Evolving Ensemble Models for Image Segmentation Using Enhanced Particle Swarm Optimization

**TECK YAN TAN[1], LI ZHANG [1], (Member, IEEE), CHEE PENG LIM[2],
BEN FIELDING[1], YONGHONG YU[3], AND EMMA ANDERSON[1]**

[1]Computational Intelligence Research Group, Department of Computer and Information Sciences, Faculty of Engineering and Environment, University of Northumbria, Newcastle upon Tyne NE1 8ST, U.K.
[2]Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds, VIC 3216, Australia
[3]College of Tongda, Nanjing University of Posts and Telecommunications, Nanjing 210028, China

Corresponding author: Li Zhang (li.zhang@northumbria.ac.uk)

ABSTRACT In this paper, we propose particle swarm optimization (PSO)-enhanced ensemble deep neural networks and hybrid clustering models for skin lesion segmentation. A PSO variant is proposed, which embeds diverse search actions including simulated annealing, levy flight, helix behavior, modified PSO, and differential evolution operations with spiral search coefficients. These search actions work in a cascade manner to not only equip each individual with different search operations throughout the search process but also assign distinctive search actions to different particles simultaneously in every single iteration. The proposed PSO variant is used to optimize the learning hyper-parameters of convolutional neural networks (CNNs) and the cluster centroids of classical Fuzzy C-Means clustering respectively to overcome performance barriers. Ensemble deep networks and hybrid clustering models are subsequently constructed based on the optimized CNN and hybrid clustering segmenters for lesion segmentation. We evaluate the proposed ensemble models using three skin lesion databases, i.e., PH2, ISIC 2017, and Dermofit Image Library, and a blood cancer data set, i.e., ALL-IDB2. The empirical results indicate that our models outperform other hybrid ensemble clustering models combined with advanced PSO variants, as well as state-of-the-art deep networks in the literature for diverse challenging image segmentation tasks.

INDEX TERMS Convolutional neural network, ensemble model, Fuzzy C-Means clustering, image segmentation, particle swarm optimization.

I. INTRODUCTION

Malignant melanoma is the most deadly skin cancer. There are 132,000 new cases of melanoma diagnosed worldwide each year (World Health Organization). Automatic early diagnosis of melanoma is critical in administering effective treatment and increasing the survival rate. Skin lesion segmentation, which delineates the lesion foreground from the skin, is a vital step in delivering a subsequent reliable and robust diagnosis. However, despite great efforts, the retrieval of distinguishing complete lesion/tumour boundaries for diverse cases is still a challenging task owing to large variations in size, shape, texture, occlusion, and fuzzy indistinguishable boundaries of tumours.

In this research, we propose evolving ensemble deep networks and hybrid clustering models to undertake skin lesion segmentation. A cascade Particle Swarm Optimization (PSO) algorithm is proposed to optimize the learning hyper-parameters of deep Convolutional Neural Networks (CNNs) and the cluster centroids of Fuzzy C-Means (FCM) clustering, respectively, to enhance the segmentation performance. Such processes are used to overcome local optima traps, initialization sensitivity of FCM clustering, and the difficulties in optimal hyper-parameter selection of deep pixelwise classification networks. In order to overcome the bias and variance of individual segmenters, ensemble deep networks and hybrid clustering models are subsequently

The associate editor coordinating the review of this manuscript and approving it for publication was Kashif Munir.
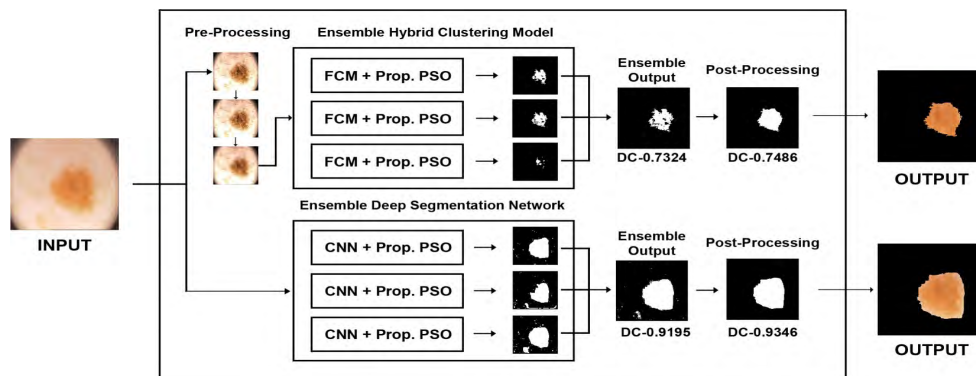
constructed based on the enhanced base CNN and clustering models, respectively. A majority voting strategy is used to combine the prediction results of each base model to produce the final pixelwise classification outcome. A series of post-processing procedures, including image dilation, hole filling, border clearance and smoothing, and small object removal, is subsequently conducted to further enhance the performance. Figure 1 shows the system architecture. The research contributions are highlighted, as follows.

1) We propose a cascade swarm intelligence (SI) algorithm to undertake skin lesion segmentation based on PSO. It is used to optimize the hyper-parameters of deep CNNs and the cluster centroids of traditional FCM clustering. On top of the enhanced base of CNN and hybrid FCM models, two ensemble segmenters are subsequently constructed, i.e. ensemble deep networks and hybrid clustering models. Each ensemble segmenter consists of three base clustering models or the CNNs, respectively.

2) The proposed PSO variant embeds several search operations including Simulated Annealing (SA), Levy flight distribution, helix search actions, modified PSO and Differential Evolution (DE) with spiral search coefficients. These search strategies work in a cascade manner to increase local and global search capabilities. Specifically, for each particle, each of the above-mentioned actions is inherited in the subsequent iterations to accelerate convergence if its search process shows improvements, otherwise an alternative search behaviour is performed. This proposed cascade search mechanism not only equips each particle with different search operations during the search process, but also enables the swarm particles to conduct distinctive search actions simultaneously in any single iteration (e.g. in iteration t, particle i conducts SA, while particle j conducts helix search and particle k performs modified PSO operation with spiral search coefficients). The above proposed mechanisms work cooperatively to mitigate premature convergence of the original PSO model.

3) Firstly, the proposed cascade PSO model is used to further improve the centroids of FCM clustering to overcome sensitivity to noise and membership initialization, as well as local optima traps. Triple enhanced FCM models are generated using different experimental settings. In order to overcome bias and variance of individual segmenters, an ensemble clustering model is subsequently constructed using the above enhanced hybrid segmenters.

4) Secondly, deep learning models show superior performances in diverse computer vision tasks. However, the optimal parameter identification has been mainly conducted using an exhaustive grid-search [1]–[3], which may lead to an expensive computational cost. Therefore, we employ the proposed PSO model to provide an automatic optimization procedure to identify the optimal hyper-parameters (namely the learning rate and weight decay parameters) of a deep CNN model with dilated convolutions. Three deep pixelwise classification CNNs with different optimized hyper-parameters are generated, which are subsequently used to form a deep ensemble network. A pixelwise majority voting is conducted pertaining to both the clustering and deep ensemble models to obtain the final pixelwise classification results.

5) Three skin lesion data sets, i.e. PH2, Dermofit Image Library, and ISIC 2017, and a blood cancer data set, ALL-IDB2, are used to evaluate the proposed evolving ensemble deep CNN networks and clustering models. Both ensemble models show impressive performances and outperform state-of-the-art deep learning networks such as U-Net and other enhanced ensemble clustering models incorporating diverse advanced PSO variants, significantly.

The rest of the paper is organized as follows. Section II presents related studies on clustering and deep semantic segmentation models, and diverse PSO variants. The proposed cascade SI algorithm and the construction of evolving ensemble deep networks and clustering models are discussed comprehensively in Section III. The evaluation results and

comparison studies are provided in Section IV. Section V concludes this research and presents future directions.

II. RELATED WORK

In this section, we discuss state-of-the-art PSO variants and related studies on image segmentation using deep networks and clustering models.

A. IMAGE SEGMENTATION

Diverse threshold-, clustering- and deep learning-based segmentation methods have been proposed in the literature to retrieve complete distinguishing boundaries for tumour segmentation.

Long *et al.* [4] proposed a Fully Convolutional Network (FCN) for semantic segmentation. The FCN model is trained in an end-to-end fashion, and is able to accept input images with different sizes for dense output prediction. A skip architecture is also embedded to incorporate semantic information from a deep layer with appearance information from a shallow layer for pixelwise prediction. Ronneberger *et al.* [5] proposed the U-Net architecture for lesion segmentation. The model consists of both encoding and decoding paths. Skip-layers are also used to enable each decoding layer to accept inputs from its previous layer as well as its symmetric encoding layer. The network showed impressive performances on diverse biomedical segmentation tasks. Fernandes *et al.* [6] proposed a CNN model to measure the quality of an image-segmentation pair. The CNN employed the learned model and the backpropagation algorithm to perform image segmentation in an iterative process. Gossip Networks were proposed to enable the communication between the foreground and background streams. Besides generating segmentation masks from scratch, their deep network was capable of improving the outputs produced by other segmentation methods (e.g. U-Net). The work was evaluated using biomedical data sets pertaining to segmentation of skin lesions, teeth, iris, etc. The developed model outperformed U-Net and U-Net with dilated convolutions (Dilated-Net), respectively. Izadi *et al.* [3] employed generative adversarial networks (GANs) for skin lesion segmentation. An additional critic deep learning CNN model was formulated, on top of the segmenter FCN network (such as U-Net). The segmenter synthesized segmentation masks whereas the critic network distinguished the synthesized masks from real ground truth (GT). Their work indicated that the incorporation of the critic model with the segmenter network showed an enhanced performance in comparison with increasing the segmenter network complexity. They evaluated the added critic CNN by combining it with the U-Net segmenter, which outperformed U-Net when using the Dermofit skin lesion data set. Proposed by Vesal *et al.* [7], the SkinNet architecture was a modified version of U-Net. In comparison with U-Net, it employed dense convolution blocks, instead of convolution layers, in the encoder and decoder paths. It also incorporated dilated convolutions in the lowest encoder layer, in order to better capture global context information.

