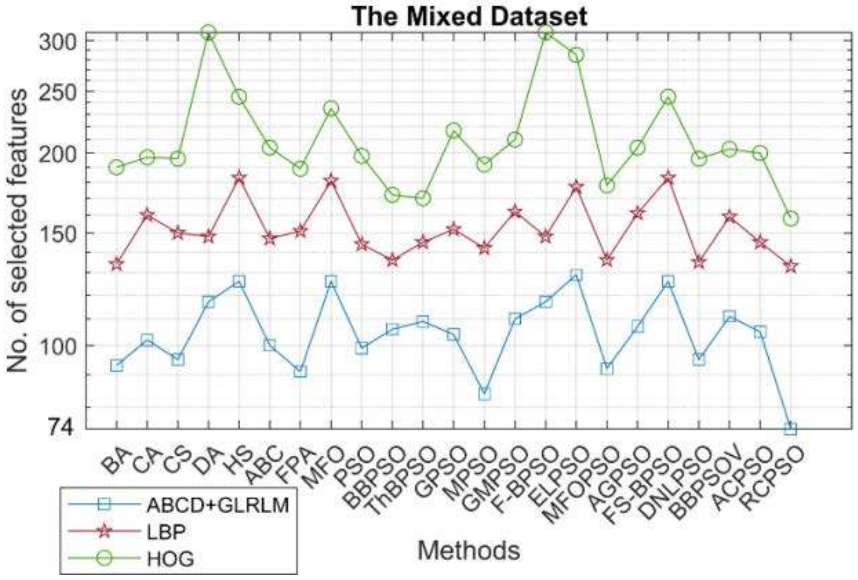

Figure 5. The number of selected features of each method for the mixed data set

Table 2 Mean classification results over 30 runs for each method for the mixed data set

	Mean	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN	**10-fold**	**0.9579**	0.9560	0.9336	0.9353	0.9362	0.9308	0.933	0.9404	0.9369	0.9381	0.9315	0.937
	Hold-out	**0.9559**	0.9541	0.9308	0.9323	0.9328	0.9276	0.9269	0.9332	0.9314	0.9367	0.9288	0.9319
ESEM SVM	**10-fold**	**0.9966**	0.9955	0.9845	0.9864	0.9825	0.9842	0.9872	0.9863	0.9869	0.989	0.9832	0.9866
	Hold-out	**0.9954**	0.9949	0.9831	0.9848	0.9809	0.9825	0.9853	0.9841	0.9821	0.9865	0.9805	0.9825

	Mean	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN	**10-fold**	0.9362	0.936	0.9322	0.9319	0.9439	0.94	0.9413	0.935	0.934	0.9352	0.9316
	Hold-out	0.9322	0.9301	0.9289	0.9283	0.9358	0.9341	0.9356	0.9311	0.9309	0.9279	0.9256
ESEM SVM	**10-fold**	0.9905	0.9882	0.9866	0.9871	0.9892	0.9893	0.9891	0.9829	0.986	0.983	0.9884
	Hold-out	0.9869	0.9861	0.9856	0.9833	0.9874	0.9846	0.9857	0.9796	0.9832	0.9825	0.985

		mRMR	ReliefF
ESEM KNN	**10-fold**	0.9082	0.9076
	Hold-out	0.9088	0.9154
ESEM SVM	**10-fold**	0.9212	0.9256
	Hold-out	0.9243	0.9287

Table 3 The Wilcoxon rank sum test results for the mixed data set for ACPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-Fold	1.20E-07	4.41E-07	7.97E-06	1.15E-07	1.10E-08	4.33E-05	3.06E-08	1.42E-05	1.65E-07	8.99E-08
ESEM KNN Hold-out	1.26E-08	2.12E-07	4.03E-07	2.38E-09	5.56E-10	7.50E-11	7.37E-10	4.80E-07	1.75E-09	1.58E-07
ESEM SVM 10-Fold	3.90E-05	1.36E-04	1.64E-05	1.48E-05	7.70E-04	2.42E-07	3.74E-04	3.45E-03	1.09E-05	9.79E-07
ESEM SVM Hold-out	9.25E-06	6.42E-06	2.96E-08	8.35E-08	7.64E-06	1.43E-10	1.33E-07	6.88E-06	2.21E-08	5.92E-09

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-Fold	2.00E-08	3.77E-07	8.25E-08	8.82E-10	1.23E-03	2.77E-07	2.19E-05	8.99E-08	8.25E-08	1.42E-08	1.61E-09
ESEM KNN Hold-out	3.39E-09	1.90E-08	4.38E-09	3.83E-10	3.91E-09	1.78E-07	1.17E-08	1.58E-07	4.38E-09	2.67E-10	8.65E-11
ESEM SVM 10-Fold	5.38E-02	3.53E-03	1.21E-05	1.85E-05	2.57E-03	2.31E-02	2.96E-05	9.79E-07	1.21E-05	4.78E-06	8.55E-04
ESEM SVM Hold-out	1.90E-04	3.13E-05	1.00E-07	1.00E-07	4.73E-04	8.48E-05	9.94E-09	5.92E-09	1.00E-07	2.00E-07	2.82E-06

	mRMR	ReliefF
ESEM KNN 10-Fold	7.22E-06	1.07E-09
ESEM KNN Hold-out	2.61E-10	1.43E-08
ESEM SVM 10-Fold	8.79E-07	2.77E-05
ESEM SVM Hold-out	1.49E-06	1.29E-09

Table 4 The Wilcoxon rank sum test results for the mixed data set for RCPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-Fold	6.24E-08	1.65E-07	7.94E-06	3.36E-07	5.73E-08	1.47E-05	3.63E-08	3.43E-06	2.77E-07	1.35E-07
ESEM KNN Hold-out	2.41E-08	4.44E-07	7.40E-07	3.11E-09	2.09E-09	6.72E-10	5.54E-09	1.10E-06	4.98E-09	3.36E-07
ESEM SVM 10-Fold	1.53E-05	5.44E-05	7.16E-06	8.03E-06	2.18E-04	3.00E-07	1.52E-04	8.68E-04	5.01E-06	6.98E-07
ESEM SVM Hold-out	9.33E-06	8.32E-06	7.67E-08	1.95E-07	9.04E-06	1.49E-09	3.32E-07	1.10E-05	6.40E-08	1.59E-08

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-Fold	1.06E-07	2.00E-07	1.46E-07	6.76E-09	4.97E-04	1.02E-06	1.86E-05	1.35E-07	1.46E-07	9.02E-08	1.48E-08
ESEM KNN Hold-out	1.55E-08	4.19E-08	7.67E-09	2.29E-09	5.05E-08	8.85E-07	6.98E-08	3.36E-07	7.67E-09	5.37E-10	2.62E-10
ESEM SVM 10-Fold	1.29E-02	7.46E-04	5.72E-06	1.36E-05	1.08E-04	1.18E-03	2.06E-05	6.98E-07	5.72E-06	2.28E-06	2.62E-04
ESEM SVM Hold-out	1.67E-04	3.23E-05	2.47E-07	2.47E-07	1.09E-04	3.59E-05	4.32E-08	1.59E-08	2.47E-07	3.64E-07	4.36E-06

	mRMR	ReliefF
ESEM KNN 10-Fold	6.73E-06	7.04E-07
ESEM KNN Hold-out	2.88E-08	2.31E-08
ESEM SVM 10-Fold	7.04E-07	1.41E-04
ESEM SVM Hold-out	3.50E-08	1.84E-04

