Northumbria Research Link

Citation: Zhang, Li and Lim, Chee Peng (2020) Intelligent optic disc segmentation using improved particle swarm optimization and evolving ensemble models. Applied Soft Computing, 92. p. 106328. ISSN 1568-4946

Published by: Elsevier

URL: https://doi.org/10.1016/j.asoc.2020.106328 <https://doi.org/10.1016/j.asoc.2020.106328>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/42920/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

CUSTOMER SERVICE EXCELLENCE CSE UniversityLibrary

Intelligent Optic Disc Segmentation Using Improved Particle Swarm Optimization and Evolving Ensemble Models

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, UK, NE1 8ST

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

Email: **li.zhang@northumbria.ac.uk; chee.lim@deakin.edu.au**

Abstract.
In this research, we propose Particle Swarm Optimization (PSO)-enhanced ensemble deep neural networks for optic disc (OD) segmentation using retinal images. An improved PSO algorithm with six search mechanisms to diversify the search process is introduced. It consists of an accelerated super-ellipse action, a refined super-ellipse operation, a modified PSO operation, a random leader-based search operation, an average leader-based search operation and a spherical random walk mechanism for swarm leader enhancement. Owing to the superior segmentation capabilities of Mask R-CNN, transfer learning with a PSO-based hyper-parameter identification method is employed to generate the fine-tuned segmenters for OD segmentation. Specifically, we optimize the learning parameters, which include the learning rate and momentum of the transfer learning process, using the proposed PSO algorithm. To overcome the bias of single networks, an ensemble segmentation model is constructed. It incorporates the results of distinctive base segmenters using a pixel-level majority voting mechanism to generate the final segmentation outcome. The proposed ensemble network is evaluated using the Messidor and Drions data sets and is found to significantly outperform other deep ensemble networks and hybrid ensemble clustering models that are incorporated with both the original and state-of-the-art PSO variants. Additionally, the proposed method statistically outperforms existing studies on OD segmentation and other search methods for solving diverse unimodal and multimodal benchmark optimization functions and the detection of Diabetic Macular Edema.

Keywords: Image Segmentation, Particle Swarm Optimization, Evolutionary Algorithm, Convolutional Neural Network and Ensemble Segmentation Model

1. INTRODUCTION

Diabetic retinopathy (DR) and glaucoma are two common ocular diseases that affect millions of people worldwide. DR, which is the most common diabetic eye disorder, is caused by the damage to the blood vessels of light-sensitive tissue in the retina. Glaucoma, which is another complicated ocular disorder, affects people of all ages. It is caused by the damage to the optic nerve connecting the eye to the brain. Both DR and glaucoma are silent diseases, and are the leading causes of irreversible blindness worldwide [1]. Therefore, early and instant detection of their onset is vital. The optic disc (OD) is a bright yellow, approximately circular object in the retina where the major blood vessels emerge. Serving as a reference location for other features in fundus images, it provides important information for the estimation of fovea locations, blood vessel segmentation, and cup-to-disc ratio calculation. These characteristics, especially the cup-to-disc ratio, play a significant role in glaucoma diagnosis. Since the disc is also similar to bright lesions that occur in retinal diseases such as DR, it is important to segment the OD regions in the fundus images before performing disease grading.

Owing to the low-contrast blurry disc boundaries, vasculature occlusion, retinal atrophy, and bright lesion distractors, determining the disc boundaries is a challenging task. Many segmentation methods have been proposed, which can be classified into two categories, i.e. shape and texture-based approaches. Shape-based approaches include template-based, morphological, and deformable methods. Cheng et al. [2] conducted OD segmentation by eliminating peripapillary atrophy using edge detection, constraint elliptical Hough transform, peripapillary atrophy detection, and ellipse fitting correction. Morales et al. [3] employed the Principal Component Analysis (PCA) and mathematical morphological operations for determining OD contours. Sarathi et al. [4] utilized adaptive threshold based region growing techniques to identify the disc boundaries. Deformable models such as the Active Shape Model were adopted by Yin et al. [5] for segmenting the disc regions. Since these methods rely heavily on the brightness and the circular or elliptical characteristics of the disc regions, they either depict limited capabilities in tackling shape variations of the disc boundaries, or could be easily distracted by bright lesions in abnormal images or the optic cup (OC), a brighter circular region in the centre of the disc [6, 7]. They also depend strongly on pre-processing procedures such as OD localization, blood vessel removal, and contrast enhancement to enable OD segmentation.

On the other hand, texture-based approaches perform pixel-wise classification of OD and non-OD regions, including filter-based and clustering-based techniques. Dashtbozorg et al. [8] employed a multiresolution sliding band filter (SBF), including both a low-resolution SBF and a high-resolution SBF, to determine the OD boundaries. Xue et al. [9] extracted OD regions using K-Means (KM) clustering and subsequently applied ellipse fitting and active contour to retrieve refined OD contours. KM clustering and Connected Component Analysis were employed by Sharma and Verma [10] to segment the disc regions, after eliminating noise and enhancing image contrast using colour band selection and spatial average filtering. Hamednejad and Pourghassem [11] employed the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm for OD segmentation. They first employed a masking procedure and combinational filtering to retrieve the initial region of interest (ROI) sub-images, which were subsequently converted to the LAB colour space before passing on to the DBSCAN model for OD segmentation. Post-processing operations such as median filter and morphological operations were also used to further refine the segmented regions. Thakur and Juneja [12] proposed a hybrid model, i.e. the Level Set Based Adaptively Regularized Kernel-Based Intuitionistic Fuzzy C-Means (LARKIFCM) clustering model, for OD segmentation. A comparison between LARKIFCM and KM clustering, Fuzzy C-Means (FCM) Clustering, Intuitionistic FCM clustering, Spatial FCM clustering, as well as Adaptively Regularized Kernel-Based FCM (ARKFCM) clustering was conducted, in order to validate the model efficiency. Although the abovementioned filter-based and clustering-based approaches show enhanced performance, such techniques require pre-processing procedures, such as region cropping and contrast enhancement, to eliminate noise and non-OD structures, as well as post-processing operations such as active contour, level set, and morphological operations, to tackle the over and/or under segmentation problems. Besides that, the abovementioned clustering-based methods are constrained by the selected clustering models such as KM and FCM, which are likely to be trapped in local optima [13, 14, 15]. Similar to shape-based approaches, the abovementioned texture-based methods are also sensitive to distractors, such as blood vessels, the occurrence of bright lesions, as well as fuzzy and blurry OD boundaries.

Deep learning and transfer learning techniques have shown superior performances in addressing computer vision and medical imaging problems [16, 17, 18, 19, 20]. In comparison with the traditional shape and texture-based methods for OD segmentation, they do not require multi-stage pre- and post-processing operations. They perform the segmentation task in an end-to-end fashion. In addition, transfer learning is able to overcome data scarcity issues and yield efficient learners with a small data set from an alternative target domain. It transfers the learned features to a new task through training with a set of pre-trained weights, instead of randomly assigned ones. In this regard, Mask R-CNN [16, 17] is one of the state-of-the-art segmentation methods with superior performance in instance segmentation for the 2017 COCO suite of challenges. We adopt transfer learning based on Mask R-CNN for OD segmentation in this research, which employs a one-stage, end-to-end deep architecture with image-level inputs to overcome the constraints of multi-stage pipeline processing employed in the existing methods. Since the hyper-parameters of transfer learning, such as the learning rate and momentum, have a significant impact on the network performance, which is a bottleneck for deploying deep learning models to a new domain, we propose a variant of the Particle Swarm Optimization (PSO) algorithm to automate the selection of optimal hyper-parameters. This PSO-based hyper-parameter selection process also enables the generation of multiple distinctive fine-tuned segmenters with different optimal training hyper-parameters. To overcome bias of single segmentation models, we combine the individual segmenters to form a deep ensemble model for performance enhancement. Figure 1 depicts the overall system architecture.

The contributions of this research are highlighted, as follows.

1. Transfer learning based on the state-of-the-art object detector and segmenter, Mask-RCNN [16, 17], is used to generate the base segmenters. It is able to overcome the data scarcity issue by transferring the learned features to a new task using a smaller number of training images, in comparison with the process of training from scratch.
2. Identification of the optimal training configuration is a bottleneck in deploying deep learning models for a new domain, owing to the fact that hyper-parameters such as the learning rate and momentum have a significant impact on the network performance [15, 21]. Manual trial-and-error and exhaustive grid search procedures have been adopted in the existing methods for the selection of optimal training settings. In this research, we automate the optimal hyper-parameter identification process using a PSO variant, leading to performance enhancement.
3. Since the original PSO algorithm employs a monotonous search process guided by one leader, it is likely to converge prematurely. The proposed PSO model overcomes this shortcoming by employing a versatile search procedure guided by multiple elite solutions. Specifically, it encompasses super-ellipse search operations, a new velocity updating action with elliptical coefficients, mean and random best leader-enhanced search operations, and a circular movement for global best enhancement. The model is used to optimize the training hyper-parameters, e.g. learning rate and momentum, of the transfer learning procedure based on Mask R-CNN, in order to generate fine-tuned segmenters.
4. To overcome the bias issue of single segmenters, an ensemble model is constructed. It incorporates the results of a number of distinctive base models using a pixel-level majority voting strategy to yield the final segmentation outcome. Since the base models are trained with distinctive optimized learning parameters, they possess great diversity to enhance the performance of the ensemble model.
5. Evaluated using two public retinal image data sets, i.e. Drions and Messidor, the proposed ensemble model outperforms other deep ensemble networks and hybrid ensemble clustering models integrated with the original and advanced PSO variants, as well as other state-of-the-art studies on OD segmentation, significantly. The proposed PSO algorithm also depicts great superiority over other search methods in solving diverse benchmark optimization functions, and the detection of Diabetic Macular Edema (DME).