A further study of their work [8] employed the Faster region-based convolutional neural network (Faster-RCNN) and SkinNet for lesion detection and localization, as well as segmentation, respectively. Specifically, Faster-RCNN was used to produce the bounding boxes of the detected lesion regions, which were subsequently cropped for lesion segmentation using SkinNet. Their two-stage system showed superior performance for lesion segmentation.

Other deep learning networks have also been proposed for lesion segmentation. Venkatesh *et al.* [9] proposed a deep architecture with multi-scale residual connections for lesion segmentation, known as Multi-scale residual U-Net. Their model employed multi-scale residual connections to tackle information loss in the encoding stages in the U-Net. The model outperformed other state-of-the-art networks such as deep fully convolutional-deconvolutional neural networks [10] for diverse challenging lesion segmentation tasks. Goyal and Yap [11] conducted transfer learning based on the pre-trained FCN models for multi-class skin lesion segmentation.

There are also other state-of-the-art hybrid clustering models proposed in recent years for biomedical image segmentation. Aljawawdeh *et al.* [12] employed GA and PSO to obtain the initial centroids of the K-Means (KM) clustering model. The KM algorithm was subsequently used for melanoma/benign lesion segmentation. KM clustering and ensemble regression were used for skin lesion segmentation in Alvarez and Iglesias [13]. Louhichi *et al.* [14] employed multiple density clustering algorithms to identify the key parameters of region growing techniques for lesion segmentation. Pham *et al.* [15] produced a modified PSO model in combination with kernelized fuzzy entropy clustering (KFEC) for Magnetic Resonance Imaging (MRI) brain image segmentation. Their PSO model employed the Halton sequence for population initialization and an adaptive inertia weight to accelerate convergence. Besides KFEC, their objective function also incorporated local spatial information and bias correction for particle evaluation. Their model outperformed five other competitors when evaluated using noisy simulated and real MRI brain images. Singh and Bala [16] proposed an enhanced discrete cosine transform (DCT)-based nonlocal FCM (DCT-NLFCM) method for MRI brain image segmentation. Their model employed DCT-domain pre-processed images to attain fast segmentation in comparison with other unsupervised DCT-based methods. The empirical results indicated that their model was invariant to noise with superior segmentation performances. A multi-objective spatial fuzzy clustering model was developed by Zhao *et al.* [17] for image segmentation. Non-dominated Sorting Genetic Algorithm-II (NSGA-II) was incorporated with the clustering model to conduct image segmentation. The objective function took into account the intra-cluster fuzzy compactness as well as inter-cluster fuzzy separation derived from non-local spatial information. A cluster validity index was also produced to help retrieve the best solution among the generated non-dominated individuals. The number

of clusters was also identified in an evolving process using a real-coded variable string length method. Evaluated with images contaminated by noise, the developed model achieved better performances in comparison with those of KM, FCM and several FCM variants.

Pan *et al.* [18] proposed a bacterial foraging evolutionary algorithm for cell image segmentation. Their proposed edge detection model was able to overcome the limitation of initialization sensitivity of the traditional edge detectors. It calculated a bright pixel density map for nutrient concentration estimation. Evaluated using synthetic and real cell images, the proposed model achieved an improved segmentation accuracy in comparison with those of several well-known traditional edge detectors such as the active contour model and the Canny edge detector. Neoh *et al.* [19] conducted nucleus-cytoplasm segmentation for blood cancer detection using a hybrid model of FCM clustering integrated with the Genetic Algorithm (GA). Their fitness function took both intra- and inter-cluster variances into account. The hybrid model showed impressive performances and outperformed other state-of-the-art FCM variants for nucleus-cytoplasm segmentation using the ALL-IDB2 data set. Dai *et al.* [20] performed optic disc (OD) segmentation in fundus images using a variation model with boundary, shape, and region energies. A sparse coding based technique was initially used to perform optic disc localization. Subsequently, a region of interest was cropped based on the localized disc centre and the surrounding area of OD in the image. Blood vessel removal was also conducted using morphological operations before segmentation. The Hough transform was conducted to obtain the initial boundary information. Then, three energies, i.e. the phase-based boundary, PCA-based shape and region energies, were used to further enhance the OD segmentation results.

B. PSO VARIANTS

PSO, one of the popular SI algorithms, has been adopted in diverse single-, multi- and many-objective optimization problems [21]. The PSO operation employs the following strategy for position updating.

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

where x_{id}^{t+1} and v_{id}^{t+1} denote the position and velocity of particle x_i in the d-th dimension and the $t + 1$-th iteration, respectively, with x_{id}^t and v_{id}^t representing the associated parameters in the t-th iteration. p_{id} and p_{gd} represent the personal best experience and the swarm leader in the d-th dimension respectively. In addition, c_1 and c_2 refer to acceleration coefficients and r_1 and r_2 denote the random vectors while w represents an inertia weight that defines the impact of the previous velocity over iterations. Equations (1)-(2) indicate that all the particles in PSO are

attracted towards the global best solution. As such, it is very likely to converge prematurely.

Diverse search strategies have been proposed to overcome premature convergence of the classical PSO model [22]–[25]. Fielding and Zhang [26] proposed an enhanced PSO variant for evolving deep architecture generation for image classification. Their PSO model employed several cosine annealing mechanisms to generate adaptive acceleration coefficients to overcome local optima traps and accelerate convergence. A weight sharing mechanism between similar optimized architectures was used to reduce the training cost. In comparison with other state-of-the-art deep learning models, their work achieved impressive error rates of 4.78% and 25.42% based on the CIFAR-10 and CIFAR-100 data sets, respectively. Moreover, in comparison with related studies using 250 and 20 GPUs, their experiments were performed using a single GPU with a significantly reduced computational cost. AGPSO was proposed by Mirjalili *et al.* [27], where a number of linear and nonlinear functions were proposed for adaptive acceleration coefficient generation. Their PSO model enabled the search process to focus on the cognitive component for global exploration in early iterations and the social component for local exploitation in subsequent iterations. The classical PSO model was also used in the deep learning architecture generation for convolutional auto-encoder in Sun *et al.* [28], while the GA was employed in the block architecture generation based on ResNet blocks and DenseNet blocks in Sun *et al.* [29]. A hybrid optimization model incorporating PSO with the Firefly Algorithm (FA) was proposed by Aydilek [30]. Their model employed FA and PSO for exploitation and exploration, respectively. Specifically, according to the criterion that each individual solution in the swarm was fitter than the current swarm leader or otherwise, either the FA or PSO operation was used for position updating accordingly. The model showed superior performances for evaluation of unimodal, multimodal, hybrid, and composition benchmark functions.

Ye [31] proposed a PSO variant for configuration optimization of a deep fully-connected neural network, known as a multi-layer perceptron (MLP) for classification and regression tasks. The acceleration coefficients and the inertia weight of their PSO model were determined by linear schedules. The model was then used to optimize the number of neurons in each fully-connected layer, as well as overall hyper-parameters including the learning rate and dropout factor. A single best model from the global best solution as well as an ensemble model consisting of a configurable, pre-determined number of models derived from the final local best positions were produced. Based on experiments with different numbers of hidden layers, the depth of the networks was modified empirically. The developed model was evaluated using the MNIST data set for handwritten digit classification and the KCD data set for biological activity prediction through regression. Soon *et al.* [32] used the PSO model to optimize the learning rate and kernel configuration for a CNN for vehicle logo recognition. The optimized

deep network was subsequently used to classify vehicle logo images, which was followed by pre-processing to segment the logo areas from full vehicle images. The PSO model was used to optimize seven hyper-parameters consisting of the learning rate as well as the number and size of convolutional filters for a fixed number of three layers in the network. The outcome indicated that using an optimization approach resulted in a more efficient system than manually tuning.

III. THE PROPOSED IMAGE SEGMENTATION SYSTEM

The proposed lesion segmentation system consists of two key steps. The first step employs ensemble deep networks and clustering models for lesion segmentation, respectively. The post-processing procedures using morphological operations are used in the second step to fill in holes and smooth lesion boundaries, in an attempt to further improve the segmentation results. In order to identify the optimal hyper-parameter settings for base CNNs and overcome initialization sensitivity of classical FCM clustering base models, a cascade PSO algorithm is proposed. It is used to fine-tune the learning hyper-parameters of each base semantic segmenter network and further improve the cluster centroids identified by FCM clustering. A majority voting strategy is employed to combine pixelwise classification results of all the base models for each ensemble predictor. The ensemble segmentation output is then enhanced using the post-processing morphological operations. We introduce each key stage of the proposed semantic segmentation models comprehensively, as follows.