4.2 Evaluation of the PH2 Data Set Using Ensemble Models

Evaluation is conducted for the 3-class lesion classification using the PH2 data set. The experimental setting of the above mixed data set is applied to this experiment, i.e. a total of population (50) × iterations (500) of function evaluations are used as the stopping criterion. Moreover, the two filter-based methods, i.e. mRMR and ReliefF, are also employed for comparison. Each method selects 100 ABCD+GLRLM, 150 LBP and 225 HOG features, respectively, for performance comparison. Again each selected feature size is similar to the average number of features obtained by all the wrapper-based methods for

each feature type over 30 runs. The empirical and the statistical test results are illustrated in Tables 5-7, respectively. RCPSO achieves the best mean GM performances of 97.79% and 97.54% in combination with the SVM-based ensemble for 10-fold and hold-out validations, respectively. The mean GM scores of both the proposed models show statistically significant improvements over those of other methods in nearly all test cases. The exception is for FS-BPSO, which achieves similar result distributions to those of ACPSO integrated with the KNN-based ensemble. For SVM-based ensemble model with 10-fold cross-validation, ACPSO also shows similar mean GM results to those of CS, MPSO, FS-BPSO, AGPSO and BBPSOV respectively. Moreover, FS-BPSO obtains similar performances to those of RCPSO for KNN and SVM-based ensembles with hold-out and 10-fold validations, respectively. Both proposed models also outperform the filter-based methods, i.e. mRMR and ReliefF, with statistical significance. The empirical results indicate that ABCD+GLRLM features contribute to the ensemble classification results significantly for the PH2 data set, with HOG & LBP features boosting the ensemble performance greatly for the mixed data set.

Table 5 Mean classification results over 30 runs for each method for the PH2 data set

	Mean	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN	**10-fold**	**0.9742**	0.9726	0.9397	0.9365	0.9407	0.936	0.9394	0.9363	0.9503	0.9385	0.9333	0.945
	Hold-out	0.9732	**0.9738**	0.9521	0.9488	0.9534	0.9567	0.9528	0.9573	0.9646	0.954	0.9489	0.955
ESEM SVM	**10-fold**	**0.9779**	0.9767	0.9618	0.9611	0.9465	0.9618	0.9662	0.9593	0.9613	0.9628	0.9503	0.9657
	Hold-out	**0.9754**	0.9731	0.9472	0.9447	0.9355	0.9429	0.952	0.9409	0.9453	0.9486	0.9359	0.9532

	Mean	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN	**10-fold**	0.9487	0.9456	0.9414	0.9377	0.939	0.953	0.9385	0.9367	0.9355	0.9471	0.9372
	Hold-out	0.9618	0.9596	0.9556	0.9527	0.9607	0.9677	0.9559	0.9481	0.9493	0.9555	0.9539
ESEM SVM	**10-fold**	0.9592	0.9634	0.9529	0.9577	0.9616	0.9646	0.9615	0.9581	0.9646	0.9604	0.9642
	Hold-out	0.9505	0.9446	0.9478	0.9412	0.9435	0.9481	0.9432	0.9528	0.9524	0.9468	0.95

		mRMR	ReliefF
ESEM KNN	**10-fold**	0.8908	0.9128
	Hold-out	0.9029	0.9274
ESEM SVM	**10-fold**	0.9108	0.9411
	Hold-out	0.9182	0.9487

Table 6 The Wilcoxon rank sum test results for the PH2 data set for ACPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-Fold	3.51E-07	8.95E-07	6.86E-08	4.37E-08	2.57E-07	4.85E-08	2.07E-05	2.17E-07	1.24E-07	5.89E-08
ESEM KNN Hold-out	3.90E-07	2.62E-09	2.29E-10	2.56E-05	3.80E-07	3.11E-07	8.05E-04	6.92E-06	2.82E-09	1.54E-08
ESEM SVM 10-Fold	4.69E-03	4.55E-03	1.46E-05	1.37E-02	9.58E-02	6.69E-04	5.28E-03	5.36E-03	4.96E-04	3.96E-04
ESEM SVM Hold-out	5.49E-06	2.67E-06	3.38E-08	1.03E-07	5.69E-05	2.05E-08	1.97E-07	7.19E-06	5.29E-09	3.26E-05

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-Fold	1.70E-05	8.75E-07	1.87E-07	7.25E-09	2.29E-05	5.04E-02	2.56E-07	5.89E-08	1.87E-07	7.67E-07	4.16E-08
ESEM KNN Hold-out	8.68E-04	2.91E-04	5.24E-11	1.39E-08	3.52E-03	6.69E-01	1.44E-09	1.54E-08	5.24E-11	1.60E-06	1.85E-09
ESEM SVM 10-Fold	1.15E-03	9.51E-03	8.91E-02	2.42E-03	1.38E-02	9.72E-02	1.86E-02	3.96E-04	8.91E-02	3.86E-04	7.43E-02
ESEM SVM Hold-out	1.93E-05	1.40E-07	2.36E-05	1.64E-07	1.17E-06	1.34E-05	1.36E-07	3.26E-05	2.36E-05	9.71E-07	1.84E-05

	mRMR	ReliefF
ESEM KNN 10-Fold	3.96E-08	1.12E-05
ESEM KNN Hold-out	7.80E-05	2.31E-08
ESEM SVM 10-Fold	1.58E-04	1.84E-04
ESEM SVM Hold-out	3.50E-08	2.31E-08

Table 7 The Wilcoxon rank sum test results for the PH2 data set for RCPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-Fold	1.49E-06	2.19E-05	4.15E-08	1.23E-08	2.83E-06	4.39E-08	4.24E-04	2.24E-07	1.28E-06	3.66E-08
ESEM KNN Hold-out	1.13E-06	1.42E-08	2.69E-09	5.40E-05	1.20E-06	1.36E-06	1.50E-03	1.62E-05	1.28E-08	6.69E-08
ESEM SVM 10-Fold	4.31E-04	5.74E-04	5.93E-06	1.14E-03	1.16E-02	1.10E-04	6.95E-04	6.59E-04	5.18E-05	4.81E-05
ESEM SVM Hold-out	5.47E-05	1.32E-05	1.27E-07	8.46E-07	7.49E-04	8.48E-08	2.76E-06	6.58E-05	1.95E-08	1.31E-04

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-Fold	1.51E-05	5.60E-06	8.32E-07	1.69E-09	5.31E-06	3.49E-02	9.62E-07	3.66E-08	8.32E-07	1.87E-06	3.86E-08
ESEM KNN Hold-out	1.29E-03	4.74E-04	4.24E-10	6.66E-08	2.77E-03	3.93E-01	1.44E-08	6.69E-08	4.24E-10	4.85E-06	1.49E-08
ESEM SVM 10-Fold	9.33E-05	1.09E-03	1.09E-02	1.54E-04	7.15E-03	5.69E-02	1.79E-03	4.81E-05	1.09E-02	6.01E-05	1.15E-02
ESEM SVM Hold-out	4.70E-04	9.04E-07	3.01E-04	1.84E-06	2.91E-07	1.39E-05	6.53E-07	1.31E-04	3.01E-04	5.61E-06	1.80E-04

	mRMR	ReliefF
ESEM KNN 10-Fold	6.15E-05	2.69E-05
ESEM KNN Hold-out	5.42E-05	2.69E-05
ESEM SVM 10-Fold	9.58E-04	5.69E-05
ESEM SVM Hold-out	1.58E-05	5.69E-05

4.3 Evaluation Using the Deep CNN Model

We also apply both proposed PSO models to adaptive fine-tuning of the learning rate and the regularization coefficient of CNNs to further ascertain its efficiency. The optimal settings of the two hyper-parameters usually rely on the training set and the network structure. The proposed PSO models are capable of identifying optimal network hyper-parameters for diverse network and training set-up. The PH2 data set is used in this experiment. The training, validation, and test ratios are 60:20:20. All the images are re-sized to 200×200. These resolutions are selected owing to the best trade-off between efficiency and performance. The network is trained from scratch for each set of the recommended hyper-parameters. We train and validate the CNN model using the training and validation sets, respectively. The optimal hyper-parameters identified in the training stage are used for test set evaluation.