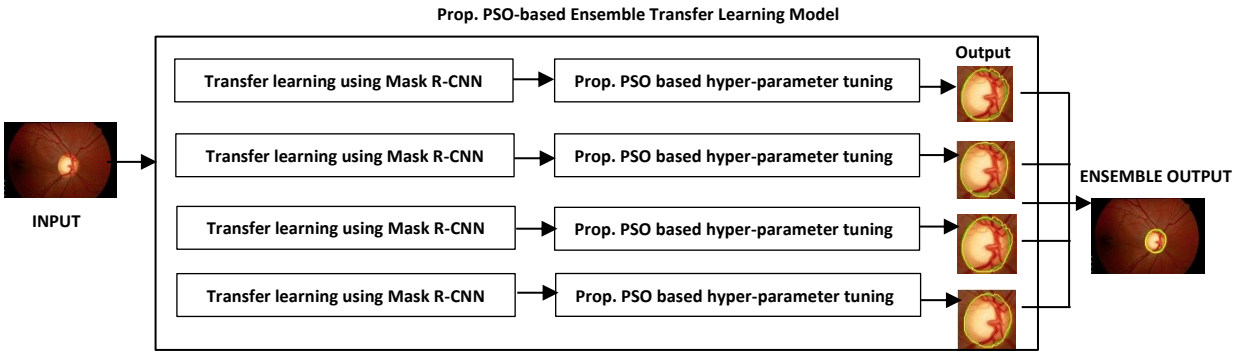


Figure 1 The overall system architecture

The paper is organized as follows. The related studies on image segmentation using fundus images and state-of-the-art PSO variants are described in Section 2. We present the proposed PSO model and the ensemble transfer learning network with automatic hyper-parameter identification in Sections 3 and 4, respectively. A comprehensive evaluation of the proposed ensemble segmenter is presented in Section 5. Conclusions and suggestions for further research are given in Section 6.

2. RELATED WORK

In this section, we discuss the state-of-the-art related studies on OD segmentation and other medical imaging problems using retinal fundus images. We also review recently proposed diverse PSO variants in the literature.

2.1 Image Segmentation Using Retinal Fundus Images

Zhang et al. [22] conducted automatic OD localization and segmentation using Faster R-CNN and a shape constrained level set algorithm. Their method firstly performed OD localization using Faster R-CNN integrated with a ZF net. The Hessian matrix method was used to remove the blood vessels in the detected bounding boxes. OD segmentation was subsequently carried out by using a shape constrained level set mechanism. Based on a total of 120 images from the Messidor data set, their model achieved an average precision score of 99.9% for OD localization and an average Intersection over Union (IoU) score of 85.4% for OD segmentation. Dai et al.

[7] proposed a variational model incorporated with different energy measures for determining the OD boundary. Disc localization was firstly conducted using sparse coding and a circular Hough transform. Based on the detected disc centre, the ROI-based sub-image was extracted. Morphological closing and opening operations were subsequently applied to remove the blood vessels from the sub-image. Then, multiple types of energy, including the phase-based boundary energy, the PCA-based shape energy, and the region energy, were jointly minimized to yield the disc contours. Their model yielded an impressive performance in comparison with those of existing studies when evaluated using the Messidor, ONHSD, and Drions data sets.

Rehman et al. [23] proposed statistical and textural feature classification for OD segmentation, in an attempt to tackle the challenges of deformable and irregular shapes of the disc regions. A number of pre-processing operations, such as bilateral filtering, image localization and cropping, colour channel selection and histogram matching, were applied to highlight the disc regions and eliminate the vascular structures. A Simple Linear Iterative Clustering algorithm was then used to segment the ROI into super-pixels. Subsequently, a set of 17 statistical and textural attributes, i.e. 5 intensity, 6 texton-map histogram, and 6 fractal features, pertaining to each super-pixel, was extracted. Minimum redundancy-based feature selection was also performed. Super-pixel-based OD segmentation was implemented using several classification algorithms, such as Random Forest, Support Vector Machine (SVM), AdaBoost, and RusBoost. Evaluated with several fundus image data sets, Random Forest depicted an improved disc boundary classification performance in comparison with those of related studies. Sarathi et al. [4] developed a pipeline processing method for OD localization and segmentation. A double windowing-based method was firstly used to locate the disc as well as the disc centre. Then, an inpainting algorithm was proposed to eliminate the blood vessel structures. The detected disc centre was used as the initial seed of an adaptive threshold-based region growing method for segmentation of disc boundaries. Post-processing operations such as ellipse fitting and spline interpolation were also performed. Their method achieved a superior performance when evaluated with the Messidor, Drive and private databases. Abed et al. [24] employed multi-stage pre-processing operations and meta-heuristic search algorithms for OD detection. Several operations, such as rescaling and conversion to grayscale, median filtering, adaptive histogram equalization, background subtraction, multiple filters and masking processes, were adopted as the pre-processing procedures to remove false peaks and smooth the images. Search methods such as PSO, Cuckoo Search (CS), Firefly Algorithm (FA), Artificial Bee Colony (ABC), and Bat Algorithm (BA) were used to identify the maximum grayscale value in the image where the OD resides. The FA model with optimal parameter settings integrated with the above pre-processing procedures yielded the best OD detection performance for several fundus image data sets.

Thakur and Juneja [12] conducted OD segmentation by using a hybrid model, i.e. the LARKIFCM clustering model. It integrated a level set algorithm with Adaptively Regularized Kernel-Based Intuitionistic FCM (ARKIFCM) clustering. The ARKIFCM clustering model was firstly used to obtain the initial OD boundaries, where a new objective function was developed based on those of ARKFCM and Intuitionistic FCM. The level set algorithm was then used to tackle both the under- or over-segmentation problems of ARKIFCM clustering. Evaluated using several publicly available data sets, their model achieved an enhanced performance as compared with those of other clustering techniques. Wang et al. [25] proposed a deep learning architecture on top of a U-Net for OD segmentation. Their model conducted OD segmentation using not only the fundus images but also the vessel density map derived from the retinal images, respectively. An overlapping mechanism was used to identify the most likely OD region from the output candidate regions from the above two segmentation processes. The U-Net model with the recommended OD regions as the inputs was used to generate the final segmentation outputs. The model achieved an improved performance as compared with those of U-Net and other related studies, pertaining to the evaluation of a private database and six publicly available data sets. Tan et al. [26] proposed a 7-layer Convolutional Neural Network (CNN) model for segmentation of OD as well as fovea and blood vessels, simultaneously. After background normalization, their model categorized each pixel into any of the four classes, i.e. OD, fovea, retinal vasculature, and background. The network achieved an impressive segmentation performance for the DRIVE database.

Li et al. [6] performed OD segmentation based on learning of a sequence of supervised descent directions between OD landmarks and their appearance. They firstly conducted OD localization and detection using a brightness based method. Landmarks were derived and used as the OD shape, while histograms of the oriented gradients (HOG) features were subsequently extracted as the OD region. Segmentation was conducted by learning a function to establish the connection between the OD landmarks and their appearance, i.e. by moving the initial OD landmarks to the optimal positions. Evaluated using Origa, MESSIDOR, three releases of RIM-ONE, and Drions data sets, their model achieved a superior performance in OD segmentation. Zahoor and Fraz [27] utilized a polar transform-based adaptive thresholding algorithm to segment OD boundaries. They initially employed circular Hough transform and morphological operations to locate the disc and eliminate blood vessels,

respectively. After cropping the ROI sub-image, the polar transform operation was applied to the ROI to convert it into distinct rectangular tiles, which were subsequently thresholded to retrieve the disc contour. The model was evaluated using the Messidor, Drions, DIRETDB1, HRF, Drive and RIM-ONE databases, and achieved a promising performance.

An enhanced U-Net model incorporated with adversarial training was proposed by Liu et al. [28] for OD segmentation. The enhanced U-Net was the main segmentation network. The training process was conducted by integrating an adversarial network with the enhanced U-Net to improve the higher-order consistencies between the ground truth and the segmented masks yielded by U-Net. Moreover, the enhanced U-Net used strided convolutions, instead of max-pooling layers, as well as inserted a batch normalization layer after each convolution layer to accelerate convergence. The method improved the performance on the DRISHTI-GS and RIM-ONEv3 databases. Guo et al. [29] employed a fractional-order Darwinian PSO (FODPSO) algorithm for OD segmentation. Morphological operation and median filtering were initially applied to eliminate the vascular structures. Disc segmentation was conducted using FODPSO by extracting comparatively brighter regions from the fundus images. Canny edge detection and Least-squares ellipse fitting were subsequently applied to determine the OD regions. The FODPSO-based segmentation and post-localization processes were conducted for a number of iterations to refine the segmented disc contours. Several fundus image databases, such as Drion, Messidor, and ORIGA, were employed to evaluate the model efficiency. Lim et al. [30] implemented a 9-layer CNN model with exaggerated feature inputs for OD and OC segmentation. The segmentation results were then used to calculate the cup-to-disc ratio, in order to facilitate early diagnosis of glaucoma. Firstly, the Daubechies wavelet transform method was employed to detect the region containing the disc. The most visually-relevant features were subsequently extracted by removing blood vessels and recovering the locations of vessel kinks. Next, the CNN model was used to classify each pixel into three classes, i.e. the background, rim, and cup classes, in order to produce a pixel-level probability map. The final disc and cup boundaries were obtained by segmenting this probability map. The method achieved mean error rates of 0.112 and 0.0843 for the Messidor and Seed-DB data sets, respectively, for OD segmentation.

2.2 PSO Variants

As one of the popular swarm intelligence (SI) algorithms, PSO [31] has been used in solving diverse mathematical, engineering, design, network, robotic, and image processing optimization problems. In PSO, the particles explore the search space by following the personal and global best experiences, as defined in Equations (1)-(2).

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

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

where x_{id}^{t+1} and v_{id}^{t+1} denote the position and velocity of particle x_i in the d^{th} dimension of the $(t+1)^{th}$ iteration, respectively. The acceleration coefficients, c_1 and c_2, play a significant role in balancing the search between social and cognitive components, while r_1 and r_2 are random vectors with each element having a value between 0 and 1. The personal and global best experiences are denoted as p_i and p_g, respectively. The inertial weight, w, determines the impact of the previous velocity on the current one. Since the search process is led by only one leader, the PSO operation is more likely to be trapped in local optima.