A. THE PROPOSED PSO MODEL

In this research, a cascade PSO algorithm is proposed for image segmentation. It is used to enhance FCM centroids as well as devise the optimal learning hyper-parameters for each CNN segmenter. The proposed PSO algorithm includes not only diverse search operations such as SA and Levy flight based local exploitation, but also helix, PSO and DE strategies for global exploration. A hierarchical mechanism is used to assign different individuals with different search actions in each iteration to mitigate premature convergence. Moreover, the enhanced helix, PSO and DE operations follow two remote swarm leaders simultaneously in a spiral fashion. The pseudo-code of the proposed PSO algorithm is shown in Algorithm 1. Our PSO model not only has better chances of finding global optima to retrieve effective learning hyper-parameters and cluster centroids, but also provides diverse efficient base segmenters for deep and clustering ensemble pixelwise prediction model construction.

The proposed algorithm initializes a swarm of particles, which are subsequently evaluated. In addition to the swarm leader, a second swarm leader with a similar fitness score but low position proximity to that of the best leader is also identified. The MATLAB function, *corr*2, is used to determine the position proximity between two particles. It generates a score in the range of $[-1, 1]$, where '1' and '−1' specify the two particles are exactly the same or distinctive entirely in positions.

Algorithm 1 The Proposed Cascade PSO Algorithm
1   **Start**
2   Initialize a population randomly (e.g. 50 particle);
3   Evaluate the population;
4   **For** (each iteration **do**) {
5   Find the swarm leader, g_{best};
6   Find a second leader with a similar fitness but remote in position to g_{best};
7   **For** (each particle **do**) {
8   **If** (Iteration == 1 or Flag == SA)
9   Conduct SA operation as defined in Equation (3);
10   **Else If**(Flag == helix/PSO)
11   Randomly select one of the following operations to update the particle position;
12   1. Conduct the proposed helix operation as defined in Equations (4), (5) and (7);
13   2. Conduct a modified PSO operation as defined in Equations (9), (2) and (7);
14   **Else If**(Flag == DE)
15   Conduct a modified DE operation using three parents as defined in Equation (11)-(13) and (7);
16   **Else If**(Flag == Levy)
17   Conduct long jumps using Levy flight distribution as defined in Equation (14);
18   **Else If**(Flag == Scatter)
19   Re-initialize the particle by scattering all the dimensions as defined in Equation (15);
20   Flag = SA;
21   **End If**
22   Conduct fitness evaluation;
23   Update the personal best p_{best} and global best g_{best} if the new position has better fitness;
24   **If** (SA has improvement)
25   Flag = SA;
26   **Else If** (SA search has no improvement)
27   Flag = Proposed helix/modified PSO;
28   **Else If** (PSO/helix search has no improvement)
29   Flag = Modified DE;
30   **Else If** (DE-based search has no improvement)
31   Flag = Levy flight;
32   **Else If** (Levy search has no improvement)
33   Flag = Scatter;
34   **End If**
35   } **End For**
36   } **End For**
37   Output g_{best};

After initialization, in the 1st iteration, an SA operation is performed by the overall population. The particles that do not show improvements employ an alternative helix or PSO

search action in the 2nd iteration to avoid stagnation, while the SA operation is inherited by those with improved fitness performance. Similarly in the 3rd iteration, if the previously employed helix or PSO search movements fail to improve the fitness scores of some particles, a modified DE-based search process is applied to further diversify the search process, whereas those with improved fitness scores continue with the previous search strategies, i.e. PSO, helix, or SA actions. In the 4th iteration, if the DE-based search process is not effective for enhancement of some particles, a Levy random walk is conducted subsequently. Similarly, for those with performance improvements, the previous search actions (e.g. DE or others) are inherited and remain intact. Finally, if the Levy-based jump fails to lead some individuals to optimal regions, a scattering action is conducted to re-initialize such particles, otherwise the previous search operations are repeated to accelerate convergence. Such a cascade search strategy is applied to each swarm particle to equip it with different search actions across iterations. Meanwhile, it also assigns different search strategies to different particles simultaneously within single iteration (e.g. particle 1 performs SA, while particles 2, 3, 4, and 5 perform PSO, DE, Levy distribution, and scattering operations, respectively, in the t-th iteration). To increase local and global search capabilities, spiral search parameters are proposed to enable the individuals to follow promising solutions in a spiral manner. The algorithm terminates when the maximum number of iterations is reached. We introduce each proposed search operation in detail below.

1) THE SA-BASED LOCAL EXPLOITATION

After initialization of the swarm, the SA operation [33] is assigned to each particle in the 1st iteration. It employs both weak and promising solutions to guide the search process. A transition probability, p, defined in Equation (3), is employed in SA to determine the chances for a weak solution to be accepted.

$$p = exp\left(-\frac{\Delta f}{T}\right) > r \qquad (3)$$

where Δf and T denote the fitness score change and the temperature for controlling the annealing process, respectively. When the transition probability, p, is greater than a randomly generated threshold, r, the weak solution is accepted. To balance between local exploitation and global exploration, we also employ a geometric cooling schedule, i.e. $T = \alpha T$, to decrease the temperature, T, by a cooling factor, $\alpha \in [0, 1]$. SA shows great capabilities in overcoming local optima traps, but at the expense of a large number of function evaluations.

2) THE PROPOSED HELIX MOVEMENT

A novel helix search algorithm is proposed to diversify the search process. It enables each individual to follow the first and the second swarm leaders simultaneously in randomly selected sub-dimensions in a spiral manner. Equations (4)-(5) define the proposed spiral movement. Its search parameters are drawn from a three-dimensional helix distribution defined

in Equations (6)-(8).

$$x_{id}^{t+1} \begin{cases} x_{id}^t + h\left(p_{gd} - x_{id}^t\right) \\ \quad + \zeta\left(rand - 0.5\right) \times scale, & \text{if } d \in [1, rand_{dims}] \\ x_{id}^t + h\left(p_{sd} - x_{id}^t\right) \\ \quad + \zeta\left(rand - 0.5\right) \times scale, & \text{otherwise} \end{cases} \qquad (4)$$

$$\zeta = \zeta \times \left(\frac{10^{-4}}{0.9}\right)^{\frac{1}{maxi_Gen}} \qquad (5)$$

$$g = 0.0065 \times t \times \cos\left(2 \times t\right) g \in [-2.522, 2.532], \\ t \in [0, 389.5575] \qquad (6)$$

$$h = 0.0065 \times t \times \sin\left(2 \times t\right) h \in [0, 2.517] \qquad (7)$$

$$l = t \qquad (8)$$

where h denotes the proposed helix search coefficient. As mentioned earlier, it is randomly extracted from a spiral distribution defined in Equation (7). In addition, p_{gd} and p_{sd} denote the first and second swarm leaders in the d-th dimension, respectively. According to Equation (4), the search agent, x_i, not only follows the global best solution, p_g, in some randomly selected sub-dimensions, but also spirally moves towards the second swarm leader, p_s, in the remaining sub-dimensions. The *scale* parameter is a constant vector with each element defined as the difference between the upper and lower boundaries of each dimension. As illustrated in Equation (5), ζ is an adaptive search parameter to fine-tune the random search steps. It enables the adoption of a larger random search step in early iterations to increase global exploration, and a smaller search step in subsequent iterations to perform fine-tuning.

3) THE MODIFIED PSO OPERATION

The PSO model shows superior search capabilities in solving single-, multi-, and many-objective optimization problems. However, owing to the search process led by a single swarm leader, it is likely to be trapped in local optima. In order to prevent from stagnating prematurely, as indicated in the helix movement, two remote swarm leaders are employed to lead the search process simultaneously. Equation (9) defines this new PSO search mechanism.

$$v_{id}^{t+1} = \begin{cases} wv_{id}^t + hr_1\left(p_{id} - x_{id}^t\right) \\ \quad + hr_2\left(p_{gd} - x_{id}^t\right), & \text{if } d \in [1, rand_{dims}] \\ wv_{id}^t + hr_1\left(p_{id} - x_{id}^t\right) \\ \quad + hr_2(p_{sd} - x_{id}^t), & \text{otherwise} \end{cases} \qquad (9)$$

where h denotes the proposed helix search coefficient. Equation (9) indicates that each particle moves towards p_g in randomly selected sub-dimensions and p_s in the remaining sub-dimensions, respectively, to avoid local optima traps. Instead of using fixed acceleration coefficients as in classical PSO, the random search parameters extracted from a helix distribution defined in Equation (7) are used to guide the social and cognitive search components. These search parameters enable the particles to follow the local and global promising

solutions in a spiral manner to increase search exploration and, at the same time, fine-tune the solution vectors.

4) THE MODIFIED DE OPERATION

DE possesses efficient convergence capabilities in solving diverse optimization problems. It comprises three primary steps, i.e. mutation, crossover, and selection. There are a variety of mutation schemes in DE. A popular mutation scheme of DE, i.e. the DE/Best/1/Bin scheme, is defined in Equation (10).

$$D_i^{t+1} = x_{best}^t + F(x_l^t - x_m^t) \tag{10}$$

where D_i^{t+1} represents a donor vector in the $t+1$-th iteration, while x_{best}^t, x_l^t and x_m^t denote the current best solution and two randomly selected distinctive solution vectors in the t-th iteration, respectively. F is the differential weight.