A total of 30 successive runs are conducted for the training, validation, and test processes with the following experimental settings, i.e. population=15, dimension=2, and iterations=10. Except for ThBPSO, F-BPSO and FS-BPSO which are dedicated to feature selection problems, we employ all other methods for hyper-parameter fine-tuning in deep CNNs. The same number of function evaluations, i.e. population (15) × iterations (10), is used for each method to ensure a fair comparison. As indicated in Tables 8-10, both proposed PSO models show statistically significant improvements over nearly all other methods, except for FPA, MFO and BBPSOV which achieve similar results to those of the proposed models. In addition, the proposed models outperform the network with the default setting provided by MATLAB, i.e. the learning rate = 0.01 and L2Regularization = 0.0001, statistically.

Table 8 Mean classification results over 30 runs for CNN hyper-parameter fine-tuning using the PH2 data set

	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Mean	**0.8945**	0.8926	0.8214	0.8458	0.8327	0.8589	0.8240	0.7848	0.8371	0.8153	0.8491	0.8197

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	Default
Mean	0.8225	0.8534	0.8501	0.8033	0.8295	0.8454	0.7968	0.8567	0.8480

Table 9 The Wilcoxon rank sum test results for CNN hyper-parameter fine-tuning for ACPSO using the PH2 data set

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Wilcoxon	4.08E-04	2.17E-02	5.57E-03	6.42E-03	2.44E-02	4.91E-06	1.15E-03	2.29E-04	1.49E-01	1.31E-01

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	Default
Wilcoxon	1.99E-03	5.91E-03	5.44E-04	1.13E-04	5.58E-04	8.82E-03	5.60E-07	1.01E-01	2.62E-04

Table 10 The Wilcoxon rank sum test results for CNN hyper-parameter fine-tuning for RCPSO using the PH2 data set

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Wilcoxon	5.98E-04	2.34E-02	5.97E-03	7.93E-03	1.87E-02	8.02E-06	1.83E-03	4.40E-04	1.43E-01	1.07E-01

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	Default
Wilcoxon	2.09E-03	5.23E-03	8.19E-04	1.62E-04	9.52E-04	1.05E-02	6.36E-07	1.08E-01	4.60E-04

Table 11 shows comparison with existing studies using the PH2 data set. Since different training and test sets and evaluation strategies have been used for the evaluation of each work, it serves as a rough performance indication. Our proposed SVM-based ensemble models combined with RCPSO-based feature selection are among the top performers, serving as superior alternatives for lesion classification. Although the deep CNN model employs a lower image resolution (200×200), a small experimental set-up and a lower training ratio, its performance is comparable to those of related studies using higher image resolutions [2, 44, 45, 47, 51].

Table 11 Comparison with related work using the PH2 data set

Studies	Methodology	Types	Strategy	Results
Adjed et al. [2]	Wavelet and Curvelet Transforms+LBP+SVM	3	random sampling	0.8607
Barata et al. [44]	Colour Constancy + Bag-of-Features + K-means + SVM	3	10-fold	0.8430
Alfed et al. [45]	Colour Histograms + Colour Moments + HOG + Codebook Generation + SVM/AdaBoost	3	5-fold	0.8800
Waheed et al. [46]	Colour (HSV)+ Gray-Level Co-occurrence Matrix + SVM	3	3-fold	0.9600
Marques et al. [47]	Texture + Colour (RGB, HSV, L*a*b)	3	N/A	0.7910
Bi et al. [48]	Multistage Fully Convolutional Networks	3	N/A	0.9066
Eltayef et al. [49]	Properties of Pigment Network + Neural Networks	3	N/A	0.9000
Pennisi et al. [50]	Artefact Removal + Skin Detection + Lesion Segmentation & Binary Mask + AdaBoost	3	N/A	0.9360
Barata et al. [16]	Global Methods	3	N/A	0.96 (SE) 0.80 (SP)
Barata et al. [16]	Local Features + The Bag-of-Features Classifier	3	N/A	1.00 (SE) 0.75 (SP)
Barata et al. [51]	Lesion or Pigment Ratio + Boosting Algorithm	3	10-fold	0.8620
This research	ABCD+GLRLM/LBP/HOG + RCPSO-based Feature Selection + SVM-based Ensemble Classifier	3	10-fold Hold-out	**0.9779** **0.9754**

4.4 Evaluation Using UCI and Other Image Data Sets

To further evaluate the model efficiency, we employ a blood cancer microscopic image data set, i.e. ALL-IDB2 [52-53] and two UCI data sets [54], i.e. breast cancer and epileptic seizure, for evaluation. Specifically, the two UCI data sets are used to assess the model

efficiency for discriminative feature selection while the ALL-IDB2 microscopic image data set is used to evaluate the model capability in optimal hyper-parameter identification of deep networks.

4.4.1 Evaluation Using the UCI Data Sets for Feature Selection

We first employ the two UCI data sets, i.e. breast cancer and epileptic seizure [54], for feature selection. Specifically, the breast cancer data set comprises 569 samples, 32 features, and 2 target classes, i.e. positive and negative. The epileptic seizure data set includes a total of 11,500 samples with 179 features. It contains 5 target classes, namely eyes open, eyes closed, recording the EEG activity from the healthy brain area, recording the EEG activity from the area where the tumor is located, and recording seizure activity. Owing to the large sample size in this seizure data set, we randomly select a subset of 6000 instances for evaluation. An aspect ratio of 80:20 is used as the training and test split for both UCI data sets for discriminative feature selection.

The experimental setting of the aforementioned ensemble lesion classification is also applied to this experiment, i.e. a total of population (50) × iterations (500) of function evaluations are used as the stopping criterion. In other words, all the methods terminate when the maximum number of function evaluations is reached, in order to ensure a fair comparison. A set of 30 runs has been conducted by each method for each data set. The two filter-based methods, mRMR and ReliefF, are also employed for performance comparison. We employ 22 and 60 numbers of features for both filter-based methods for the breast cancer and epileptic seizure data sets, respectively, which are similar to the average numbers of selected features by the wrapper-based methods over 30 runs for both data sets.

Tables 12-14 depict the empirical and Wilcoxon rank sum test results for the breast cancer data set. The proposed RCPSO and ACPSO models outperform, with statistical significance results, all 21 baseline wrapper-based methods over 30 runs in discriminative feature selection. They also outperform, again with statistical significance results, the two filter-based methods, i.e. mRMR and ReliefF. RCPSO yields the best GM scores, i.e. 98.70% and 98.73%, in combination with the SVM-based ensemble classifier for 10-fold

and hold-out validations, respectively.