Diverse state-of-the-art PSO variants have been proposed to overcome the limitations of the original model. They are used to solve a variety of single-, multi-, and many-objective optimization problems. Tan et al. [15] proposed a hybrid learning PSO model, namely HLPSO, for skin lesion segmentation and classification. It incorporated local exploitation based on Gaussian, Cauchy, and Levy distributions, and global exploration based on variants of the FA model and a spiral search action to guide the search process. Crossover and mutation operations were also used to enhance search diversification. The HLPSO model integrated with KM clustering was used for skin lesion segmentation, where the intra- and inter-cluster variances were embedded into the objective function. Subsequently, two models were devised for melanoma classification. The first lesion classification strategy employed the HLPSO-based feature selection and ensemble models while the second classification mechanism adopted the evolving CNN architectures with optimized topologies and hyper-parameters devised by HLPSO. The hybrid clustering model and deep networks showed superior capabilities in lesion segmentation and classification, and outperformed a number of state-of-the-art models based on the ISIC 2017 data set and a mixed data set. He et al. [32] proposed a modified PSO model with a cooperative strategy and a damping factor to accelerate convergence and avoid stagnation. The damping factor was used to determine the impact of each particle's previous position with respect to its new position in the next iteration. Besides the global best solution, a neighbourhood best solution was identified with the search neighbourhood established by FCM clustering. Both such local and global best solutions were then used for updating the position of the

current particle. A fitter offspring solution from the above processes was inherited to the next generation. To accelerate convergence, the weakest particle in the swarm was identified and re-initialized in each iteration. The PSO model showed an enhanced performance as compared with those from other PSO variants when evaluated using 24 benchmark functions. An adaptive learning PSO algorithm, namely ALPSO, was proposed by Tan et al. [21]. ALPSO was incorporated with ensemble clustering and deep CNN models for skin lesion segmentation. It included several search mechanisms, such as a helix search action and modified Differential Evolution (DE) and PSO with random coefficients, to diversify the search process. The model operated in a cascade manner by assigning a different search operation to each swarm particle in any single iteration. It was used to optimize the hyper-parameters of the base CNN segmenters and the centroids of the base FCM models for lesion segmentation. Evaluated with several dermoscopic skin lesion data sets and a microscopic blood cancer database, the model outperformed the U-Net, Dilated-Net and other hybrid ensemble clustering models, significantly.

Xu et al. [33] proposed a PSO model, namely two-swarm learning PSO (TSLPSO), for solving diverse challenging benchmark functions. A dimensional learning mechanism was firstly proposed. A dimension-by-dimension learning mechanism using the personal best solution and the swarm leader was devised to yield a new offspring particle to lead the search process. Two sub-swarms were subsequently generated, in which the search process of one sub-swarm was led by the above dimensional learning mechanism while the other was guided by an existing comprehensive learning strategy, in order to balance between intensification and diversification. The model showed impressive capabilities in solving diverse shifted and rotated numerical optimization problems. Chen et al. [34] proposed a hybrid PSO model integrating DE-based global exploration and Quasi-Newton-based local exploitation. Dynamic sub-swarms were employed to diversify the search processes. The model also adopted a new velocity updating strategy. Instead of using the personal and global best solutions, it employed purely the neighbourhood best experience of each particle for exploration of the search space. The neighbourhood leader was further enhanced by the DE operation to overcome stagnation. Local exploitation using the quasi-Newton method was subsequently conducted for the top 25% of the neighbourhood best leaders. Tested with 13 basic and 28 CEC2013 benchmark functions, the PSO model depicted great efficiency in solving diverse challenging numerical optimization problems.

A PSO variant with self-adapt control parameters was proposed by Isiet and Gadala [35]. An evolutionary state factor was firstly calculated for each particle by taking into account the fitness difference between the swarm leader and the personal best experience as well as the fitness of the current particle. This parameter was then employed to influence the generation of the linear adaptive acceleration coefficients, in order to achieve a better trade-off between local exploitation and global exploration. The velocity updating operation was modified by inserting a repulsion component to increase diversity and robustness. Sub-dimensions of the particle velocity were then re-initialized to diversify the search when these elements reached zero velocity scores. Tested with benchmark constrained optimization problems and a real-world design problem, the PSO model showed great superiority over other state-of-the-art PSO variants. Fielding et al. [36] proposed another PSO variant with adaptive cosine coefficients, namely CPSO, for evolving deep architecture generation. Cosine annealing operations were proposed to produce ascending and descending acceleration coefficients. Such characteristics enabled the search process to focus on global exploration during the early iterations and local exploitation during the final iterations. The model was then used to optimize the quantities of different types of block architectures for deep network generation. They subsequently employed two different strategies for deep ensemble model generation. The first was built by combining the networks recommended by all the personal best experiences, while the second was constructed by integrating distinctive best-performed individual block architectures recommended by a historical weight sharing look-up table. Evaluated using the CIFAR-10 data set, the best single deep network recommended by the global best solution using the adaptive PSO model achieved an error rate of 4.78%. The ensemble deep networks constructed by the local best experiences and look-up table nominations achieved better error rates of 4.39% and 4.27%, respectively. Other state-of-the-art SI algorithms [13-15, 20, 37-45] were also proposed, which showed impressive performances in solving diverse other classification and clustering optimization problems, such as discriminative feature selection, evolving deep architecture generation, clustering centroid enhancement, and classifier ensemble reduction.

3. THE PROPOSED PSO ALGORITHM

In this research, we automate hyper-parameter selection using an enhanced PSO model in the transfer learning process for OD segmentation. The enhanced PSO model comprises super-ellipse search actions, a new operation with elliptical acceleration coefficients, search operations driven by random and mean best individuals, and a spherical random walk mechanism for swarm leader enhancement. Specifically, it employs accelerated and refined super-ellipse search operations to balance between local exploitation and global exploration. Multiple global optimal solutions are used to diversify the search process. Spiral coefficients and two randomly recruited

particles are used to improve the swarm leader and overcome stagnation. The model is capable of assigning distinctive search actions to different particles in any single iteration. We introduce each key search operation in the following sub-sections.

Algorithm 1 – The Proposed PSO Algorithm	
1	**Start**
2	Initialize a particle swarm randomly;
3	Evaluate all the particles in the swarm;
4	**For** (each iteration **do**) {
5	Find the global best solution, g_{best};
6	**For** (each particle **do**) {
7	//Assign distinctive actions to different particles in each iteration
8	Generate a random value, p;
9	**If** ($p < 0.2$)
10	Randomly select one search parameter among those generated by Equations (3)-(7);
11	Conduct the accelerated search operation as defined in Equation (8) with the selected elliptical coefficient;
12	**Else If** ($p < 0.4$)
13	Randomly select one search parameter among those generated by Equations (3)-(6) & (9);
14	Conduct the refined search operation as defined in Equation (10) with the selected elliptical coefficient;
15	**Else If** ($p < 0.6$)
16	Randomly select two search parameters among those generated by Equations (3)-(6) & (9);
17	Conduct the modified PSO operation as defined in Equations (11) & (2) with the selected elliptical coefficients;
18	**Else If** ($p < 0.8$)
19	Randomly select one search parameter among those generated by Equations (3)-(6) & (9);
20	Conduct the search operation using the average best leader as defined in Equation (12) with the selected elliptical coefficient;
21	**Else**
22	Randomly select one search parameter among those generated by Equations (3)-(6) & (9);
23	Conduct the search operation using a random leader as defined in Equation (13) with the selected elliptical coefficient;
24	**End IF**
25	} **End For**
26	Conduct the fitness evaluation;
27	Update the personal best p_{best} and global best g_{best};
28	Randomly select one search parameter among those generated by Equations (3)-(6) & (14);
29	Generate an offspring solution g'_{best} of the global best solution g_{best} using the search operation defined in Equation (15) and the selected elliptical coefficient;
30	Replace g_{best} with the new offspring solution g'_{best} if the new solution has a better fitness score;
31	} **End For**
32	Output g_{best};
33	**End**

A number of search mechanisms are introduced to diversify the search process, i.e. an accelerated super-ellipse action, a refined super-ellipse operation, a modified PSO strategy, random leader as well as average leader enhanced search operations. After initializing a particle swarm, we randomly select any of the above search operations for position updating. These search actions adopt elliptical search coefficients generated by two instantiations of a super-ellipse formula to balance between intensification and diversification. The elliptical shapes generated by these two instantiations simulate spider networks with irregular boundary patterns. Randomly selected values from the generated shapes are assigned as the search coefficients in the above operations, leading the particles to move towards the local and global best experiences in an irregular spiral manner. Specifically, one instantiation of the employed formula produces comparatively larger search parameters to accelerate convergence, while the other yields comparatively smaller but more irregular search coefficients to enhance search diversification. After a number of iterations, the swarm leader, g_{best}, is identified. An offspring solution, g'_{best}, of the global best solution is subsequently produced using a local search mechanism, which incorporates two randomly recruited particles and the circular search parameters. The

offspring solution, g'_{best}, is used to replace g_{best} if it has a better fitness score. Overall, the proposed model is able to equip each swarm particle with a distinctive search action in any single iteration to increase search diversity. The pseudo-codes of the proposed PSO model are illustrated in Algorithm 1.

3.1 The Accelerated Elliptical Search Operation

Before explaining the proposed search operations, we introduce a super-ellipse formula which lays the foundation of the proposed mechanisms. Proposed by Gielis [46], Equations (3)-(4) define the super-ellipse formula, which is a generalization of the super-ellipse. This super-ellipse formula offers significant capabilities in simulating many complex shapes and curves found in nature. As an example, different settings of parameters, m, k, n_1, n_2, n_3 and n_4, in Equations (3)-(4) produce different curves with distinctive shapes. This super-ellipse formula is employed to simulate irregular reticulated networks with different sizes and boundary patterns and to drive distinctive local and global search operations.