In order to mitigate local optima traps, in this research, again we take two remote leaders into account to guide the search process while a random scale factor based on the helix distribution is also embedded. This improved mutation scheme is defined in Equation (11).

$$D_{id}^{t+1} = \begin{cases} p_{gd} + h(x_{ld}^t - x_{md}^t), & \text{if } d \in [1, rand_{dims}] \\ p_{sd} + h(x_{ld}^t - x_{md}^t), & \text{otherwise} \end{cases} \tag{11}$$

In comparison with the original mutation scheme shown in Equation (10), the donor vector illustrated in Equation (11) is generated using two remote swarm leaders. Specifically, when d belongs to the randomly selected sub-dimensions, the new solution is generated using the current global best solution, p_g, otherwise, it is produced using the second leader, p_s. The differential weight parameter is randomly generated using the helix distribution as defined in Equation (7). The resulting search process, therefore, enables the donor solution to conduct local exploitation around the two swarm leaders in randomly selected sub-dimensions simultaneously in a spiral fashion.

Subsequently, a crossover operation defined in Equation (12) is performed.

$$k_{id}^{t+1} = \begin{cases} D_{id}^{t+1}, & \text{if } rand \leq pC_r \\ x_{id}^t, & \text{otherwise} \end{cases} \tag{12}$$

The generation of a new solution, k_i^{t+1}, is controlled by a crossover parameter, pC_r. It determines whether each dimension of the new solution is inherited from that of the donor vector, D_i^{t+1}, or the previous solution, x_i^t. This new offspring solution, k_i^{t+1}, is then evaluated using the objective function. If it has an improved fitness score than that of the previous solution, x_i^t, it is selected, otherwise the previous solution, x_i^t, is selected, and passed on to the next iteration as defined in Equation (13).

$$x_i^{t+1} = \begin{cases} k_i^{t+1}, & \text{if } f\left(k_i^{t+1}\right) \geq f(x_i^t) \\ x_i^t, & \text{otherwise} \end{cases} \tag{13}$$

This modified DE operation increases search capabilities by enabling the particles to perform local exploitation of the two best leaders simultaneously in a helix manner.

5) RANDOM WALK USING LEVY FLIGHT

A random search strategy based on the Levy flight distribution is also incorporated, in order to overcome local optima traps, which is defined in Equation (14).

$$x_{id}^{t+1} = x_{id}^t + \left(x_d^{max} - x_d^{min}\right) \times \vartheta \tag{14}$$

where x_d^{max} and x_d^{min} indicate the upper and lower boundaries in the d-th dimension, respectively, and ϑ denotes the Levy flight operation. A new offspring solution is generated using this Levy random walk strategy, which is used to replace the previous solution if it has a better fitness score. This Levy operation depicts great search capabilities in exploring an unknown large-scale search space.

6) THE SCATTERING BEHAVIOUR

A scattering action is also implemented to re-initialize any weak particles showing limited performance improvements in the last few iterations. This scattering operation is defined in Equation (15).

$$x_{id}^{t+1} = x_d^{min} + \left(x_d^{max} - x_d^{min}\right) \times rand \tag{15}$$

where *rand* denotes a randomly generated vector with each element in the range of [0, 1]. It is used to increase swarm diversity and reduce the probabilities of converging prematurely.

As discussed earlier, the series of proposed search actions work in a cascade manner to increase diversification and accelerate convergence. Specifically, it assigns different search operations to different particles in each single iteration. It also enables each particle to switch between several different search actions across iterations to alleviate stagnation. Furthermore, the proposed cascade PSO model is used to enhance FCM cluster centroids and hyper-parameters in CNN models to improve segmentation performance. Next, we introduce deep and clustering image segmentation processes, as follows.

B. THE CLUSTERING-BASED IMAGE SEGMENTATION

The proposed PSO algorithm is first used to enhance the centroids of FCM clustering for image segmentation. Clustering models such as FCM and KM are sensitive to membership and centroid initialization and image noise. Their clustering processes also tend to be trapped in local optima. These clustering methods are usually used to provide an initial analysis of data distribution. Evolutionary algorithms show great superiority in attaining global optima, therefore the proposed PSO model is employed to further enhance the centroids identified by the FCM model and to increase segmentation robustness.

The detailed hybrid clustering process incorporating FCM with the proposed PSO model is introduced below. A series

of pre-processing procedures is first applied, including hair removal, sharpness and contrast enhancement. In addition, the search process of the proposed PSO model employs two dimensions to represent the centroids of two clusters, e.g. skin vs lesion, or nucleus vs background. We also assign the initial centroids produced by FCM to one of the particles as the initial seed solution, with the rest of the swarm randomly initialized. The search process of cluster centroid enhancement is conducted by the abovementioned proposed search strategies, until the maximum number of iterations is reached. The final swarm leader representing the best centroids is used to perform the lesion/tumour segmentation. Equation (16) defines the fitness function for particle evaluation, which takes both intra- and inter-cluster variances into account [19].

$$fitness\,(x_i) = M_{intra-cluster}/M_{inter-cluster} \qquad (16)$$

where $M_{intra-cluster}$ and $M_{inter-cluster}$ denote the intra-cluster and inter-cluster measures, respectively. The optimization process minimizes the fitness function by preferring clusters with high compactness and large separation. Such enhanced hybrid clustering model is used as the base evaluator for ensemble clustering segmenter construction subsequently.

C. EVOLVING DEEP NETWORKS FOR IMAGE SEGMENTATION

Deep learning models show impressive performances for diverse computer vision tasks. Optimal hyper-parameter identification has a considerable impact on deep network performances. For instance, a large or small learning rate leads to a suboptimal result or a significant computational cost in the training stage, while the regularization factor (i.e. weight decay) has a significant impact in reducing model overfitting. Some related research studies employ random or exhaustive grid-search mechanisms as popular strategies for parameter selection [1]–[3]. However, such methods are more likely to be trapped in local optima or at the expense of a large computational cost. In order to deal with these limitations, in this research, we employ the proposed cascade model to optimize the two primary hyper-parameters of a CNN segmenter, i.e. the learning rate and the regularization factor (denoted as L2Regularization), to enhance performance.

To better capture global information and avoid information loss, we employ a deep semantic network with dilated convolutions as the base segmenter. Dilated convolutions [34] refer to the technique for 'dilating' convolutional kernels, in order to increase the receptive field size of the convolutional layer without introducing more parameters or reducing the network entropy. Dilating the kernels effectively 'spaces out' the individual filter values over the convolutional window, thereby maintaining the same dimensionality of the kernel itself but also incorporating more global information in each window position of the convolution operation. A dilated convolution has the same number of parameters as a non-dilated convolution, which leads to no increase in the overall number of network parameters. We therefore employ such a semantic network as the base evaluator.

TABLE 1. The Deep CNN Architecture.

	filterSize	numFilters	DilationFactor	Padding
Image input layer	-	-	-	-
Convolution layer	3	64	1	same
batchNormalization Layer	-	-	-	-
reluLayer	-	-	-	-
Convolution layer	3	64	2	same
batchNormalization Layer	-	-	-	-
reluLayer	-	-	-	-
Convolution layer	3	64	4	same
batchNormalization Layer	-	-	-	-
reluLayer	-	-	-	-
Convolution layer	1	2	-	0
softmaxLayer	-	-	-	-
pixelClassificationLayer	-	-	-	-

The topology of the base deep segmenter network is illustrated in Table I. It constitutes one image input layer, which accepts an image input size of 128×128. Three blocks of convolutional layers are subsequently attached. Specifically, each convolutional layer has a kernel size of 3×3 and 64 filters with increasing dilation factors. In order to pad the inputs to be the same size as the outputs for segmentation tasks, the option of 'same' is used for the padding setting. In addition, the final pixel classification block consists of one convolutional layer with q (i.e. number of classes) filters and a kernel size of 1×1, as well as a softmax layer and a pixelClassificationLayer. Moreover, since the majority of the image pixels belong to the background class, the image segmentation tasks tend to have severe class imbalance problems. An inverse frequency weighting method is also used to generate class weights as the inverse of the class frequencies to deal with the class imbalance bias.

The search of optimal settings for the two training parameters in the CNN segmenter employs the following setting, i.e. dimension = 2, population = 15 and iteration = 10, which yields the best trade-off between performance and cost. The optimal parameter search process is guided by the proposed PSO mechanism. The stochastic gradient descent with momentum optimizer is used as the solver for network training, with the maximum number of epochs = 100 and the mini-batch size = 64. Unlike the above proposed clustering-based segmentation, the pre-processing procedures are not required for the deep ensemble networks with optimal parameter settings. Moreover, we have constructed a baseline network with the default settings of the learning rate and the L2Regularization factor, i.e. 0.01 and 0.0001, respectively, in accordance with MATLAB. The CNN segmenter with the identified optimized hyper-parameters is also compared with this default baseline model, and other existing deep architectures for performance comparison.

In this research, we employ the Sørensen-Dice similarity coefficient, defined in Equation (17), as the objective function. This Dice similarity coefficient is used to evaluate the

superiority of the segmentation results against the ground truth. It is calculated using twice the intersection divided by the union of the ground truth and generated masks.

$$D\left(m, m'\right) = \frac{2|m \cap m'|}{|m| + |m'|} \tag{17}$$

where m and m' represent the ground truth and the predicted masks, respectively.