Table 12 Mean classification results over 30 runs for each method for the breast cancer data set

	Mean	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN	**10-fold**	**0.9678**	0.9676	0.9600	0.9582	0.9583	0.9588	0.9583	0.9598	0.9592	0.9597	0.9606	0.9582
	Hold-out	**0.9691**	0.9688	0.9647	0.9622	0.9620	0.9629	0.9623	0.9634	0.9635	0.9633	0.9642	0.9622
ESEM SVM	**10-fold**	**0.9870**	0.9864	0.9795	0.9795	0.9769	0.9767	0.9784	0.9794	0.9841	0.9785	0.9785	0.9795
	Hold-out	**0.9873**	0.9856	0.9816	0.9816	0.9793	0.9795	0.9808	0.9818	0.9805	0.9808	0.9805	0.9816

	Mean	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN	**10-fold**	0.9603	0.9592	0.9606	0.9592	0.9620	0.9565	0.9601	0.9584	0.9605	0.9440	0.9591
	Hold-out	0.9635	0.9630	0.9639	0.9630	0.9615	0.9635	0.9634	0.9628	0.9610	0.9484	0.9628
ESEM SVM	**10-fold**	0.9810	0.9796	0.9776	0.9811	0.9789	0.9766	0.9778	0.9802	0.9775	0.9760	0.9784
	Hold-out	0.9833	0.9818	0.9798	0.9831	0.9815	0.9825	0.9800	0.9823	0.9785	0.9780	0.9805

	Mean	mRMR	ReliefF
ESEM KNN	**10-fold**	0.9598	0.9603
	Hold-out	0.9610	0.9623
ESEM SVM	**10-fold**	0.9784	0.9803
	Hold-out	0.9793	0.9812

Table 13 The Wilcoxon rank sum test results for the breast cancer data set for ACPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-Fold	4.75E-07	6.70E-07	4.90E-07	4.50E-07	1.45E-07	9.60E-07	7.16E-07	7.97E-07	2.46E-06	5.30E-07
ESEM KNN Hold-out	3.27E-04	6.32E-05	9.65E-06	5.86E-04	5.09E-06	3.85E-04	1.24E-05	1.06E-05	4.80E-03	6.29E-05
ESEM SVM 10-fold	3.58E-07	4.45E-08	2.73E-08	6.06E-07	2.64E-07	1.39E-06	5.55E-05	1.38E-08	4.34E-06	5.83E-07
ESEM SVM Hold-out	4.73E-06	4.02E-06	9.72E-07	3.65E-07	4.76E-07	1.36E-04	2.09E-02	4.79E-07	4.61E-07	3.96E-06

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-fold	3.00E-07	2.76E-07	1.61E-04	2.97E-07	2.47E-07	6.88E-07	5.24E-07	9.49E-07	2.85E-06	1.07E-07	1.72E-07
ESEM KNN Hold-out	1.27E-05	8.28E-06	3.18E-02	8.96E-06	6.99E-06	2.29E-05	1.89E-03	6.91E-06	7.71E-06	2.84E-07	8.83E-06
ESEM SVM 10-fold	3.20E-07	4.52E-08	6.47E-07	4.40E-07	2.48E-07	9.75E-07	6.14E-07	9.02E-07	6.18E-07	9.36E-07	9.18E-07
ESEM SVM Hold-out	1.31E-03	5.88E-05	9.87E-07	6.19E-04	3.88E-07	1.51E-03	3.51E-08	5.62E-05	2.60E-07	1.63E-07	5.53E-07

	mRMR	ReliefF
ESEM KNN 10-Fold	3.00E-07	1.45E-07
ESEM KNN Hold-out	1.06E-05	6.32E-05
ESEM SVM 10-fold	6.06E-07	4.40E-07
ESEM SVM Hold-out	5.88E-05	1.24E-05

Table 14 The Wilcoxon rank sum test results for the breast cancer data set for RCPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-fold	4.11E-07	4.41E-08	1.60E-09	1.60E-07	4.26E-09	2.26E-08	7.34E-11	2.63E-10	2.09E-06	4.41E-08
ESEM KNN Hold-out	3.27E-03	6.26E-05	8.84E-06	5.85E-04	4.91E-06	3.85E-04	1.22E-05	1.00E-05	4.80E-03	6.26E-05
ESEM SVM 10-fold	6.87E-08	1.44E-08	4.22E-10	2.79E-11	4.20E-08	4.94E-07	5.49E-05	3.61E-09	3.36E-06	1.44E-08
ESEM SVM Hold-out	3.80E-06	3.80E-06	7.25E-09	2.37E-08	1.79E-07	1.35E-04	2.09E-02	1.79E-07	1.98E-07	3.80E-06

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-fold	8.43E-10	9.97E-11	1.61E-04	4.92E-10	8.33E-09	5.41E-09	1.12E-07	1.01E-10	2.29E-06	2.90E-11	6.91E-11
ESEM KNN Hold-out	1.22E-05	8.08E-06	3.18E-02	8.08E-06	6.84E-06	2.22E-05	1.89E-03	6.36E-06	7.64E-06	1.23E-11	8.08E-06
ESEM SVM 10-fold	1.90E-07	2.86E-08	1.97E-09	7.59E-08	1.55E-08	4.94E-07	1.42E-08	2.85E-08	1.67E-09	3.33E-10	1.60E-07
ESEM SVM Hold-out	1.31E-03	5.80E-05	2.35E-07	6.18E-04	3.80E-07	1.51E-03	5.36E-09	5.59E-05	6.45E-08	3.34E-11	5.88E-08

	mRMR	ReliefF
ESEM KNN 10-Fold	5.41E-09	1.01E-10
ESEM KNN Hold-out	8.84E-06	8.08E-06
ESEM SVM 10-fold	4.94E-07	7.59E-08
ESEM SVM Hold-out	2.35E-07	1.79E-07

For the epileptic seizure data set, Tables 15-17 present the experimental and statistical test results. RCPSO reveals the best GM performance, i.e. 95.78% and 95.92%, in combination with the SVM-based ensemble classifier for 10-fold and hold-out validations, respectively, over 30 runs. Both ACPSO and RCPSO models show statistically significant superiority over mRMR, ReliefF, and 21 search methods, except that RCPSO shows similar result distributions to those of AGPSO for the SVM-based ensemble model with 10-fold validation over 30 runs.