Motivated by the super-ellipse formula, we propose an accelerated super-ellipse search action. The search coefficient of this operation is produced by an instantiation of the super-ellipse formula defined in Equations (3)-(7). The super-ellipse formula defined in Equations (3)-(7) simulates a simplified, and comparatively wider, spider network, as illustrated in Figure 2. This search action leads the particles to move towards the swarm leader in an irregular accelerated reticulated manner.

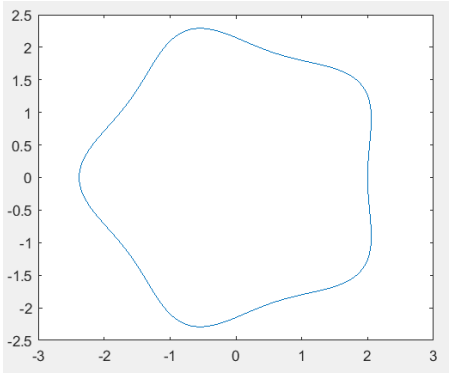


Figure 2 A comparatively wider spider network generated by Equations (3)-(7)

Firstly, Equations (3)-(6) are used to generate a set of 3,142 unique points to represent the overall elliptical contour, as illustrated in Figure 2. Specifically, there are 3,142 positive and negative values in the range of [-2.3784, 2.055] to form the curve in the x axis. We randomly select one of the generated 3,142 unique values in the x axis as the search coefficient. Owing to the wider span of the yielded scores in the x axis, i.e. [-2.3784, 2.055], this instantiation of the super-ellipse formula provides better chances of employing a comparatively larger and accelerated search coefficient, in comparison with that produced by a uniform distribution within the range of [0, 1], to accelerate convergence.

$$\tau(\vartheta) = \left(\left|\frac{\cos\left(\frac{n_1\vartheta}{4}\right)}{m}\right|^{n_3} + \left|\frac{\sin\left(\frac{n_1\vartheta}{4}\right)}{k}\right|^{n_4}\right) \quad \vartheta = [0: 0.002: 2\pi] \tag{3}$$

$$\alpha = \tau^{-\frac{1}{n_2}} \tag{4}$$

$$x = \alpha \times \cos(\vartheta) \tag{5}$$

$$y = \alpha \times \sin(\vartheta) \tag{6}$$

$$m = k = 2, n_1 = 5, \text{ and } n_2 = n_3 = n_4 = 4 \tag{7}$$

In Equation (3), τ and ϑ represent the radius and angle, respectively, with x and y denoting the generated points in the x and y coordinates using Equations (5) and (6), respectively. The starting and ending angles of ϑ are 0 and 2π, with an updating angle of 0.002.

The selected search parameter is subsequently used to guide the current particle to move towards the global best solution, as defined in Equation (8).

$$x_{id}^{t+1} = x_{id}^{t} + s_1 \times (g_{best} - x_{id}^{t}) \tag{8}$$

where s_1 is a randomly selected score among the generated 3,142 values using Equations (3)-(5). This search action enables the particles to follow the best leader with an accelerated search step. It also assigns both positive and negative coefficients to increase search exploration.

3.2 The Refined Elliptical Search Operation

Next, a refined super-ellipse search action is proposed, which employs a comparatively smaller but more irregular search coefficient. Another instantiation of the super-ellipse formula as defined in Equations (3)-(6) with the parameters provided in Equation (9) is used to generate a new elliptical curve with comparatively more irregular borders as the source for coefficient generation.

$$m = k = 1, n_1 = 16, n_2 = n_3 = 0.5, and\ n_4 = 16 \tag{9}$$

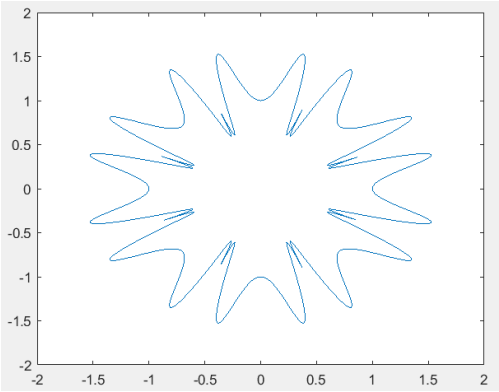


Figure 3 A comparatively more irregular border pattern generated by Equations (3)-(6) and (9)

Figure 3 illustrates the generated elliptical flight using Equations (3)-(6) and (9), which is composed of a total of 3,142 unique points. A new set of 3,142 positive and negative values in the x axis in the range of [-1.5253, 1.5253] is used as the source for coefficient generation. We randomly select one of these 3,142 values, and assign it as the search coefficient in Equation (10) to perform position updating in the search space.

$$x_{id}^{t+1} = x_{id}^{t} + s_2 \times (g_{best} - x_{id}^{t}) \tag{10}$$

where s_2 is a randomly selected score from the 3,142 values in the x axis. This operation enhances the search diversity through the assignments of both positive and negative search coefficients. As indicated in Figure 3, this search action empowers the particles to follow the best leader using a refined step along a more irregular ellipse shape.

3.3 The Modified PSO Operation Using Super-Ellipse Coefficients

A modified PSO operation is also proposed by assigning the super-ellipse acceleration coefficients produced by Equations (3)-(6) and (9). Specifically, two randomly selected scores among the generated 3,142 elements within [-1.5253, 1.5253] in the x axis in Figure 3, as discussed in Section 3.2, are used as the acceleration coefficients. This modified velocity updating operation is defined in Equation (11).

$$v_{id}^{t+1} = w \times v_{id}^{t} + |s_3| \times r_1 \times (p_{id} - x_{id}^{t}) + |s_4| \times r_2 \times \left(p_{gd} - x_{id}^{t}\right) \tag{11}$$

where s_3 and s_4 are two randomly selected scores from the generated values in the x axis, as illustrated in Figure 3. In comparison with other proposed search operations, the absolute values of these selected scores among the 3,142 elements are used as the search coefficients to accelerate convergence. This mechanism increases search diversification by enabling the particles to orbit around the personal and global best experiences in a comparatively more chaotic and irregular manner, in order to search for global optimality.

3.4 The Average Best Leader-enhanced Search Operation

To increase search diversity and avoid stagnation, we propose another search operation driven by an average best leader. This operation adopts the average position of all the neighbouring fitter solutions of the current particle to lead the search process. Specifically, we identify all the particles with better fitness scores than that of the current individual in the swarm. The mean position of these fitter solutions is calculated and used to lead the search process. Equation (12) defines this search operation.

$$x_{id}^{t+1} = average\_best_d^t + s_5 \times (average\_best_d^t - x_{id}^t) \quad (12)$$

where $average\_best_d^t$ denotes the mean position of all fitter solutions in the d^{th} dimension during the t^{th} iteration. The search parameter, s_5, is a randomly selected score among the 3,142 elements within [-1.5253, 1.5253] in the x axis produced by Equations (3)-(6) and (9). This search mechanism enables the current particle to conduct local exploitation of the mean best position along a reticulated flight.

3.5 The Random Best Leader-enhanced Search Operation

To reduce the probabilities of premature convergence, another search operation driven by a random best leader is also proposed. This operation randomly selects one of the fitter neighbouring solutions of the current particle to explore the search space. Specifically, all the particles with better fitness scores than that of the current individual are retrieved. Then, we randomly select one of these fitter solutions to lead the search process. Equation (13) defines this new search operation.

$$x_{id}^{t+1} = random\_best_d^t + s_6 \times (random\_best_d^t - x_{id}^t) \quad (13)$$

where $random\_best_d^t$ indicates a randomly selected fitter individual than the current particle in the swarm in the d^{th} dimension during the t^{th} iteration. The search parameter, s_6, is a randomly selected coefficient among the 3,142 values within [-1.5253, 1.5253] in the x axis produced by Equations (3)-(6) and (9). This search action increases diversification by following multiple distinctive leaders for exploration of the search space.

3.6 Swarm Leader Enhancement

After identifying the global best solution, a search operation, which adopts two randomly selected particles and a super-ellipse coefficient generated using Equations (3)-(6) and Equation (14), is employed to further enhance the best particle, g_{best}. Equation (15) defines this search action for the global best enhancement.

$$m = k = 1, n_1 = 5, and\ n_2 = n_3 = n_4 = 4 \quad (14)$$

$$g'_{best} = g_{best} + s_7 \times (x_h - x_l) \quad (15)$$

where x_h and x_l denote two randomly selected particles in the swarm, while s_7 denotes the randomly selected super-ellipse coefficient.

Moreover, another instantiation of the super-ellipse formula defined in Equations (3)-(6) with the parameter settings provided in Equation (14) is employed to generate a comparatively constrained elliptical curve, as compared with the one produced in Section 3.1. Figure 4 illustrates the corresponding flight with 3,142 generated unique points and with the value range of [-1.1892, 1.0275] for the x axis using the equations. The search coefficient, s_7, in Equation (15) is assigned with a randomly selected value among the 3,142 elements in the range of [-1.1892, 1.0275] along the x axis. A new offspring solution, g'_{best}, is subsequently generated using Equation (15). It is used to replace g_{best} if it has a better fitness score. This search operation implements local exploitation with both positive and negative coefficients to overcome local optima traps.

Overall, the above proposed search mechanisms empower the swarm particles with distinctive search actions in any single iteration to increase diversification.

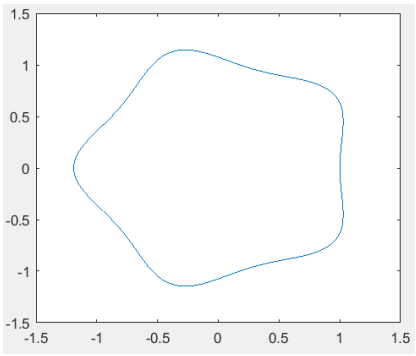


Figure 4 A comparatively constrained reticulated network generated by Equations (3)-(6) and (14)

4. THE PROPOSED ENSEMBLE IMAGE SEGMENTATION MODEL

We employ transfer learning to fine-tune Mask R-CNN for OD segmentation using the fundus image data sets. The proposed PSO model is used to optimize the hyper-parameters, i.e. learning rate and momentum, of the transfer learning process. Optimal distinctive base segmenters are subsequently generated using different identified optimized settings. An ensemble model with a pixel-level majority voting strategy is constructed to combine the results of the base segmenters and derive the final segmented OD boundaries. We discuss the proposed ensemble transfer learning model for OD segmentation in detail, as follows.