D. ENSEMBLE DEEP NETWORK AND CLUSTERING MODEL CONSTRUCTION & POST-PROCESSING

In order to overcome bias and variance of single base evaluators, ensemble models combining several base evaluators are produced to enhance segmentation performance. To produce diverse different evolving base evaluators, different experimental settings are employed for the base model generation for both deep and clustering algorithms, i.e. inertia weight = 0.6/0.4/0.2, population = 5/10/15 (deep) or 20/30/50 (clustering), and iteration = 3/5/10 (deep) or 10/15/20 (clustering). For constructing the base CNN segmenters, the training and validation data sets are shuffled once for each fitness evaluation as well as the final model fine-tuning step. In this way, the diversity of each single base CNN model is further enhanced. All experiments pertaining to the proposed deep ensemble networks were performed on a single NVIDIA GTX 1080Ti consumer GPU.

As mentioned earlier, triple CNNs with different optimized hyper-parameters and triple hybrid clustering models with different optimized centroids are produced using diverse experimental settings. These CNN and clustering base evaluators are used to construct an ensemble deep network as well as an ensemble clustering model, respectively. Each base evaluator conducts the segmentation tasks individually. A majority voting mechanism is then used to combine the base model results to generate the predicted mask for the ensemble segmenter.

A set of post-processing procedures is subsequently undertaken to improve ensemble segmentation performance. For instance, the generated binary gradient masks for lesion segmentation sometimes show interior gaps or holes as well as 'salt-and-pepper' effects. Morphological operations, such as image dilation, hole filling, border clearance, small object removal and border smoothing, are used to deal with such problems. MATLAB functions, such as strel, imfill, imclearborder and bwareaopen, are used in these processes. Some example segmentation results of the proposed deep and clustering ensemble models are illustrated in Figures 2-3. We present the detailed evaluation results of the proposed evolving ensemble deep networks and hybrid clustering models in the following section.

IV. EVALUATION

We employ three skin lesion data sets, i.e. PH2 [35], Dermofit Image Library [36], and ISIC 2017 [37], to evaluate the proposed ensemble deep networks and clustering models. A blood cancer data set, i.e. ALL-IDB2 [38], is also employed
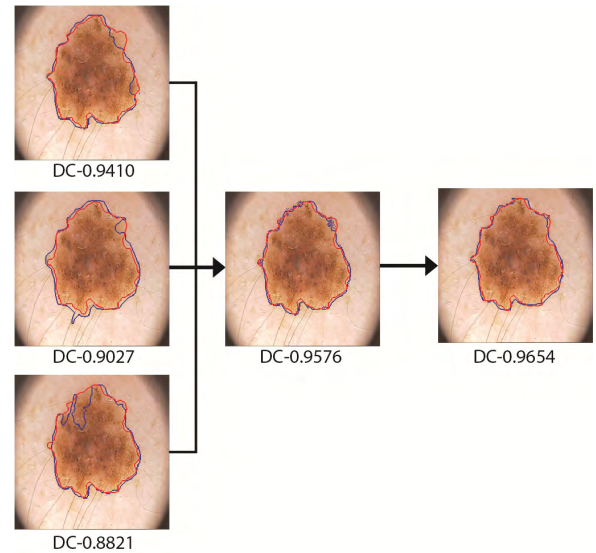

FIGURE 2. Example segmentation results for the proposed deep ensemble network (blue line: predicted mask, red line: GT. From left to right, the results of three base models, the ensemble model output and the final predicted mask after post-processing).

FIGURE 3. Example segmentation results for the proposed clustering ensemble model (blue line: predicted mask, red line: GT. From left to right, the results of three base models, the ensemble model output and the final predicted mask after post-processing).
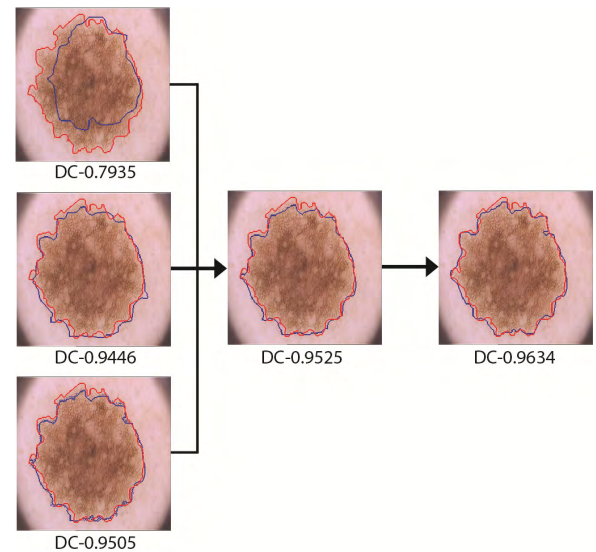
for nuclei segmentation, in order to indicate the robustness of the proposed models. The skin lesion PH2 data set is composed of 200 images with 80 common nevi (benign), 80 atypical nevi, and 40 melanoma cases. The ISIC data set has a total of 2,750 dermoscopic images, while there are 1,300 skin lesion images from 10 classes in the Dermofit Image Library data set. Moreover, a total of 180 WBC sub-images with nucleus and cytoplasm regions are extracted from the ALL-IDB2 data set in the previous study of Srisukkham *et al.* [39].

In our experiments, we employ all the images from the PH2, ISIC 2017 and ALL-IDB2 data sets, as well as the

TABLE 2. Data set description for image segmentation.

Data set	No. of images
PH2	200
ISIC 2017	2,750
Dermofit Image Library	1,300
ALL-IDB2	180

TABLE 3. Parameter settings of each algorithm.

Algorithms	Parameters
PSO [21]	maximum velocity=0.6, inertia weight=1.0, acceleration constants $c_1=c_2=2$
FA [40]	attractiveness =1.0, randomization parameter =0.25, absorption coefficient =1.0
ELPSO [41]	$c_1=c_2=2$, standard deviation of Gaussian mutation=1, scale factor of Cauchy mutation=2, scale factor of DE-based mutation=1.2, inertia weight=0.9
AGPSO [27]	maximum velocity=0.6, adaptive decreasing c_1 and increasing c_2 over generations, inertia weight=$0.9 - (0.9 - 0.4) \times k$/max_gen, where k and max_gen represent the current and maximum iteration numbers, respectively.
GMPSO [42]	maximum velocity = 0.6, inertia weight = 0.9, acceleration constants $c_1=c_2=2$, standard deviation of Gaussian distribution=1, scaling factor of Cauchy distribution =2, crossover probability =0.7, mutation probability =0.3
BBPSOV [39]	Logistic map used as the search parameters
Prop. PSO	maximum velocity=0.6, inertia weight=0.6/0.4/0.2, using random helix coefficients as the search parameters

extraction of a total of 160 images (76 melanoma and 84 benign cases) from the Dermofit data set for performance evaluation. Details of data sets are provided in Table 2. For each data set, we employ a 60-20-20 split to form the training, validation, and test sets, respectively.

To indicate the efficiency of the proposed PSO models, classical optimization methods i.e., PSO [21] and FA [40], as well as advanced PSO variants, i.e. AGPSO [27], BBPSOV [39], ELPSO [41], and GMPSO [42], are implemented for performance comparison. These methods are combined with FCM to conduct hybrid clustering, where the evolutionary algorithms are used to further enhance the cluster centroids identified by the FCM model, respectively. Besides that, single and ensemble conventional KM and FCM clustering models are implemented for performance comparison. Deep single and ensemble networks with default hyper-parameter settings are also developed for performance comparison with the proposed deep networks based on the identified optimal hyper-parameters. In order to increase the base model diversity of the deep networks with default hyper-parameters for ensemble construction, a number of different learning epochs are employed, i.e. 100, 120, and 150, whereas the proposed networks with optimized parameter settings employ a maximum number of 100 epochs consistently.

We employ an image resolution of 128×128 in our experiments, in order to achieve the best trade-off between performance and computational efficiency. Such an image resolution has also been employed by other related studies [6]. Since the proposed and other PSO models are stochastic methods, a set of 10 runs is employed for evaluation of each data set. The mean Dice coefficients over 10 runs are used as the criteria for performance comparison.

Moreover, we employ the same number of function evaluations for all the search methods as the stopping criteria for clustering base model generation. Although the proposed PSO model embeds diverse search strategies, it includes the same number of function evaluations, i.e. population size $\times$ maximum number of iterations, as that of the original PSO model. On the contrary, ELPSO [41] and GMPSO [42] employ additional numbers of function evaluations for the enhancement of the global best solution or the subswarm leader respectively. The iteration numbers or the population sizes of these two search methods are therefore reduced to ensure the conduction of the same number of function evaluations as those of other optimization algorithms.

A. PARAMETER SETTINGS

Table 3 presents the parameter settings of the proposed and other search methods in the experimental studies. We extract

the parameter settings of related methods from their original studies reported in the literature. Since the proposed PSO model employs random search parameters from a helix distribution, it does not require any parameter setting.

B. THE PH2 SKIN LESION DATA SET

We first demonstrate the efficiency of the proposed ensemble deep networks and clustering models using the PH2 data set. This PH2 data set, as well as the ISIC and Dermofit skin lesion data sets, contain diverse lesion segmentation challenges. As an example, each of the abovementioned data sets contains dermoscopic images with dark hairs, which partially occlude the lesion boundaries, and vignettes around illuminated lesion regions that lead to difficult dark shadows in image corners. The lesion images also show a variety of skin pigmentations and sometimes with a low contrast.

As mentioned earlier, since the proposed PSO and other SI algorithms are stochastic methods, we conduct 10 runs of both proposed ensemble models for performance comparison. The mean Dice coefficients over 10 runs of the proposed models and other related methods for the PH2 data set are provided in Table 4. The best Dice coefficient in each column is highlighted in bold.