Table 15 Mean classification results over 30 runs for each method for the epileptic seizure data set

	Mean	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN	**10-fold**	0.8259	**0.8262**	0.8061	0.8041	0.8014	0.8077	0.8033	0.7996	0.8026	0.8013	0.8030	0.8041
	Hold-out	**0.8298**	0.8277	0.8114	0.8095	0.8074	0.8131	0.8092	0.8052	0.8085	0.8069	0.8085	0.8095
ESEM SVM	**10-fold**	**0.9578**	0.9573	0.9380	0.9364	0.9394	0.9385	0.9401	0.9370	0.9374	0.9376	0.9376	0.9364
	Hold-out	**0.9592**	0.9589	0.9395	0.9378	0.9408	0.9398	0.9413	0.9383	0.9388	0.9388	0.9389	0.9378

	Mean	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN	**10-fold**	0.8024	0.8030	0.8003	0.7997	0.8090	0.7961	0.8046	0.7983	0.8105	0.8033	0.8015
	Hold-out	0.8082	0.8085	0.8058	0.8054	0.8130	0.8025	0.8104	0.8040	0.8135	0.8089	0.8068
ESEM SVM	**10-fold**	0.9382	0.9401	0.9353	0.9357	0.9410	0.9352	0.9375	0.9375	0.9425	0.9375	0.9399
	Hold-out	0.9395	0.9414	0.9367	0.9370	0.9425	0.9365	0.9388	0.9389	0.9433	0.9388	0.9411

	Mean	mRMR	ReliefF
ESEM KNN	**10-fold**	0.8192	0.8119
	Hold-out	0.8199	0.8137
ESEM SVM	**10-fold**	0.9364	0.9388
	Hold-out	0.9387	0.9393

Table 16 The Wilcoxon rank sum test results for the epileptic seizure data set for ACPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-fold	3.74E-05	1.00E-04	3.50E-06	3.63E-04	2.88E-06	6.24E-07	7.60E-06	6.02E-06	6.58E-06	1.03E-04
ESEM KNN Hold-out	6.08E-05	3.34E-04	3.81E-06	7.54E-04	8.30E-06	2.28E-06	2.53E-06	2.74E-06	3.68E-06	3.41E-04
ESEM SVM 10-fold	1.23E-05	2.97E-06	3.10E-04	1.22E-06	2.18E-04	1.95E-06	6.36E-06	8.56E-06	4.95E-07	7.34E-06
ESEM SVM Hold-out	5.18E-05	3.63E-06	8.77E-04	9.52E-06	5.93E-04	4.96E-06	8.81E-06	7.55E-06	5.14E-07	5.55E-06

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-fold	8.04E-06	9.47E-06	9.92E-06	1.98E-06	2.28E-05	7.48E-06	8.09E-06	2.99E-06	2.85E-04	6.56E-05	1.52E-05
ESEM KNN Hold-out	2.32E-05	2.39E-06	3.54E-06	8.10E-06	3.03E-04	6.16E-06	1.34E-05	1.39E-06	5.59E-05	9.16E-05	1.21E-05
ESEM SVM 10-fold	1.22E-06	4.12E-06	7.59E-06	3.62E-06	2.13E-02	2.03E-06	4.54E-06	4.30E-06	3.98E-02	6.64E-06	6.34E-07
ESEM SVM Hold-out	4.62E-06	2.62E-06	9.74E-06	9.04E-06	4.10E-02	8.35E-06	5.50E-06	2.49E-06	2.13E-02	2.94E-06	7.97E-06

	mRMR	ReliefF
ESEM KNN 10-fold	9.92E-06	2.18E-04
ESEM KNN Hold-out	2.62E-06	6.08E-05
ESEM SVM 10-fold	3.50E-06	4.96E-06
ESEM SVM Hold-out	2.32E-05	2.53E-06

Table 17 The Wilcoxon rank sum test results for the epileptic seizure data set for RCPSO

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
ESEM KNN 10-fold	2.77E-05	9.79E-05	2.38E-07	3.56E-04	2.68E-06	5.46E-09	1.73E-07	6.05E-07	7.04E-07	9.79E-05
ESEM KNN Hold-out	5.38E-05	3.33E-04	4.16E-07	7.48E-04	5.53E-06	1.39E-08	3.33E-07	6.88E-07	9.05E-07	3.33E-04
ESEM SVM 10-fold	6.74E-06	4.44E-07	3.01E-04	3.08E-08	2.13E-04	6.01E-08	1.55E-09	1.41E-09	2.03E-07	4.44E-07
ESEM SVM Hold-out	4.66E-05	2.65E-06	8.69E-04	2.04E-07	5.88E-04	7.76E-08	8.68E-10	6.39E-09	3.10E-07	2.65E-06

	ThBPSO	GPSO	MPSO	GMPSO	F-BPSO	FS-BPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV
ESEM KNN 10-fold	3.57E-06	4.31E-08	2.03E-07	9.06E-08	1.73E-05	2.15E-10	1.61E-06	3.35E-08	2.77E-04	5.61E-05	6.74E-06
ESEM KNN Hold-out	1.95E-05	6.85E-07	1.29E-06	2.20E-06	3.01E-04	9.30E-10	3.75E-06	1.85E-07	5.38E-05	8.17E-05	1.01E-05
ESEM SVM 10-fold	5.09E-08	1.36E-07	1.10E-08	1.07E-09	2.13E-02	3.02E-11	1.61E-06	2.03E-07	**3.98E-01**	1.70E-08	2.57E-07
ESEM SVM Hold-out	4.90E-07	3.07E-07	5.01E-08	3.05E-09	4.10E-02	4.66E-11	3.84E-06	2.64E-07	2.13E-02	2.71E-08	6.76E-07

	mRMR	ReliefF
ESEM KNN 10-fold	4.96E-06	2.97E-06
ESEM KNN Hold-out	3.57E-06	1.36E-07
ESEM SVM 10-fold	1.29E-06	1.23E-05
ESEM SVM Hold-out	3.57E-06	2.20E-06

Figure 6 illustrates the average numbers of selected features of each method for both breast cancer and seizure data sets over 30 runs. For the breast cancer data set, ACPSO, MPSO, and RCPSO select the smallest optimal feature subsets in comparison with those of other methods. For the seizure data set, MPSO, ThBPSO, GPSO and RCPSO obtain the smallest optimal feature subsets with the rest of the methods identifying comparatively

larger feature subsets. In short, both proposed models achieve the best trade-off between classification performance and computational efficiency. They obtain smaller or comparable sizes of feature subsets in comparison with those of all the baseline methods for both UCI data sets, while maintaining competitive classification performance.

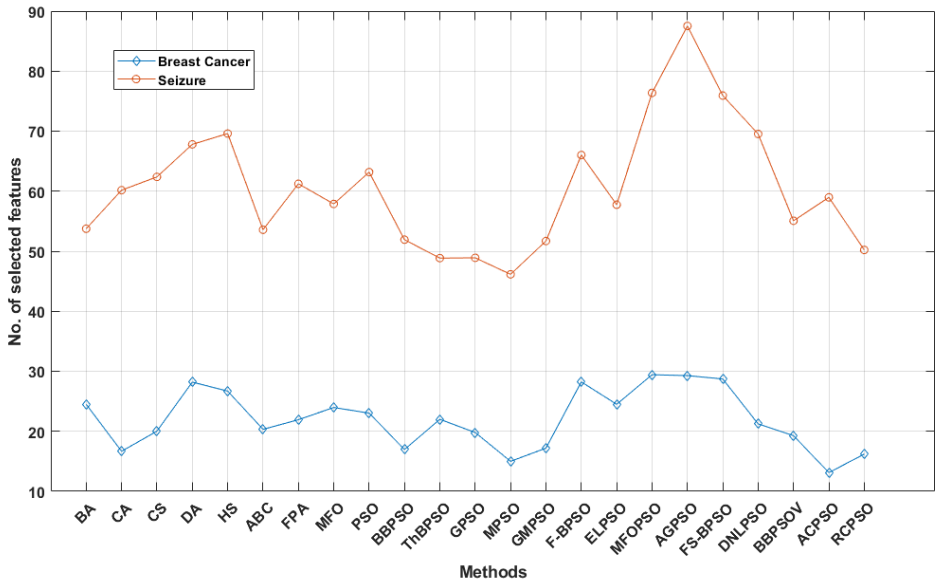

Figure 6 The numbers of selected features of each method for the breast cancer and seizure data sets