4.1 Optimized Base Segmenter Generation Using the Proposed PSO Model

Before we explain the generation of the base segmenter using Mask R-CNN integrated with the proposed PSO model, we introduce state-of-the-art studies on deep learning-based object detectors. Proposed by Girshick et al. [47], the region-based CNN (R-CNN) model acts as a CNN-based object detector. R-CNN employs a multi-stage process for object detection, which includes region proposal generation, CNN-based deep feature extraction, classification of each region using class-specific SVM models, and regression-based bounding-box prediction. R-CNN is the earliest model employing a multi-stage pipeline training process for demonstrating the superiority of a CNN model in object detection against existing methods based on simpler HOG-like features on the PASCAL VOC 2012 data set.

Fast R-CNN [48] improves the training drawbacks of R-CNN. It conducts training with a single-stage end-to-end deep architecture with a multi-task loss. Fast R-CNN employs a fully convolutional network to extract a feature map for the input image. An ROI pooling layer is then used to extract a fixed-length feature vector for each region proposal from the feature map for the subsequent classification and bounding-box regression operations.

Faster R-CNN [49] replaces selective search in Fast R-CNN with a Region Proposal Network (RPN) for more efficient and accurate region proposal generation. Specifically, Faster R-CNN comprises a unified network by merging RPN and Fast R-CNN, where the convolutional features are shared between these two components to enable nearly cost-free region proposal generation. It conducts object detection using a two-step process, i.e. RPN-based bounding-box generation as well as object detection and classification based on the generated region proposals using Fast R-CNN.

Mask R-CNN [16, 17] is an extension of Faster R-CNN with additional instance segmentation on top of object localization and recognition. It achieves this by adding a mask branch for binary segmentation mask generation of each ROI. Mask R-CNN [16, 17] achieved a superior performance on the 2017 COCO suite of challenges in all the three tracks, i.e. instance segmentation, bounding-box object detection, and person keypoint detection. Similar to Faster R-CNN, it employs a two-stage process consisting of RPN-based region proposal prediction and object mask generation as well as object detection and classification using Fast R-CNN.

Mask R-CNN with the ResNet-101 and Feature Pyramid Network (FPN) backbone achieved significantly better performances in instance segmentation than those of other state-of-the-art methods, such as the Multi-task Network Cascade model (MNC) [50], and Fully Convolutional Instance-aware Semantic segmentation (FCIS) [51], i.e. the winners of the COCO 2015 and 2016 segmentation challenges, respectively. Therefore, we employ Mask R-CNN with the ResNet-101+FPN backbone for OD segmentation in this research. Comprehensive reviews of Mask R-CNN and other object detectors are provided in [18, 19].

We perform transfer learning on top of the pre-trained Mask R-CNN model using two retinal image data sets. The proposed PSO model is used to optimize the learning rate and momentum of the transfer learning process owing to the significant impact of both hyper-parameters toward the network performance. As an example, a large learning rate is more likely to produce sub-optimal solutions, while a small learning rate is more inclined to result in a longer training process. Moreover, the momentum setting plays a significant role in avoiding local optima. It determines the contribution of the previous gradient step to the current iteration. The settings of '0' and '1' indicate no contribution and maximum contribution from the previous step, respectively. It increases the step size of moving toward the global optimum if the gradient keeps pointing to the same direction, whereas it smooths out the random variations if the gradient direction is constantly changing. A large momentum combined with a large learning rate is more likely to go over the global optimum, while a small momentum may not be sufficient enough to escape from the sub-optimal solutions. The optimal setting of the learning rate and momentum is dependent on the problem domains, such as the number of local optima, deepness of each local optimum, and smoothness of the problem functions. As such, we provide an automatic process of identifying the optimal setting of both hyper-parameters using the proposed PSO model in this research.

In our transfer learning process, the training stage starts with the pre-trained weights based on the COCO data set. Owing to the pre-training stage using diverse real-life images from the COCO data set, the pre-trained weights possess rich feature representations of generic images. The transfer learning process re-trains the last few layers of the Mask R-CNN model, and generates weights that are embedded with the features learned from the retinal image domain.

For each data set, an aspect ratio of 60-20-20 is employed to form the training, validation, and test data sets. The fitness function illustrated in Equation (16), which employs the IoU score as the performance indicator, is used to guide the search process.

$$score = \frac{Area\ (R_s \cap R_g)}{Area\ (R_s \cup R_g)} \tag{16}$$

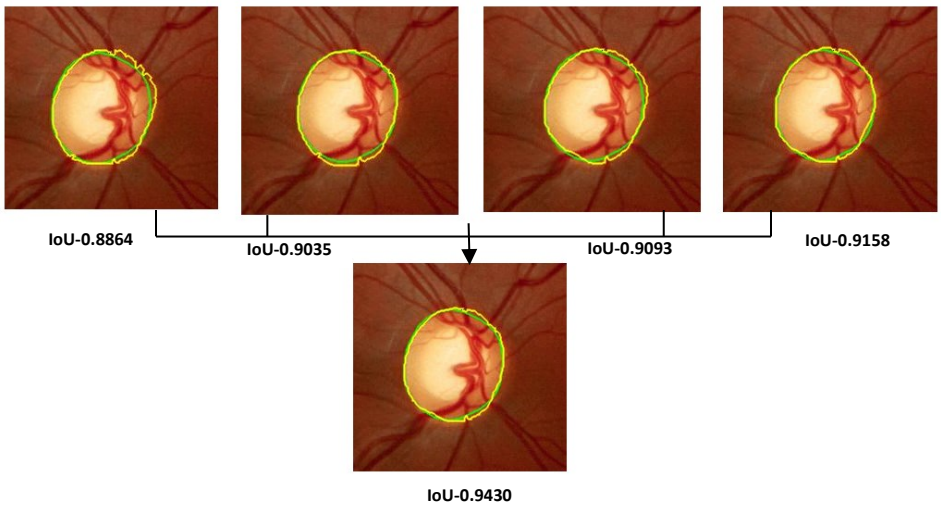
where R_s and R_g denote the segmented OD region and the ground truth (GT), respectively. The IoU score is calculated using the intersection divided by the union of the segmented region and the GT mask. Many related studies [3, 6, 7, 22, 30] employ this IoU score to evaluate the segmentation results against the ground truth. As such, it is adopted in this research, in order to facilitate a valid comparison with other related studies. Equipped with elliptical accelerated and refined search actions, random and mean leader enhanced operations and swarm leader enhancement, the proposed PSO model is able to yield impressive capabilities in identifying effective learning hyper-parameters for deep CNN models. The details are as follows.

The search process of optimal hyper-parameter selection is conducted. A particle swarm is initialized randomly. Each particle is used to represent the possible assignments of the two hyper-parameters. The proposed search operations guide the particles to move towards the promising search regions. The global best solution denotes the identified best hyper-parameters. The training and validation sets of each image data set are used for the optimal hyper-parameter identification using the proposed model.

The transfer learning process with the identified optimal training configurations is subsequently used to train Mask R-CNN on top of the pre-trained weights. To maximize the performance, we combine the training and validation sets into a larger training set for this fine-tuning process of transfer learning. Significant larger numbers of training steps (e.g. 100) and epochs (e.g. 50) are also used for this final operation of transfer learning. The generated new segmenters with different training epochs (e.g. 20, 30, 40, and 50 epochs) are then used for OD segmentation using the test data set. We repeat the process to generate a number of optimized base segmenters with different optimal hyper-parameters and different numbers of epochs.

4.2 Ensemble Segmentation Model Generation

Individual segmenters are produced using the abovementioned transfer learning process incorporated with the proposed PSO-based parameter identification. The generated optimized base segmenters are subsequently used to construct an ensemble network.



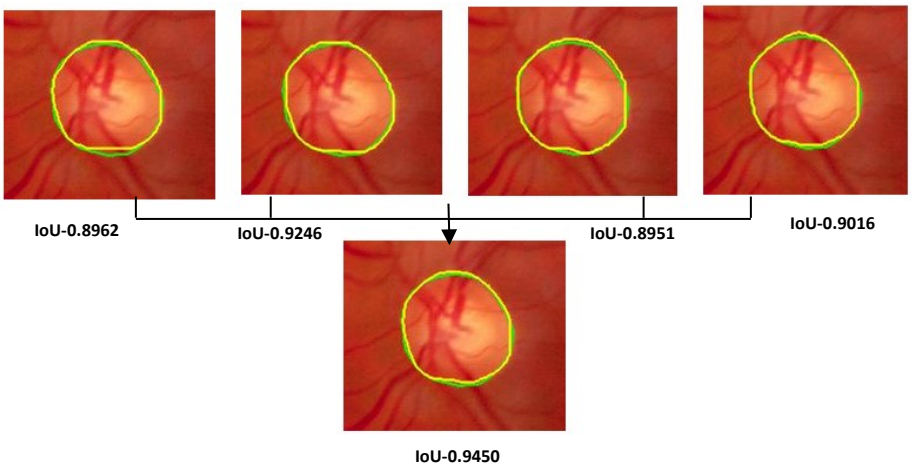


Figure 5 Two example segmentation processes for the proposed deep ensemble transfer learning network (yellow line: predicted boundaries, green lines: GT annotations). In each example, the top row presents the segmentation outputs of four base segmenters while the bottom row illustrates the output of the ensemble network.

To increase the base model diversity, we employ the newly generated weights trained using different hyper-parameters and different numbers of epochs (e.g. 20, 30, 40, and 50 epochs) as the base evaluators. In this research, we employ a set of four individual segmenters as the base models. As an example, we incorporate two individual segmenters trained with one set of identified hyper-parameters and learning epochs of 20 and 30 with another two segmenters trained with another set of identified hyper-parameters and learning epochs of 40 and 50 to construct one ensemble model, and vice versa. A pixel-level majority voting method is conducted based on the outputs of the base models to determine the final segmentation outcome for each retinal image. A total of 30 runs have been performed for hyper-parameter identification to construct 30 ensemble models. The mean result of 30 ensemble models is used as the main criterion for comparison. Two example ensemble segmentation processes are illustrated in Figure 5. As indicated in Figure 5, the ensemble models overcome the bias and variance of their corresponding base evaluators to obtain improved segmentation performances.