Table 4 illustrates the Dice coefficient results for both base (i.e. single) and ensemble models. The proposed base CNN model is produced by optimizing the hyper-parameters of the network using the proposed PSO algorithm (row 1), while other clustering base models are generated by combining FCM with the proposed PSO or other search models. Single classical KM and FCM models, and the deep CNN with default hyper-parameter settings are also implemented for comparison. As mentioned earlier, to ensure a fair comparison, for each hybrid clustering base model generation, all

TABLE 4. Evaluation results for the PH2 data set.

PH2	Base model	Ensemble
Deep + Prop. PSO	**0.8187**	**0.8748**
Prop. PSO	0.8168	0.8339
Deep (default)	0.7413	0.7521
FA	0.6501	0.6543
PSO	0.6558	0.6585
AGPSO	0.7465	0.7486
ELPSO	0.6847	0.6896
BBPSOV	0.7485	0.7511
GMPSO	0.7074	0.7132
FCM	0.5839	0.5912
KM	0.5711	0.5881

TABLE 5. Comparison between the proposed research and related studies using the PH2 data set.

Studies	Methodology	Dice score
Ronneberge et al. [5]	U-Net	0.8370
Fernandes et al. [6]	U-Net with Gossip Networks	0.8409
Chen et al. [43]	Dilated-Net	0.8552
Fernandes et al. [6]	Dilated-Net with Gossip Networks	0.8641
Neoh et al. [19]	FCM + GA with inter- and intra-cluster variance measures	0.6854
This research	Ensemble with Prop. PSO-based evolving FCM clustering	0.8339
This research	Ensemble with Prop. PSO-based evolving CNNs	**0.8748**

the search methods employ the same number of function evaluations as the stopping criterion, e.g. population size (50) × maximum number of iterations (20). The optimized hyper-parameter identification of the deep base CNNs employs a fewer number of function evaluations, e.g. population size (15) × maximum number of iterations (10).

We first compare all the base model performances in the first column in Table 4. Among all the methods, the base CNN with optimized hyper-parameter achieves the best Dice result of 0.8187. This is followed by the proposed PSO model combined with FCM clustering with a mean Dice coefficient of 0.8168. Both proposed base models outperform single hybrid clustering methods of PSO, FA, AGPSO, ELPSO, BBPSOV, and GMPSO, respectively. All the hybrid clustering base evaluators outperform the single base KM and FCM significantly. The proposed base CNN model with optimized hyper-parameters also outperforms single networks with default hyper-parameter settings, significantly, owing to the identification of effective learning hyper-parameters during the search process.

Moreover, as indicated in the second column in Table 4, for the ensemble model performance comparison, the proposed deep CNN ensemble model with optimized hyper-parameters achieves the best average Dice coefficient of 0.8748. This is followed closely by the proposed clustering ensemble model with a mean Dice coefficient of 0.8339. Both proposed models outperform all other hybrid ensemble clustering methods and the ensembles with classical KM and FCM methods significantly. The empirical results also indicate that the ensemble deep CNNs with default hyper-parameter settings are composed of base models with limited diversity. On the contrary, our deep ensemble models possess great base model diversity and depict significant performance superiority over the ensemble networks with default settings. In short, both proposed deep and clustering ensemble segmenters outperform all other baseline ensemble models significantly for lesion segmentation.

Table 5 illustrates the comparison details between the proposed research and other related studies reported in the literature for the PH2 data set. All the methods employ the same experimental settings, i.e. 60-20-20 for the training, validation, and test sets, respectively. Note that our results

are obtained by averaging the Dice coefficients over 10 independent runs. As indicated in Table 5, U-Net and Dilated-Net show competitive performance for lesion segmentation. Another two enhanced networks, i.e. U-Net with Gossip Networks and Dilated-Net with Gossip Networks, were proposed by Fernandes *et al.* [6]. Instead of performing segmentation from scratch, both new models further enhanced the segmentation results based on the outputs of U-Net and Dilated-Net, respectively, using Gossip Networks. Therefore, they achieved better performances in comparison with those of U-Net and Dilated-Net. In comparison with these state-of-the-art deep architectures, our proposed deep ensemble networks with optimized learning parameters conduct segmentation from scratch and achieve competitive results. U-Net and Dilated-Net employ skip-layers and dilated kernels to reduce information loss in the encoding-decoding mechanism, whereas the proposed ensemble deep networks employ three base CNNs with great diversity as well as dilated convolutions in each base evaluator to overcome information loss, bias, and variance of single segmenters. Moreover, our proposed clustering ensemble model also outperforms the hybrid image segmentation method in Neoh *et al.* [19], where FCM is integrated with GA, for lesion segmentation.

C. THE ISIC SKIN LESION DATA SET
To indicate efficiency of the proposed ensemble segmenters, we also employ the ISIC 2017 data set for performance comparison. The ISIC 2017 data set shows diverse challenges. Besides the presence of dark hair, dark image corners, low contrast, and a variety of skin pigmentations, the data set also contains lesion images with annotation marks and rulers for lesion scaling.

Table 6 shows the mean Dice coefficients for each base evaluator and ensemble model for the ISIC 2017 data set. The proposed ensemble deep networks and clustering models illustrate impressive performances and outperform all other default ensemble deep networks and hybrid ensemble models significantly. Specifically, the proposed ensemble deep network embeds distinctive evolving base CNN models with different optimized learning parameters to boost the ensemble performance. It achieves a superior mean Dice score

TABLE 6. Evaluation results for the ISIC 2017 data set.

ISIC	Base model	Ensemble
Deep+Prop. PSO	**0.6922**	**0.7677**
Prop. PSO	0.6917	0.7198
Deep (default)	0.6140	0.6986
FA	0.5143	0.5264
PSO	0.5412	0.5575
AGPSO	0.5458	0.5560
ELPSO	0.5086	0.5284
BBPSOV	0.5327	0.5524
GMPSO	0.5132	0.5240
FCM	0.4762	0.4873
KM	0.4517	0.4589

TABLE 7. Comparison between the proposed research and related studies using the ISIC 2017 data set.

Studies	Methodology	Dice score (%)
Ronneberge et al. [5]	U-Net	0.7135
Fernandes et al. [6]	U-Net with Gossip Networks	0.7652
Chen et al. [43]	Dilated-Net	0.7206
Fernandes et al. [6]	Dilated-Net Gossip Networks	0.7611
Neoh et al. [19]	FCM + GA with inter- and intra-cluster variance measures	0.5438
This research	Ensemble with Prop. PSO-based evolving FCM clustering	0.7198
This research	Ensemble with Prop. PSO-based evolving CNNs	**0.7677**

of 0.7677 for foreground background pixel classification, and outperforms the model with default learning settings significantly. The proposed PSO model shows great efficiency in identifying optimal clustering centroids in comparison with other classical and advanced search methods. It leads to the superior performance of the proposed base and ensemble clustering models in comparison with those of other search methods.

The comparison between the proposed models and other state-of-the-art related studies using the ISIC 2017 data set is also illustrated in Table 7. Similar to our experiments, all the related studies employed a total of 2,750 images in the data set, and used 60-20-20 as the training, validation and test data sets. As indicated in Table 7, our proposed ensemble deep networks show better Dice results than those of U-Net [5], Dilated-Net [43], U-Net with Gossip Networks [6] and Dilated-Net with Gossip Networks [6], respectively, owing to the employment of dilated convolutions and optimal hyper-parameter settings in the base networks and the ensemble mechanism by combining complementary base evaluators with great diversity. Our proposed models also outperform the hybrid FCM+GA clustering segmenter of Neoh *et al.* [19], considerably.

D. THE DERMOFIT SKIN LESION DATA SET

The Dermofit skin lesion data set is also employed for evaluation. Table 8 shows the mean Dice coefficients for the base and ensemble models over 10 runs. This data set poses diverse challenges for lesion segmentation including occlusion, illumination changes, annotation marks, and diverse skin tones.

TABLE 8. Evaluation results for the dermofit data set.

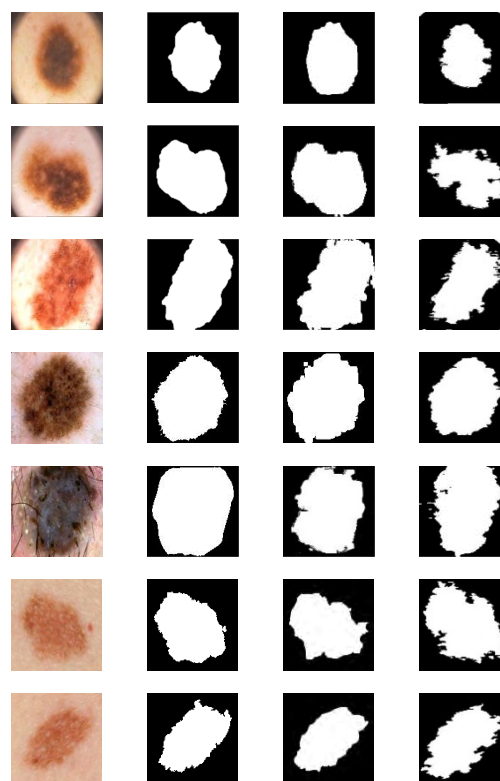| Dermofit	Base model	Ensemble
Deep+Prop. PSO	**0.8278**	**0.8916**
Prop. PSO	0.8012	0.8318
Deep (default)	0.7239	0.7583
FA	0.6556	0.6616
PSO	0.6780	0.6889
AGPSO	0.7544	0.7589
ELPSO	0.7222	0.7269
BBPSOV	0.7649	0.7737
GMPSO	0.7273	0.7323
FCM	0.6434	0.6519
KM	0.6391	0.6417

FIGURE 4. Example lesion segmentation results (from left to right, the input lesion images, the GT masks, the generated masks using the proposed deep ensemble networks and clustering ensemble models, respectively. Row 1-3 from PH2, row 4-5 from ISIC 2017 and row 6-7 from Dermofit).