We subsequently conduct experiments to investigate the performance contribution of the proposed sub-dimension based search mechanism in the proposed ACPSO and RCPSO models. We compare the full versions of the proposed ACPSO and RCPSO models and both models without the sub-dimension based search operations but with all other proposed mechanisms embedded. All the four data sets are used for comparison, i.e. the mixed and PH2 skin lesion data sets and two UCI data sets. A series of 30 runs is conducted for each algorithm with each data set. Table 18 depicts the detailed evaluation results. As indicated in Table 18, the proposed sub-dimension based search strategy performs fine-tuning and local exploitation of the swarm particles to further enhance g_{best}. Therefore, it shows sufficient capabilities in improving the feature selection outcomes consistently for all the evaluated scenarios using all the data sets. Specifically, for the PH2 data set, equipped with the sub-dimension based search mechanism, the proposed RCPSO model yields 1.27% and 1.09% improvements in comparison with the

version without such operations for the KNN-based ensembles with 10-fold and hold-out validations, respectively. Similarly, equipped with sub-dimension based local exploitation, ACPSO also yields 1.18% and 1.23% improvements in comparison with the version without such operations for the above test cases with the PH2 data set. Overall, ACPSO shows slightly more improvements than RCPSO with the inclusion of the sub-dimension strategy for g_{best} enhancement as compared with the versions without such procedures. In short, the proposed sub-dimension based search operation is able to produce enhanced capabilities in improving the feature selection results for all the test cases.

Table 18 The mean evaluation results of the sub-dimension based search operation over 30 runs

		Mean	RCPSO (1)	RCPSO without sub_dimen (2)	Increase b/w (1) and (2)	ACPSO (3)	ACPSO without sub_dimen (4)	Increase b/w (3) and (4)
Mixed	**KNN**	**10-fold**	0.9579	0.9489	0.0090	0.9560	0.9478	0.0082
		Hold-out	0.9559	0.9497	0.0062	0.9541	0.9458	0.0083
	SVM	**10-fold**	0.9966	0.9933	0.0033	0.9955	0.9906	0.0049
		Hold-out	0.9954	0.9935	0.0019	0.9949	0.9912	0.0037
PH2	**KNN**	**10-fold**	0.9742	0.9615	0.0127	0.9726	0.9608	0.0118
		Hold-out	0.9732	0.9623	0.0109	0.9738	0.9615	0.0123
	SVM	**10-fold**	0.9779	0.9701	0.0078	0.9767	0.9698	0.0069
		Hold-out	0.9754	0.9722	0.0032	0.9731	0.9699	0.0032
Breast	**KNN**	**10-fold**	0.9678	0.9666	0.0012	0.9676	0.9664	0.0012
		Hold-out	0.9691	0.9686	0.0005	0.9688	0.9670	0.0018
	SVM	**10-fold**	0.9870	0.9858	0.0012	0.9864	0.9857	0.0007
		Hold-out	0.9873	0.9865	0.0008	0.9856	0.9843	0.0013
Seizure	**KNN**	**10-fold**	0.8259	0.825	0.0009	0.8262	0.8235	0.0027
		Hold-out	0.8298	0.8279	0.0019	0.8277	0.8263	0.0014
	SVM	**10-fold**	0.9578	0.9507	0.0071	0.9573	0.9498	0.0075
		Hold-out	0.9592	0.9577	0.0015	0.9589	0.9556	0.0033
Average					0.0044			0.0050

4.4.2 Evaluation Using the ALL-IDB2 Data Set for Hyper-parameter Fine-tuning

In order to further assess robustness of the proposed models, the ALL-IDB2 microscopic image data set [52] is employed for optimal hyper-parameter identification of deep networks. This data set has a total of 180 microscopic white blood cell sub-images with 120 positive (i.e. acute lymphoblastic leukaemia) and 60 negative samples [52, 53]. An aspect ratio of 60-20-20 is employed for training, validation, and test evaluations. All the images are re-sized to 200×200. The same settings as those of CNN hyper-parameter fine-tuning for the PH2 data set are adopted in this experiment, i.e. the maximum number of

fitness evaluations=population (15) × iterations (10), and runs=30. The CNN network is trained from scratch for each set of the recommended hyper-parameters. Tables 19-21 show the empirical and Wilcoxon rank sum test results over 30 runs. The proposed RCPSO and ACPSO models outperform, with statistical significance results, all the baseline methods in nearly all test cases in hyper-parameter fine-tuning. The exception is for BBPSOV, which shows similar results to those of ACPSO.

Table 19 Mean classification results over 30 runs for CNN hyper-parameter fine-tuning using the ALL-IDB2 data set

	RCPSO	ACPSO	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Mean	0.9113	0.8799	0.7679	0.7112	0.7520	0.7569	0.7962	0.7616	0.7616	0.7836	0.7537	0.7883

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	default
Mean	0.7395	0.8119	0.7710	0.7301	0.7128	0.6892	0.7474	0.8151	0.7477

Table 20 The Wilcoxon rank sum test results for CNN hyper-parameter fine-tuning for ACPSO using the ALL-IDB2 data set

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Wilcoxon	3.65E-05	1.27E-05	1.92E-04	4.27E-04	3.43E-03	3.03E-03	1.62E-03	1.20E-02	2.57E-03	1.07E-02

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	default
Wilcoxon	2.52E-05	3.15E-02	1.02E-02	8.36E-05	4.07E-04	4.29E-06	5.21E-04	7.88E-02	8.69E-03

Table 21 The Wilcoxon rank sum test results for CNN hyper-parameter fine-tuning for RCPSO using the ALL-IDB2 data set

	PSO	BBPSO	BA	CA	CS	DA	HS	ABC	FPA	MFO
Wilcoxon	7.89E-07	1.67E-07	6.53E-06	5.24E-05	2.00E-05	1.65E-04	1.18E-04	5.61E-04	1.27E-04	3.52E-05

	GPSO	MPSO	GMPSO	ELPSO	MFOPSO	AGPSO	DNLPSO	BBPSOV	default
Wilcoxon	2.29E-07	9.67E-04	8.26E-04	3.82E-06	5.37E-06	1.46E-07	6.86E-06	1.50E-03	3.50E-05

4.5 Computational Cost

The computational costs for the proposed models and all the baseline methods for

discriminative feature selection and optimal hyper-parameter identification of deep networks are summarized in Tables 22-23.

Since all the methods use the same maximum number of function evaluations in each optimization task, i.e. feature selection and optimal hyper-parameter selection, at the training stage and since fitness evaluation is the most time-consuming component, all the methods have the same computational cost primarily for each optimization task. Nevertheless, owing to the differences of the internal search mechanisms in each search method, the computational cost of each search method varies slightly.

The average training computational costs for discriminative feature selection pertaining to the skin lesion, breast cancer and seizure data sets over 30 runs are provided in Table 22. As illustrated in Table 22, the costs of the three computationally most efficient models for each data set are highlighted in bold. For the mixed skin lesion data set, FPA, ACPSO and BA yield the least computational costs, while RCPSO, PSO, CS, DA, GMPSO, MFOPSO, BBPSOV and MPSO produce medium computational costs, with the rest of the models depicting computationally expensive costs. For the PH2 data set, FPA, ACPSO and DA are computationally the most efficient methods. This is followed by BA, RCPSO, CS, and GMPSO, with the remaining methods showing comparatively higher computational costs.