In short, the transfer learning process incorporating the proposed PSO-based hyper-parameter identification method is able to produce a variety of optimal base evaluators to increase ensemble diversity and enhance performance. We discuss the detailed evaluation results in Section 5.

5. EVALUATION

We employ two retinal image data sets, i.e. Drions [52] and Messidor [53], to evaluate the proposed ensemble model in OD segmentation. The Drions data set contains a total of 110 fundus images. For each image, there are two GT annotations for the OD boundaries provided by two medical experts. The image resolutions are 600×400, and there are various challenging characteristics such as strong pallor distractors, low-contrast blurry boundaries, and blood vessel interference. To enable a valid comparison with other related studies [6, 7], we employ the annotations of the first annotator as the GT and calculate the IoU scores against these GT disc boundaries.

Another public data set employed for evaluation is the Messidor data set, which consists of 1,200 retina images in three resolutions, i.e. 2240×1488, 2304×1536 and 1440×960. The GT annotations of the OD boundaries are also provided by medical experts [54]. We resize the images of this data set to 600×400 resolutions, in order to achieve a reasonable trade-off between performance and computational efficiency. There are two types of annotations provided for each image, i.e. diabetic retinopathy (DR) grading and the risk of DME. We employ the ratio of 60-20-20 as the training, validation, and test sets for both Drions and Messidor databases in hyper-parameter optimization.

5.1 Evaluation Using the Drions Data Set for OD Segmentation

For a comprehensive evaluation, we have implemented the baseline ensemble models with the default settings as well as those with optimal hyper-parameters identified by the original and other PSO methods. As an example, CPSO [20] is implemented, where the search coefficients are generated by the cosine annealing process. Motivated by CPSO, another adaptive baseline PSO variant, EPSO, with search parameters generated

by exponential functions is implemented for comparison. Equations (17)-(18) define the descending and ascending coefficient generation by the exponential functions in EPSO.

$$c_1 = 0.5 + \frac{exp(4.8-k)}{6\times max\_iter} \tag{17}$$

$$c_2 = 0.5 + \frac{exp(k-5.2)}{6\times max\_iter} \tag{18}$$

where k and max_iter represent the current and maximum numbers of iterations, respectively. Both CPSO and EPSO models are selected for comparison because they employ ascending and decreasing search parameters, as compared with the random search coefficients in the proposed model, to balance between intensification and diversification. They are subsequently employed for optimal hyper-parameter identification during the transfer learning process.

We employ the following experimental settings for optimal hyper-parameter identification using the transfer learning process, i.e. the maximum number of function evaluations=15 (population) × 10 (maximum number of iterations), dimension=2, and trial=30. In other words, all the search methods terminate when the maximum number of function evaluations is reached for optimal hyper-parameter selection, in order to ensure a fair comparison. After identifying the optimal network configurations, we train Mask R-CNN using a total of 50 epochs and 100 learning steps per epoch with a larger training set by combining the training and validation sets. We generate four distinctive optimized base models as the base segmenters. The segmentation results of the base models are combined using a pixel-level majority voting strategy to obtain the final segmentation mask of each test image. We employ all the 110 images from the Drions data set for OD segmentation, with an aspect ratio of 60-20-20 for forming the training, validation, and test data sets. A total of 30 trials are conducted, and the average IoU scores over 30 runs for the ensemble networks constructed by all the search methods are employed for comparison.

Besides the adoption of transfer learning incorporating the above PSO variants, hybrid clustering methods are implemented for comparison, i.e. PSO+FCM, Genetic Algorithm (GA)+FCM [14], and Adaptive Learning PSO (ALPSO)+FCM [21]. Specifically, in GA+FCM hybrid clustering [14], the GA is used to improve the centroids identified by FCM clustering in conjunction with the intra- and inter-clustering measurement in the objective function. Similarly, in ALPSO+FCM [21], ALPSO, which comprises a helix search action, Simulated Annealing (SA), and modified DE and PSO with spiral coefficients, is used to enhance the centroids of FCM clustering in skin lesion segmentation, where the between- and within-cluster variances are considered in the fitness function. Finally, PSO in conjunction with FCM clustering is implemented for comparison, namely a PSO+FCM hybrid model. We employ the following comparatively larger experimental settings for clustering-based OD segmentation using the recommendations in their original studies, i.e. the maximum number of function evaluations=20 (population) × 20 (maximum number of iterations), number of clusters=2, and runs=30. The three hybrid clustering methods terminate when the maximum number of function evaluations is reached. Ensemble hybrid clustering models are subsequently constructed with the above three types of clustering models as the base segmenters, respectively. Specifically, four instantiations of the GA+FCM model are used to construct one ensemble clustering model. Similar conditions are applied to the ensemble models built by ALPSO+FCM and PSO+FCM, respectively. A majority voting mechanism is used to combine the results from the base segmenters and to yield the final segmentation outcome of each ensemble clustering method. The parameter settings of all the search methods are shown in Table 1, which are consistent with the recommendations in their original studies.

Table 1 Parameter settings of different search methods

Models	Parameter settings
Prop. PSO	maximum velocity=0.6, inertia weight=0.6, using super-ellipse search coefficients as search parameters
CPSO [20]	ascending and descending coefficients generated by cosine annealing functions
EPSO	ascending and descending coefficients generated by exponential annealing functions
ALPSO [21]	maximum velocity=0.6, inertia weight=0.6, using random helix coefficients as the search parameters
GA	crossover probability=0.7, mutation probability=0.3
PSO	maximum velocity=0.6, inertia weight=0.5, acceleration constants c_1= c_2= 1.5

Table 2 illustrates the average IoU scores of the ensemble models with the hyper-parameters identified by the proposed, the original, and other PSO methods, respectively, as well as those with default hyper-parameter settings, over a set of 30 runs. The mean results of three ensemble hybrid FCM clustering methods incorporating PSO, GA and ALPSO, respectively, over a set of 30 runs are also provided for comparison.

Table 2 Evaluation results for the Drions data set over a set of 30 runs

Models	Methodologies	IoU scores
Prop. PSO-based Ensemble	Prop. PSO + ensemble Mask R-CNN transfer learning model	0.916
CPSO-based Ensemble [20]	CPSO + ensemble Mask R-CNN transfer learning model	0.9088
EPSO-based Ensemble	EPSO + ensemble Mask R-CNN transfer learning model	0.9001
PSO-based Ensemble	PSO + ensemble Mask R-CNN transfer learning model	0.9052
Ensemble Model with Default Settings	Ensemble Mask R-CNN transfer learning model with default parameter settings	0.8847
Ensemble Clustering Using ALPSO+FCM [21]	Ensemble clustering with ALPSO+FCM as the base	0.564
Ensemble Clustering Using GA+FCM [14]	Ensemble clustering with GA+FCM as the base	0.5206
Ensemble Clustering Using PSO+FCM	Ensemble clustering with PSO+FCM as the base	0.511

Table 3 Comparison with related studies for the Drions data set

Relate studies	Methodologies	IoU scores
Prop. PSO-based Ensemble	Prop. PSO + ensemble Mask R-CNN transfer learning model	0.916
Dai et al. [7]	Variational model with multiple energies	0.9081
Zahoor and Fraz [27]	A polar transform-based adaptive thresholding	0.886
Li et al. [6]	A brightness based method for OD localization + OD segmentation based on learning the supervised descent directions	0.8532
Morales et al. [3]	PCA and mathematical morphology	0.8424
Rehman et al. [23]	Superpixel based feature classification	0.821
2nd observer	2nd human expert annotation	0.9202

As illustrated in Table 2, the optimal ensemble segmenter built by the proposed PSO model achieves the best IoU score of 0.916 over a set of 30 runs, in comparison with those using the default parameter settings and the optimal hyper-parameters identified by CPSO, EPSO, and PSO, respectively. The proposed PSO model with the accelerated and refined search coefficients for simulating diverse elliptical shapes is able to balance well between intensification and diversification, and depict great efficiency in overcoming premature convergence. Among the baseline methods, CPSO with adaptive cosine coefficients and the original PSO model produce more effective training hyper-parameters than those identified by EPSO with the exponential functions. This is likely to be caused by the constrained search exploration capabilities of EPSO with drastically descending c_1 and slowly ascending c_2 produced by the exponential formula. The optimised ensemble segmenters produced by all the search methods yield better IoU scores than those of the ensemble models formulated based on the default training configurations [16, 17], where learning rate=0.001 and momentum=0.9.

With respect to the three hybrid ensemble clustering methods with a set of GA+FCM, ALPSO+FCM, PSO+FCM, as the base segmenters, respectively, without blood vessel and false peak removal, the segmentation tasks pose great challenges to these ensemble clustering methods. They only segment partial OD regions in most of the test cases, owing to various interferences that severely affect their IoU scores.

Table 3 shows a comparison with related studies pertaining to OD segmentation for the Drions data set. The proposed ensemble model is among the top performers, and shows comparable performance with those from other state-of-the-art studies. Moreover, the proposed ensemble model shows great efficiency in OD segmentation, and yields a comparable IoU score (i.e. 0.916) to that (i.e. 0.9202) with respect to the overlapping comparison between the two human annotators. Some examples of the segmentation results from the proposed ensemble model are provided in Figure 6.

Figure 6 Example segmentation results of the proposed ensemble model (yellow lines: predicted boundaries, green lines: GT annotations)

5.2 Evaluation Using the Messidor Data Set for OD Segmentation

To further ascertain the model efficiency, the Messidor data set is employed. The data set has 1,200 images. We employ the ratio of 60-20-20 for forming the training, validation, and test sets. The aforementioned experimental settings for both types of ensemble model construction are also employed for the optimal hyper-parameter identification. Transfer learning with 50 epochs and 100 steps per epoch is performed using the optimized training configuration identified by the proposed PSO model and other search methods. Table 4 shows the detailed evaluation results of the Messidor data set over a set of 30 runs.