As indicated in Table 8, the proposed ensemble deep networks and clustering models, with optimized learning parameters and cluster centroids respectively, show superior capabilities in retrieving distinguishing lesion boundaries and outperform all the baseline ensemble methods significantly. A similar case is also observed for the proposed base models with optimized training parameters and cluster centroids. Figure 4 shows the example lesion segmentation results for the PH2, ISIC 2017 and Dermofit data sets.

E. THE ALL-IDB2 DATA SET

In order to indicate the robustness and efficiency of the proposed ensemble models, we conduct segmentation of nuclei

TABLE 9. Evaluation results for the ALL-IDB2 data set.

ALL-IDB2	Base model	Ensemble
Deep+Prop. PSO	**0.8255**	**0.8703**
Prop. PSO	0.8176	0.8490
Deep (default)	0.7474	0.8388
FA	0.7073	0.7094
PSO	0.7083	0.7134
AGPSO	0.7435	0.7528
ELPSO	0.7348	0.7433
BBPSOV	0.7661	0.7757
GMPSO	0.7428	0.7508
FCM	0.6620	0.6726
KM	0.6583	0.6611

TABLE 10. Comparison between the proposed research and related studies using the ALL-IDB2 data set.

Studies	Methodology	Dice score (%)
Neoh et al. [19]	FCM + GA with inter- and intra-cluster variance measures	0.7212
This research	Ensemble with Prop. PSO-based evolving FCM clustering	0.8490
This research	Ensemble with Prop. PSO-based evolving CNNs	**0.8703**

TABLE 11. Statistical test results for the proposed deep ensemble model.

	PH2	ISIC 2017	Dermofit	ALL-IDB2
Deep (default)	8.94E-06	5.54E-06	1.40E-05	9.70E-05
FA	1.17E-04	2.08E-05	9.55E-05	1.26E-05
PSO	8.14E-06	2.87E-05	1.09E-04	5.27E-06
AGPSO	1.20E-04	5.93E-05	3.60E-05	5.54E-05
ELPSO	4.84E-05	5.81E-05	3.41E-05	9.62E-05
BBPSOV	8.24E-05	4.17E-05	1.46E-05	1.62E-05
GMPSO	3.23E-05	1.09E-04	3.98E-05	1.11E-05
FCM	2.76E-05	1.04E-04	5.74E-06	7.47E-05
KM	3.80E-05	1.94E-05	3.59E-05	8.38E-05

TABLE 12. Statistical test results for the proposed clustering ensemble model.

	PH2	ISIC 2017	Dermofit	ALL-IDB2
Deep (default)	1.77E-04	2.03E-04	1.69E-04	1.99E-04
FA	1.63E-04	1.04E-04	1.19E-04	1.57E-04
PSO	1.22E-04	9.03E-05	1.70E-04	8.24E-05
AGPSO	1.68E-04	1.62E-04	9.20E-05	7.31E-05
ELPSO	6.73E-05	7.69E-05	1.45E-04	1.41E-04
BBPSOV	1.51E-04	1.09E-04	6.46E-05	9.38E-05
GMPSO	9.10E-05	1.67E-04	7.33E-05	7.43E-05
FCM	9.48E-05	1.30E-04	7.33E-05	1.71E-04
KM	1.15E-04	1.17E-04	1.57E-04	1.81E-04

regions from the WBCs using the ALL-IDB2 data set. A set of 180 sub-images for acute lymphoblastic leukaemia (ALL) diagnosis from the ALL-IDB2 data set was extracted by Srisukkham *et al.* [39]. These sub-images contain 60 lymphocyte and 120 lymphoblast cases. We conduct the nucleus-background segmentation for a series of 10 runs using the 180 WBC sub-images for each model. These cell images also pose great challenges for nucleus segmentation including irregularity of the nucleus regions, variations in terms of the nucleus to cytoplasm ratio, existence of nucleoli and vacuoles, nucleus and cytoplasm colour, and chromatin patterns [19], [39].

The experimental results in Table 9 indicate that the proposed deep and clustering ensemble models show significant superiority over other baseline ensemble methods. Table 10 shows the comparison with related research. Our proposed single and ensemble segmenters outperform the hybrid FCM+GA clustering model in Neoh *et al.* [19] with inter- and intra-cluster variance measures, significantly. Specifically, our proposed deep and clustering base models outperform the compared method by 10.43% and 9.64%, respectively, while the proposed deep and clustering ensemble models improve the Dice coefficients by 14.91% and 10.46% over those of the compared method, respectively. In comparison with the GA used in the compared method, our proposed hierarchal search mechanism, and the enhanced helix, PSO, and DE search actions with spiral coefficients, account for superiority of the proposed PSO model in retrieving distinguishing nucleus regions.

To further indicate the significance of the proposed ensemble models, statistical tests, i.e. Wilcoxon rank sum tests, have also been conducted. Tables 11-12 show the

statistical test results for the proposed deep and clustering ensemble models respectively in comparison with all the baseline ensemble methods over a series of 10 independent runs pertaining to each test data set. The statistical test results shown in Tables 11-12 are all less than 0.05, which reject the null hypothesis that the baseline ensemble segmenters have similar result distributions to those of the proposed ensemble evaluators. In other words, this indicates that the proposed deep and clustering ensemble models outperform the baseline ensemble segmenters statistically significantly for all the test data sets.

We provide a theoretical comparison with related studies, as follows. ELPSO employs a series of probability distribution strategies to enhance the swarm leader, while AGPSO applies linear and non-linear functions for adaptive acceleration coefficient generation. Evading and attraction search actions are embedded in BBPSOV to increase search diversity, whereas probability distributions and GA are used to enhance sub-swarm leaders in GMPSO. However, the search processes of these related studies are guided by single global best solution (e.g. FA, PSO, AGPSO, BBPSOV and ELPSO) or single subswarm leader (e.g. GMPSO). In addition, the whole population or the subswarm in these methods purely relies on single search operation in each iteration to guide the search process, therefore lack of search diversity and more likely to be trapped in local optima.

On the contrary, the proposed model incorporates SA, helix search, modified PSO and DE operations, as well as random walk and scattering actions. To avoid stagnation, two remote swarm leaders are used to guide the search process of these operations simultaneously. Moreover, random search parameters extracted from a helix distribution are used to

diversify the search process. Such helix-driven search parameters enable each individual to perform local exploitation in a spiral manner around attractions in the modified DE, while they also equip each particle with spiral movement towards the local and global promising solutions in enhanced PSO and helix actions to increase search diversification. Furthermore, these proposed search mechanisms, such as SA, Levy flight, scattering, helix, PSO, and DE search actions, work hierarchically to overcome stagnation. They not only equip each particle with a series of different search actions throughout the search process, but also enable the overall swarm to adopt diverse distinctive search operations in each single iteration. These proposed search mechanisms account for superiority of the proposed PSO model in selection of optimal cluster centroids and hyper-parameters in comparison with other classical and advanced search methods.

The empirical results also indicate that U-Net and Dilated-Net show impressive performances. Besides that, the recent research used additional networks such as Gossip networks [6] or GAN-motivated critic networks [3] on top of U-Net to enhance performance. Their findings indicate that adding such additional networks provides complementary or extra information for the segmenter in comparison with increasing the segmenter network complexity. Motivated by this, we propose the modified PSO model for hyper-parameter fine-tuning to generate diverse optimized CNNs. Our model not only has better chances of finding global optima to retrieve effective learning hyper-parameters, but also provides diverse efficient base networks for deep ensemble pixelwise prediction model construction. These base models serve as complementary pixelwise predictors. They complement each other via the majority voting process to draw the final prediction. The post-processing procedures such as the removal of 'salt-and-pepper' effects and small objects also show great efficiency in further enhancing performance. Moreover, in comparison with other recent studies, where grid-search was employed for optimal parameter selection [1]–[3], the proposed PSO model is used to identify the optimal hyper-parameters for each base segmenter with a significantly reduced computational cost.

V. CONCLUSIONS

In this research, we conduct skin lesion segmentation using PSO-enhanced deep and clustering ensemble models with optimized hyper-parameters and cluster centroids. The proposed PSO model employs SA, Levy distribution, as well as helix, PSO, and DE operations with spiral search parameters to increase diversification. Each particle in the search actions follows two remote swarm leaders simultaneously in a spiral fashion to avoid stagnation. Moreover, the above search operations work in a hierarchical manner to assign distinctive search actions to different particles in each iteration, as well as a series of search actions to one specific particle throughout the search process. The model not only shows impressive performance with respect to hyper-parameter fine-tuning and cluster centroid enhancement for base segmentors,

but also generates diverse base evaluators with great diversity. Deep and clustering ensemble models have subsequently been constructed. Evaluated using PH2, ISIC 2017, Dermofit and ALL-IDB2, they outperform deep networks with default hyper-parameter settings and hybrid clustering methods integrated with diverse PSO variants, significantly. They also show competitive performances in comparison with those of state-of-the-art deep networks in the literature.