Table 22 The average training computational costs (in seconds) for discriminative feature selection over a series of 30 runs

cost	**RCPSO**	**ACPSO**	**PSO**	**BBPSO**	**BA**	**CA**	**CS**	**DA**	**HS**	**ABC**	**FPA**	**MFO**
Mixed	4431.38	**4412.68**	4482.48	4544.43	**4416.49**	4524.41	4467.68	4416.59	4555.87	4508.26	**4406.30**	4548.06
PH2	3821.65	**3808.58**	3880.73	3935.38	3815.23	3923.22	3863.04	**3809.33**	3948.80	3906.03	**3800.89**	3946.44
Breast	**1912.39**	1931.31	1982.12	2043.81	1916.40	2024.15	1966.81	**1916.17**	2055.60	2007.34	**1905.31**	2047.11
Seizure	**6722.81**	**6766.82**	6946.64	7005.57	6891.09	6934.96	**6762.72**	6811.01	7363.88	6947.75	6858.62	6982.98

cost	**ThBPSO**	**GPSO**	**MPSO**	**GMPSO**	**F-BPSO**	**FS-BPSO**	**ELPSO**	**MFOPSO**	**AGPSO**	**DNLPSO**	**BBPSOV**
Mixed	5673.82	4550.73	4493.38	4482.70	4600.50	4567.03	4591.98	4486.07	4583.58	4600.36	4486.99
PH2	4562.87	3941.51	3888.44	3876.72	3993.51	3964.36	3991.91	3882.95	3979.46	3997.77	3879.59
Breast	3872.92	2050.52	1992.57	1982.50	2099.55	2066.63	2091.52	1985.59	2082.64	2100.09	1986.58
Seizure	8823.82	7026.52	6983.78	6957.73	7029.82	7037.60	7776.87	7287.87	6852.72	6783.82	6899.99

For the breast cancer data set, FPA, RCPSO and DA yield the lowest training computational costs for discriminative feature selection. This is closely followed by BA, ACPSO and CS. For the seizure data set, RCPSO, CS and ACPSO are computationally the most optimal models, with the rest of the methods showing more expensive computational costs. In short, the proposed ACPSO and RCPSO models have better or comparable training computational costs as compared with those of most of the baseline methods in feature selection.

Table 23 shows the average training computational costs for optimal hyper-parameter identification in deep CNNs for PH2 and ALL-IDB2 data sets over 30 runs. For both data sets, as shown in Table 23, HS, BBPSO and ACPSO have the most optimal training computational costs. RCPSO, PSO, DA, MFO, MPSO, and AGPSO show medium computational costs, with the remaining methods producing computationally expensive costs. Overall, ACPSO and RCPSO depict better or comparable training computational costs as compared with those of most of the baselines methods in optimal hyper-parameter selection.

Table 23 The average training computational costs (in seconds) for hyper-parameter fine-tuning in deep networks over a series of 30 runs

cost	**RCPSO**	**ACPSO**	**PSO**	**BBPSO**	**BA**	**CA**	**CS**	**DA**	**HS**	**ABC**	**FPA**	**MFO**
PH2	344.2298	**187.2935**	320.8068	**146.913**	420.2805	418.7278	369.0928	315.2647	**141.5545**	470.0163	421.0371	315.3719
ALL	373.6267	**296.3508**	387.6023	**207.8309**	519.337	514.9375	496.0683	387.7138	**173.7336**	571.4474	514.0983	390.4159

cost	**GPSO**	**MPSO**	**GMPSO**	**ELPSO**	**MFOPSO**	**AGPSO**	**DNLPSO**	**BBPSOV**
PH2	427.4532	316.5954	392.3869	482.9665	426.0251	313.64	429.2886	420.9588
ALL	519.5351	387.5924	473.3786	581.1205	516.6073	389.8353	513.4372	515.6972

5. CONCLUSIONS

In this research, we propose an intelligent skin lesion classification system. It consists of ABCD+GLRLM, LBP and HOG feature extraction, ACPSO and RCPSO feature selection, and deep and ensemble classifiers. The proposed ACPSO model employs both global search using adaptive decreasing and increasing acceleration coefficients as well as in-depth sub-dimension based local search mechanisms to attain global optima. The RCPSO model simulates mid-air hovering flight of hummingbirds, and uses the random

coefficients generated by three non-linear functions to increase both intensification and diversification capabilities.

Optimal hyper-parameter identification of a deep CNN network is performed using both proposed PSO models. The empirical results indicate efficiency of the proposed ACPSO and RCPSO algorithms for discriminative lesion feature selection and optimal hyper-parameter identification in deep networks. Both ACPSO and RCPSO models outperform nearly all the classical methods and the state-of-the-art PSO variants, statistically. The CNN model with the identified best training configurations also outperforms the model with the default hyper-parameter settings provided by MATLAB, significantly. The experiments also indicate efficiency of different types of lesion features contributing to melanoma classification. To further evaluate model efficiency and flexibility, two UCI data sets (i.e. breast cancer and epileptic seizure) and the ALL-IDB2 microscopic image data set are also used for evaluation. The proposed models outperform all the baseline methods for feature selection and optimal hyper-parameter identification of deep networks in most of the test cases for these data sets, as ascertained by the empirical and statistical test results.

In future work, other medical image data sets will be used to evaluate the proposed PSO models. Optimization of the deep network structures [55-57] will also be explored to further evaluate efficiency of the resulting models.

ACKNOWLEDGEMENT

We appreciate the support for this research received from the European Union (EU) sponsored (Erasmus Mundus) cLINK (Centre of excellence for Learning, Innovation, Networking and Knowledge) project (EU Grant No. 2645).

REFERENCES

[1] T.Y. Tan, L. Zhang and M. Jiang. 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.

[2] F. Adjed, S.J.S. Gardezi, F. Ababsa, I. Faye and S.C. Dass. Fusion of structural and textural features for melanoma recognition. *IET Computer Vision*. 12 (2), 3, 185–195. 2018.

[3] A. Sáez, J. Sánchez-Monedero, P.A. Gutiérrez and C. Hervás-Martínez. Machine Learning Methods for Binary and Multiclass Classification of Melanoma Thickness From Dermoscopic Images. *IEEE Transactions on Medical Imaging*. 35 (4) 1036–1045. 2016.

[4] D.R. Nayak, R.Dash, and B. Majhi. Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection. *Neurocomputing*. 282 (2018) 232–247.

[5] M. Issa, A.E. Hassanien, D. Oliva, A. Helmi, I. Ziedan, A. Alzohairy. ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment. *Expert Systems with Applications*. 99 (2018) 56–70.

[6] R.S. Patwal, N. Narang, and H. Garg. A novel TVAC-PSO based mutation strategies algorithm for generation scheduling of pumped storage hydrothermal system incorporating solar units. *Energy*. 142 (2018) 822–837.

[7] B. Xue, M. Zhang, and W.N. Browne. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing*. 18 (2014) 261–276.

[8] Y. Lu, M. Liang, Z. Ye, and L. Cao. Improved particle swarm optimization algorithm and its application in text feature selection. *Applied Soft Computing*. 35 (2015) 629–636.