Table 4 Evaluation results of the Messidor data set over a set of 30 runs

Models	Methodologies	IoU scores
Prop. PSO-based Ensemble	Prop. PSO + ensemble Mask R-CNN transfer learning model	0.915
CPSO-based Ensemble [20]	CPSO + ensemble Mask R-CNN transfer learning model	0.9049
EPSO-based Ensemble	EPSO + ensemble Mask R-CNN transfer learning model	0.8984
PSO-based Ensemble	PSO + ensemble Mask R-CNN transfer learning model	0.9020
Ensemble Model with Default Settings	Ensemble Mask R-CNN transfer learning model with default parameter settings	0.8823
Ensemble Clustering Using ALPSO+FCM [21]	Ensemble clustering with ALPSO+FCM as the base	0.4719
Ensemble Clustering Using GA+FCM [14]	Ensemble clustering with GA+FCM as the base	0.5120
Ensemble Clustering Using PSO+FCM	Ensemble clustering with PSO+FCM as the base	0.4582

Table 5 Comparison with related studies for the Messidor data set

Relate studies	Methodologies	IoU scores
Prop. PSO-based Ensemble	Prop. PSO + ensemble Mask R-CNN transfer learning model	0.915
Li et al. [6]	A brightness based method for OD localization + OD segmentation based on learning the supervised descent directions	0.9128
Dai et al. [7]	Variational model with multiple energies	0.91
Dashtbozorg et al. [8]	A multiresolution sliding band filter	0.89

Sarathi et al. [4]	Adaptive threshold based region growing technique followed by strategic blood vessel inpainting	0.89
Lim et al. [30]	A 9-layer CNN model with exaggerated feature inputs	0.888
Giachetti et al. [55]	Symmetry-based OD localization + a coarse-to-fine scheme for OD segmentation based on information related to inpainted images and vessel masks	0.88
Cheng et al. [56]	Superpixel classification based OD and OC segmentation	0.88
Marin et al. [57]	Opening-closing morphological operations + a 2-step automatic thresholding procedure based on Circular Hough Transform and the Prewitt edge detector	0.87
Aquino et al. [58]	Morphological and edge detection techniques followed by the Circular Hough Transform + a voting-type algorithm	0.86
Zhang et al. [22]	Faster R-CNN and a shape constrained level set algorithm	0.854
Yu et al. [59]	Directional matched filtering and a hybrid level-set model	0.844
Zahoor and Fraz [27]	A polar transform-based adaptive thresholding	0.844
Roychowdhury et al. [60]	A Gaussian mixture model classifier with region-based features	0.84
Morales et al. [3]	PCA and mathematical morphology	0.8228
Rehman et al. [23]	Superpixel based feature classification	0.747

As illustrated in Table 4, the proposed PSO model is able to identify more efficient hyper-parameters than those produced by CPSO, EPSO and PSO, respectively. Therefore, our resulting ensemble segmentation model achieves a superior average IoU score of 0.915, as compared with those of the baseline ensemble methods, over a set of 30 runs. In addition, the optimized ensemble segmenters produced by all the PSO models show better performances than those of ensemble models with the default hyper-parameter settings. This is owing to the less effective training configurations and limited diversity of the ensemble segmenters with the default settings. CPSO produces more effective learning hyper-parameters than those from EPSO and PSO, respectively. This Messidor data set also poses great challenges to the three hybrid ensemble clustering methods, owing to the presence of certain characteristics such as low-contrast boundaries, vasculature occlusion, and bright lesion distractors, which severely affect the clustering segmentation performance.

Table 5 depicts a comparison with state-of-the-art studies for the Messidor data set. The existing methods adopt diverse deep networks, morphological operations, edge detectors, and filtering techniques for OD segmentation. Our ensemble transfer learning segmenter based on Mask R-CNN and the proposed PSO-based hyper-parameter identification depicts superior capabilities and flexibilities in OD segmentation without any pre- and post-processing procedures. By using only the image-level inputs, the proposed model is among the top performers, as compared with other existing studies. The empirical results also indicate that the deep features learned offer great superiority over hand-crafted features [4, 23, 27, 55, 57, 58] in OD segmentation.

To further indicate the significance of the proposed PSO-based ensemble segmenter, the Wilcoxon rank sum test is conducted for both Drions and Messidor data sets for OD segmentation over 30 runs. As shown in Table 6, all the statistical test results are lower than 0.05, which indicate that the proposed PSO-based ensemble model outperforms all hybrid ensemble clustering methods and the baseline deep ensemble models with the default and optimal hyper-parameters identified by all other PSO variants, statistically.

Table 6 The Wilcoxon rank sum test results for OD segmentation for Drions and Messidor data sets

	CPSO-based Ensem	EPSO-based Ensem	PSO-based Ensem	Default Ensem	ALPSO+FCM-based Ensem	GA+FCM-based Ensem	PSO+FCM-based Ensem
Drions	2.74E-11	2.32E-11	2.53E-11	1.34E-11	1.08E-12	1.08E-12	1.08E-12
Messidor	4.24E-06	2.94E-11	2.97E-11	1.45E-11	1.20E-12	1.20E-12	1.20E-12

5.3 Evaluation Using the Messidor Data Set for Diagnostic Problems

Besides the OD segmentation tasks, we also evaluate the efficiency of the proposed PSO model for hyper-parameter fine-tuning in another transfer learning process for the grading of a type of retinal diseases, i.e. the risk of DME. Specifically, we construct ensemble classification models with hyper-parameters fine-tuned by the proposed PSO and other search methods, respectively, for DME grading. As mentioned earlier, the Messidor data set [53] has two types of diagnoses, i.e. DR grading and DME. With respect to DME, each image is classified into one of the three categories, i.e. '0 (no risk)', '1 (mild risk)', and '2 (severe risk)'. There are 971, 75, and 154 samples for grades 0, 1, and 2, respectively, in the Messidor data set. The 'no risk' category

represents the normal cases, while the 'mild' and 'severe' categories indicate that the shortest distance between macula and hard exudates is more than or less than one papilla diameter, respectively. Owing to the severe imbalanced sample sizes for the three cases, a total of 420 images are employed for DME detection with 191, 75, and 154 samples for '0 (no risk)', '1 (mild)' and '2 (severe)' cases, respectively.

Since VGG16 [61], a deep learning model, achieves comparatively better performances in a trial-and-error process for DME grading among a number of other deep networks including GoogLeNet, Inception-V3, ResNet101, AlexNet, and VGG19, it is selected in this research. We employ the VGG16 model as the base classifier, which is fine-tuned using the Messidor data set to perform classification of the three DME cases. The proposed PSO model is used to optimize the learning rate and momentum for this transfer learning process based on VGG16, owing to the impact of these hyper-parameters to the model performance. As discussed earlier, a large learning rate is more likely to produce sub-optimal solutions, while a small learning rate is more inclined to require a longer training time. The momentum parameter has a significant impact on avoiding the local optima traps. The default parameter settings of the transfer learning process are also used for comparison, where learning rate=0.01 and momentum=0.9 (recommended in MATLAB) with the stochastic gradient descent with momentum as the optimizer. We employ a ratio of 70-10-20 for model training, validation, and test, for DME detection using the Messidor data set.

Besides PSO, CPSO [20] and ALPSO [21], we employ other classical methods and PSO and FA variants for performance comparison, i.e. Moth-Flame Optimization (MFO) [62], Genetic PSO (GPSO) [63], dynamic neighbourhood learning PSO (DNLPSO) [64], ELPSO [65], CFA1 with Logistic map as random search parameters [66], CFA2 with Gauss map as the attractiveness coefficients [67], and variable step size FA (VSSFA) [68]. These classical and advanced search methods show competitive capabilities in solving diverse challenging optimization problems. The parameter settings of these additional baseline methods are shown in Table 7, which are consistent with the recommendations in their original studies. The following experimental settings are employed for hyper-parameter selection, i.e. the maximum number of function evaluations=150, dimension=2, and trial=30. All the search methods terminate when the maximum number of function evaluations is reached. After identifying the optimized training configurations of the base VGG16 models using the proposed PSO and other search methods, we further improve DME classification by constructing ensemble VGG16 models. In other words, we construct ensemble VGG16 models with hyper-parameters fine-tuned by each search method for DME grading.

Table 7 Parameter settings of additional baseline methods

Methods	Parameter settings
MFO [62]	Use adaptive parameter settings
GPSO [63]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1 = 2.6, c_2 = 1.5
DNLPSO [64]	c_1=c_2=1.49445, refreshing gap=3, regrouping period=5, inertia weight=0.9−(0.9−0.4)×(k−1)/(MaxGeneration−1), where k and MaxGeneration represent the current and maximum iteration numbers, respectively.
ELPSO [65]	c_1=c_2=2, standard deviation of Gaussian mutation=1, scale parameter of Cauchy mutation=2, scale factor of DE-based mutation=1.2, inertia weight=0.9−(0.9−0.4)×(k−1)/(MaxGeneration−1).
CFA1 [66]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=Logistic map
CFA2 [67]	attractiveness=Gauss map, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.5
VSSFA [68]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.4/(1+exp(0.015*(k-MaxGeneration)/3)).

For each base classifier generation, data augmentation and a fine-tuning process with a larger training sample size are used to enhance the performance. Data augmentation operations are firstly employed to diversify the training samples and avoid overfitting. Augmentation operations such as scaling (horizontal and vertical scaling of the input images by 10%), reflection (randomly flipping the images in the left-right direction with 50% probability), and translation (random horizontal and vertical translations of the images up to 30 pixels), are performed. The augmentation pre-processing operation also resizes the original images to 224×224, which is required by the adopted VGG16 model. Such data augmentation procedures are able to prevent the networks from learning the training images too tightly, in order to avoid overfitting [69]. After identifying the optimal hyper-parameters using the training (70%) and validation (10%) sets, we combine the training and validation sets into a larger training set (80%). We use this larger training set as well as the identified optimal training configurations to fine-tune the VGG16 model to avoid overfitting. The mini-batch size is increased to 10 in this model fine-tuning process, in comparison with 6 used in the hyper-parameter selection process using each search method. To further enhance the performance, for each search method, we subsequently construct an

ensemble classifier with three fine-tuned base VGG models yielded using the abovementioned base model generation process. The results obtained by the three base VGG16 models are combined using a majority voting mechanism to determine the final classification outcome. Since the three base classifiers employ different optimized training configurations, they possess great diversity for performance enhancement.