In future work, an adaptive learning schedule in deep networks will be explored using the proposed PSO model to enhance performance. The proposed models and architecture optimization with residual connections will also be explored for lesion feature extraction and classification. Furthermore, the proposed PSO model will also be evaluated using other complex computer vision tasks such as evolving deep architecture generation for object detection and localization, and image description generation [1], [2], [26].

REFERENCES

[1] P. Kinghorn, L. Zhang, and L. Shao, "A region-based image caption generator with refined descriptions," *Neurocomputing*, vol. 272, pp. 416–424, Jan. 2018.

[2] P. Kinghorn, L. Zhang, and L. Shao, "A hierarchical and regional deep learning architecture for image description generation," *Pattern Recognit. Lett.*, vol. 119, pp. 77–85, Mar. 2019.

[3] S. Izadi, Z. Mirikharaji, J. Kawahara, and G. Hamarneh, "Generative adversarial networks to segment skin lesions," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 881–884.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[6] K. Fernandes, R. Cruz, and J. S. Cardoso, "Deep image segmentation by quality inference," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[7] S. Vesal, N. Ravikumar, and A. Maier. (2018). "SkinNet: A deep learning framework for skin lesion segmentation." [Online]. Available: https://arxiv.org/abs/1806.09522

[8] S. Vesal, S. M. Patil, N. Ravikumar, and A. K. Maier, "A multi-task framework for skin lesion detection and segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis.* Cham, Switzerland: Springer, 2018, pp. 285–293.

[9] G. M. Venkatesh, Y. G. Naresh, S. Little, and N. E. O'Connor, "A deep residual architecture for skin lesion segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis.* Cham, Switzerland: Springer, 2018, pp. 277–284.

[10] Y. Yuan. (2017). "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks." [Online]. Available: https://arxiv.org/abs/1703.05165

[11] M. Goyal and M. H. Yap. (2017). "Multi-class semantic segmentation of skin lesions via fully convolutional networks." [Online]. Available: https://arxiv.org/abs/1711.10449

[12] A. Aljawawdeh, E. Imraiziq, and A. Aljawawdeh, "Enhanced k-mean using evolutionary algorithms for melanoma detection and segmentation in skin images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 477–483, 2017.

[13] D. Alvarez and M. Iglesias. (2017). "k-Means clustering and ensemble of regressions: An algorithm for the ISIC 2017 skin lesion segmentation challenge." [Online]. Available: https://arxiv.org/abs/1702.07333

[14] S. Louhichi, M. Gzara, and H. B. Abdallah, "Skin lesion segmentation using multiple density clustering algorithm MDCUT and region growing," in *Proc. IEEE/ACIS 17th Int. Conf. Comput. Inf. Sci. (ICIS)*, Singapore, Jun. 2018, pp. 74–79.

[15] T. X. Pham, P. Siarry, and H. Oulhadj, "Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation," *Appl. Soft Comput.*, vol. 65, pp. 230–242, Apr. 2018.

[16] C. Singh and A. Bala, "A transform-based fast fuzzy C-means approach for high brain MRI segmentation accuracy," *Appl. Soft Comput.*, vol. 76, pp. 156–173, Mar. 2019.

[17] F. Zhao, H. Liu, and J. Fan, "A multiobjective spatial fuzzy clustering algorithm for image segmentation," *Appl. Soft Comput.*, vol. 30, pp. 48–57, May 2015.

[18] Y. Pan, Y. Xia, T. Zhou, and M. Fulham, "Cell image segmentation using bacterial foraging optimization," *Appl. Soft Comput.*, vol. 58, pp. 770–782, Sep. 2017.

[19] S. C. Neoh *et al.*, "An intelligent decision support system for leukaemia diagnosis using microscopic blood images," *Sci. Rep.*, vol. 5, p. 14938, Oct. 2015.

[20] B. Dai, X. Wu, and W. Bu, "Optic disc segmentation based on variational model with multiple energies," *Pattern Recognit.*, vol. 64, pp. 226–235, Apr. 2017.

[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4, Nov. 1995, pp. 1942–1948.

[22] 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 Trans. Cybern.*, vol. 47, no. 6, pp. 1496–1509, Jun. 2017.

[23] T. Y. Tan, L. Zhang, S. C. Neoh, and C. P. Lim, "Intelligent skin cancer detection using enhanced particle swarm optimization," *Knowl.-Based Syst.*, vol. 158, pp. 118–135, Oct. 2018.

[24] L. Zhang, K. Mistry, S. C. Neoh, and C. P. Lim, "Intelligent facial emotion recognition using moth-firefly optimization," *Knowl.-Based Syst.*, vol. 111, pp. 248–267, Nov. 2016.

[25] D. Pandit, L. Zhang, S. Chattopadhyay, C. P. Lim, and C. Liu, "A scattering and repulsive swarm intelligence algorithm for solving global optimization problems," *Knowl.-Based Syst.*, vol. 156, pp. 12–42, Sep. 2018.

[26] B. Fielding and L. Zhang, "Evolving image classification architectures with enhanced particle swarm optimisation," *IEEE Access*, vol. 6, pp. 68560–68575, 2018.

[27] S. Mirjalili, A. Lewis, and A. S. Sadiq, "Autonomous particles groups for particle swarm optimization," *Arabian J. Sci. Eng.*, vol. 39, no. 6, pp. 4683–4697, 2014.

[28] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "A Particle swarm optimization-based flexible convolutional autoencoder for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, Dec. 2018. doi: 10.1109/TNNLS.2018.2881143.

[29] Y. Sun, B. Xue, and M. Zhang. (2018). "Automatically evolving CNN architectures based on blocks." [Online]. Available: https://arxiv.org/abs/1810.11875

[30] I. B. Aydilek, "A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems," *Appl. Soft Comput.*, vol. 66, pp. 232–249, May 2018.

[31] F. Ye, "Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data," *PLoS One*, vol. 12, no. 12, 2017, Art. no. e0188746.

[32] F. C. Soon, H. Y. Khaw, J. H. Chuah, and J. Kanesan, "Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 939–946, Oct. 2018.

[33] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[34] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, 2016, pp. 1–13.

[35] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH^2- A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.

[36] 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* (Lecture Notes in Computational Vision and Biomechanics), M. E. Celebi, and G. Schaefer, Eds. New York, NY, USA: Springer, 2013, pp. 63–86.

[37] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[38] R. D. Labati, V. Piuri, and F. Scotti, "ALL-IDB: The acute lymphoblastic leukemia image database for image processing," in *Proc. 18th IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2045–2048. [Online]. Available: http://homes.di.unimi.it/scotti/all/

[39] W. Srisukkham, L. Zhang, S. C. Neoh, S. Todryk, and C. P. Lim, "Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization," *Appl. Soft Comput.*, vol. 56, pp. 405–419, Jul. 2017.

[40] X.-S. Yang, "Firefly algorithm, Lévy flights and global optimization," in *Research and Development in Intelligent Systems*, vol. 26. New York, NY, USA: Springer, 2010, pp. 209–218.

[41] A. R. Jordehi, "Enhanced leader PSO (ELPSO): A new PSO variant for solving global optimisation problems," *Appl. Soft Comput.*, vol. 26, pp. 401–417, Jan. 2015.

[42] 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 Syst. Appl.*, vol. 42, no. 22, pp. 8678–8697, Dec. 2015.

[43] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. (2016). "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." [Online]. Available: https://arxiv.org/abs/1606.00915

TECK YAN TAN received the B.Sc. and M.Sc. degrees in computer science from the University of Northumbria, U.K., where he is currently pursuing the Ph.D. degree. His research interests include computer vision, machine learning, and evolutionary computation.
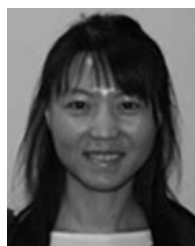
LI ZHANG (M'14) received the Ph.D. degree from the University of Birmingham. She is currently an Associate Professor and a Reader in computer science with the University of Northumbria, U.K., and also serving as an Honorary Research Fellow with the University of Birmingham, U.K.

She holds expertise in computer vision, deep learning, and evolutionary computation. She has served as an Associate Editor for *Decision Support Systems*.

CHEE PENG LIM received the B.Eng. degree in electrical from the University of Technology Malaysia, in 1992, and the M.Sc. degree in control systems and the Ph.D. degree from the University of Sheffield, Sheffield, U.K., in 1993 and 1997, respectively.

He is currently a Professor with Deakin University, Australia. He has published more than 350 technical papers in books, international journals, and conference proceedings. His research interests include soft computing, pattern recognition, medical prognosis and diagnosis, fault detection and diagnosis, and condition monitoring.

BEN FIELDING received the B.Sc. degree in computer science from University of Northumbria, Newcastle upon Tyne, U.K., in 2015, where he is currently pursuing the Ph.D. degree.

His current research interests include deep learning, computer vision, and evolutionary computation.

EMMA ANDERSON is currently a Senior Lecturer in computer science with the University of Northumbria, U.K. Her research interests include image processing, machine learning, and bioinformatics.

• • •

YONGHONG YU received the B.Sc. and M.Sc. degrees in computer science from Wuhan University, and the Ph.D. degree in computer application technology from Nanjing University. He is currently an Associate Professor with the Tongda College, Nanjing University of Posts and Telecommunications. His current research interests include machine learning, recommender systems, and image processing.