[9] L.Y. Chuang, C.H. Yang, and J.C. Li. Chaotic maps based on binary particle swarm optimization for feature selection. *Applied Soft Computing*. 11 (2011) 239–248.

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

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

[12] K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim and B. Fielding. A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. *IEEE Transactions on Cybernetics*. 47 (6) 1496–1509. 2017.

[13] N.L.A. Krisshna, V. K. Deepak, K. Manikantan, and S. Ramachandran. Face recognition using transform domain feature extraction and PSO-based feature selection. *Applied Soft Computing*. 22 (2014) 141–161.

[14] M. Mafarja and S. Mirjalili. Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*. 62 (2018) 441–453.

[15] L. Yu, H. Chen, Q. Dou, J. Qin, and P.A. Heng. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Transactions on Medical Imaging*. 36 (4) 994–1004. 2017.

[16] C. Barata, M. Ruela, M. Francisco, T. Mendonça and J.S. Marques. Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features. *IEEE Systems*. 8 (3) 965–979, 2014.

[17] N.C.F. Codella, Q.-B. Nguyen, S. Pankanti, D.A. Gutman, B. Helba, A.C. Halpern and J.R. Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*. 61 (4) 5:1– 5:15. 2017.

[18] M. Kruk, B. Świderski, S. Osowski, J. Kurek, M. Słowińska and I. Walecka. Melanoma recognition using extended set of descriptors and classifiers. *EURASIP Journal on Image and Video Processing*. 2015:43, 2015.

[19] J. Sánchez-Monedero, M. Pérez-Ortiz, A. Sáezb, P.A. Gutiérrezc, and C. Hervás-Martínez. Partial order label decomposition approaches for melanoma diagnosis. *Applied Soft Computing*. 64 (2018) 341-355.

[20] B. Bozorgtabar, S. Sedai, P.K. Roy and R. Garnavi. Skin lesion segmentation using deep convolution networks guided by local unsupervised learning. *IBM Journal of Research and Development*. 61 (4) 6:1-6:8, 2017.

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

[22] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proc. IEEE Int. Conf. Neural Network*, Volume 4, 1942–1948. 1995.

[23] L.N. De Castro and F.J. Von Zuben. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6 (3) 239–251. 2002.

[24] D. Dasgupta, S. Yu and F. Nino. Recent advances in artificial immune systems: models and applications. *Applied Soft Computing*, 11 (2) 1574–1587. 2011.

[25] B.H. Ulutas and S. Kulturel-Konak. A review of clonal selection algorithm and its applications. *Artificial Intelligence Review*, 36 (2) 117–138. 2011.

[26] V. Cutello, G. Narzisi, G. Nicosia and M. Pavone. Clonal Selection Algorithms: A Comparative Case Study using Effective Mutation Potentials, optIA versus CLONALG, In *Proceedings of the 4th International Conference on Artificial Immune Systems (ICARIS)*, Canada. Springer, LNCS 3627:13–28, 2005.

[27] H.C. Huang, S.S.D. Xu and C.H. Wu. A hybrid swarm intelligence of artificial immune system tuned with Taguchi–genetic algorithm and its field-programmable gate array realization to optimal inverse kinematics for an articulated industrial robotic manipulator. *Advances in Mechanical Engineering*. 8(1), p.1687814015626380. 2016.

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

[29] P. Kinghorn, L. Zhang and L. Shao. A Hierarchical and Regional Deep Learning Architecture for Image Description Generation. *Pattern Recognition Letters*. 2019.

[30] T.Y. Tan, L. Zhang and C.P. Lim. Adaptive Melanoma Diagnosis Using Evolving Clustering, Ensemble and Deep Neural Networks. *Knowledge-based Systems*. In Press. 2019.

[31] T. Domhan, J.T. Springenberg and F. Hutter. Speeding up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*. 3460-3468. 2015.

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

[33] L. Ballerini, R.B. Fisher, B. Aldridge and J. Rees. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In *Color Medical Image Analysis*. Springer Netherlands. 63-86. 2013.

[34] T. Mendonça, P.M. Ferreira, J.S. Marques, A.R.S. Marcal and J. Rozeira. PH2 - A dermoscopic image database for research and benchmarking. In *Proceedings of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013.

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

[36] L. Shang, Z. Zhou, and X. Liu. Particle swarm optimization-based feature selection in sentiment classification. *Soft Computing*. 20 (10) 3821–3834. 2016.

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

[38] W. Chang. A modified particle swarm optimization with multiple subpopulations for multimodal function optimization problems. *Applied Soft Computing*. 33, 170–182. 2015.

[39] S. Mirjalili, A. Lewis and A.S. Sadiq. Autonomous Particles Groups for Particle Swarm Optimization. *Arabian Journal for Science and Engineering*. 39 (6) 4683–4697. 2014.

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

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

[42] H.C. Peng, F. Long and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 27 (8) 1226–1238. 2005.

[43] I. Kononenko. Estimating attributes: analysis and extensions of RELIEF. In *European Conference on Machine Learning*. Springer, Berlin, Heidelberg. 171-182. 1994.

[44] C. Barata, M.E. Celebi and J.S. Marques. Improving dermoscopy image classification using color constancy. *IEEE Journal of Biomedical and Health Informatics*, 19(3)1146-1152. 2015.

[45] N. Alfed, F. Khelifi and A. Bouridane. Improving a bag of words approach for skin cancer detection in dermoscopic images. In *Proceedings of International Conference on Control, Decision and Information Technologies*, 24-27. 2016.

[46] Z. Waheed, A. Waheed, M. Zafar and F. Riaz. An efficient machine learning approach for the detection of melanoma using dermoscopic images. In *Proceedings of International Conference on Communication, Computing and Digital Systems*, 316-319. 2017.

[47] J.S. Marques, C. Barata and T. Mendonca. On the role of texture and color in the classification of dermoscopy images. In *Proceedings of Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 4402-4405. 2012.

[48] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham and D. Feng. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, 64 (9) 2065-2074. 2017.

[49] K. Eltayef, Y. Li and X. Liu. Detection of pigment networks in dermoscopy images. *Journal of Physics: Conference Series*. 787 (1) 012033. IOP Publishing. 2017.

[50] A. Pennisi, D.D. Bloisi, D. Nardi, A.R. Giampetruzzi, C. Mondino and A. Facchiano. Skin lesion image segmentation using Delaunay Triangulation for melanoma detection. *Computerized Medical Imaging and Graphics*. 52 (2016) 89–103.

[51] C. Barata, J.S. Marques and J. Rozeira. A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE Transactions on Biomedical Engineering*. 59 (10) (2012) 2744–2754.

[52] R.D. Labati, V. Piuri and F. Scotti. ALL-IDB: The acute lymphoblastic leukemiaimage database for image processing. In *Proceeding of IEEE International Conference on Image Processing*, Brussels Belgium. IEEE, 2045–2048, 2011. Database URL: http://homes.di.unimi.it/scotti/all/.

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

[54] K. Bache and M. Lichman. UCI Machine Learning Repository. Irvine, CA: University of California. 2013, Database URL: http://archive.ics.uci.edu/ml.

[55] D. Pandit, L. Zhang, S. Chattopadhyay, C.P. Lim and C. Liu. A Scattering and Repulsive Swarm Intelligence Algorithm for Solving Global Optimization Problems. *Knowledge-Based Systems*. 156 (2018) 12-42.

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

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