The mean diagonal classification accuracy rate of the confusion matrix of the three classes obtained for the test set is used for comparison. Such an evaluation strategy is also able to avoid any biases contributed by the dominating classes with comparatively larger sample sizes. Table 8 shows the mean result over a set of 30 runs of each search method.

Table 8 Evaluation results for DME grading over a set of 30 runs

	Prop. PSO	**PSO**	**MFO**	**CPSO**	**ALPSO**	**GPSO**	**DNLPSO**	**ELPSO**	**CFA1**	**CFA2**	**VSSFA**	**default**
mean	0.9593	0.8889	0.8151	0.9074	0.8703	0.8574	0.8148	0.9111	0.8778	0.8333	0.8555	0.7999

As indicated in Table 8, the ensemble VGG16 model with the hyper-parameters identified by the proposed PSO model outperforms those with the default and optimal hyper-parameters identified by other search methods for DME grading, significantly. The superiority of the proposed PSO model is further ascertained by the Wilcoxon rank sum test results. As presented in Table 9, the proposed ensemble VGG16 model outperforms the default ensemble model and those with training options yielded by all other baseline search methods, statistically.

Table 9 The Wilcoxon rank sum test results for DME grading

	PSO	**MFO**	**CPSO**	**ALPSO**	**GPSO**	**DNLPSO**	**ELPSO**	**CFA1**	**CFA2**	**VSSFA**	**default**
RS	7.83E-08	3.87E-09	5.36E-06	5.36E-06	6.13E-07	2.83E-09	2.91E-04	4.87E-06	2.48E-07	4.65E-08	5.51E-09

Table 10 Comparison with related studies for DME detection for the Messidor data set

Relate studies	Methodologies	Classes	Data sets used	Image size	Evaluation strategy	Accuracy on Messidor
Giancardo et al. [70]	The exudate probability map obtained using background subtraction, colour and wavelet analysis + supervised classifiers	2	Messidor	Resize the images to a height of 752 pixels maintaining the original height/width ratio.	Hold-one-out	0.91
Lim et al. [71]	Watershed transformation	3	Messidor	-	Hold-out	0.852
Akram et al. [72]	Filter bank + SVM	3	Messidor	2240×1488 2304×1536 1440×960	-	0.973
Sreejini and Govindan [73]	PSO and FCM for exudates segmentation and optic disc elimination etc. + Bayes classifier	3	Messidor	-	-	0.945
Mookiah et al. [74]	Higher-order spectra + feature selection & ranking + supervised classifiers	3	Messidor	1440×960	10-fold	0.9556
Ren et al. [75]	Vector quantization and semi-supervised learning	3	Messidor + E-ophtha EX (20%)	-	Hold-out	0.889
Ren et al. [75]	Vector quantization and semi-supervised learning	3	Messidor + E-ophtha EX (80%)	-	Hold-out	0.975
Al-Bander et al. [76]	CNN model with 3 Conv. Blocks and one FC block	3	Messidor	128×128 256×256 512×512	Hold-out	0.888
Sahlsten et al. [77]	Ensemble of Inception-v3	3	A private training set with 28,512 images + Messidor	512×512	Hold-out	0.953

Li et al. [78]	Cross-disease attention network (CANet) with the disease-specific and disease-dependent attention modules	3	Messidor	350×350	10-fold	0.912
Chen et al. [79]	A multi-task deep learning model	3	Messidor	350×350	10-fold	0.905
Prop. PSO + Ensemble VGG16	Prop. PSO + ensemble VGG16 transfer learning model	3	Messidor	224×224	Hold-out	0.9593

Table 10 shows the comparison with related research studies. Since different training databases, sample sizes, evaluation strategies, and input image resolutions have been used in the related studies, the results in Table 10 serve as a rough performance comparison between the proposed research and the existing methods for DME grading. As indicated in Table 10, many related studies employ hand-crafted features and rely largely on pre-processing and morphological operations to detect both the exudate and macula regions, while eliminating distractors such as blood vessels and OD regions, in order to assist DME detection. As an example, Lim et al. [71] conducted DME grading by employing pre-processing operations such as exudate extraction and fovea localization using contrast-limited adaptive histogram equalization, morphological opening and dilation, and watershed transformation. Akram et al. [72] conducted exudate and macula detection using a filter bank before performing feature extraction for SVM-based DME classification. Sreejini and Govindan [73] employed hybrid FCM clustering integrated with PSO for exudate segmentation, fovea and macular region localization, and OD detection, in combination with Bayes classifier-based DME identification. A multi-step pre-processing operation was adopted by Ren et al. [75], including vessel, OD and macula localization, exudate segmentation using vector quantization and semi-supervised learning based DME grading. Besides the abovementioned pre-processing procedures, Mookiah et al. [74] employed Higher Order Spectra for feature extraction, Spectral Regression Discriminant Analysis (SRDA) and minimum Redundancy Maximum Relevance (mRMR) for feature selection and discrimination of different DME stages. These related studies also employed significantly larger image input resolutions.

As mentioned above, they all require additional location information pertaining to macular and exudate from the pre-processing steps to perform the grading tasks. In addition, it is difficult to obtain the annotations of the location information of exudates and macula. Ren et al. [75] obtained such additional knowledge, i.e. the location information of exudate, by utilizing an additional data set, i.e. the E-ophtha EX data set, which was combined with the Messidor data set for model training and performance enhancement. Their best performance was achieved when 80% of a labelled data set with such additional information extracted from E-ophtha EX was used together with Messidor in the training process.

In comparison with the above morphological and texture-based methods, in this research, we conduct the DME grading task without using any pre-processing procedures, such as exudate segmentation, fovea and macular region localization, and blood vessel removal. In other words, we do not utilize any features related to the macular and exudate location information, but purely rely on the image-level inputs for discrimination of different DME stages. Besides that, the above related studies mainly employ hand-crafted features, which may show limited discriminative capabilities as compared with those in the proposed approach.

Moreover, deep CNN models were also employed for DME grading, such as Al-Bander et al. [76], Sahlsten et al. [77], Li et al. [78] and Chen et al. [79]. Sahlsten et al. [77] employed an ensemble deep network and achieved a comparatively better performance among the baseline results. Their method, however, employed a significantly larger additional private retinal image data set (28,512 images) for model training with a variety of larger image resolutions. Their base learners in the deep ensemble model were yielded by using similar training settings, therefore the lack of diversity. In this research, we employ the proposed PSO-based hyper-parameter identification incorporating the VGG16 model to generate the optimal base learners that fit the tasks at hand. The proposed search mechanisms are capable of driving the hyper-parameter search out of local optima traps to attain global optimality. Data augmentation and a final model fine-tuning process are also adopted to improve the performance of base models. The base models with distinctive training hyper-parameters embedded in the ensemble classifier show great diversity to enhance the final ensemble performance. Importantly, we do not use any additional large data sets for training, and we achieve a reasonable trade-off between the performance and computational cost with a comparatively smaller input resolution. In short, our approach offers an effective alternative to DME grading. For future work, we will embed additional lesion (exudate) segmentation masks to locate the abnormal signs so that we can further improve the performance.

5.4 Evaluation Using Test Functions

In addition to the evaluation on OD segmentation and DME detection, a set of nine mathematical benchmark functions is employed to evaluate the model efficiency. They include Ackley, Dixon-Price, Griewank, Rotated Hyper-Ellipsoid, Rosenbrock, Sphere, Sum of Different Powers, Sum Squares, and Powell numerical problems. These artificial landscapes represent a variety of challenging unimodal and multimodal optimization problems. Specifically, Ackley, Griewank, and Powell are multimodal numerical problems with multiple local minima, while the rest are unimodal benchmark functions with single global minimum [65, 80, 81]. The definitions of these test functions are provided in [15, 80].

The following experimental settings are applied, i.e. the maximum number of function evaluations=25,000, dimension=30, and trial=30. The mean, standard deviation, minimum and maximum results are computed for each method over 30 runs, where the mean results are used as the main criterion for comparison. All the methods terminate when the maximum number of function evaluations is reached. Table 11 illustrates the detailed evaluation results. The proposed PSO model with super-ellipse coefficients shows superior capabilities in solving these unimodal and multimodal mathematical landscapes. It is able to yield the most optimal (i.e. the minimum) mean results over a set of 30 runs for nearly all the test functions, except for Griewank where ALPSO obtains the best mean result.

Table 11 Evaluation results of the unimodal and multimodal benchmark functions over a set of 30 runs

		Prop. PSO	PSO	MFO	CPSO	ALPSO	GPSO	DNLPSO	ELPSO	CFA1	CFA2	VSSFA
Ackley	**mean**	**4.73E+00**	6.28E+00	9.98E+00	6.15E+00	5.76E+00	1.78E+01	5.16E+00	1.50E+01	1.57E+01	1.48E+01	1.04E+01
	Min	2.41E+00	2.20E+00	**3.87E-01**	4.22E+00	3.84E+00	1.63E+01	2.58E+00	1.32E+01	1.48E+01	1.36E+01	9.64E+00
	Max	8.48E+00	9.46E+00	1.63E+01	**7.99E+00**	8.03E+00	1.85E+01	8.59E+00	1.61E+01	1.65E+01	1.54E+01	1.09E+01
	Std	1.43E+00	1.61E+00	4.12E+00	9.33E-01	1.23E+00	4.86E-01	1.48E+00	5.97E-01	4.24E-01	4.40E-01	3.63E-01
Dixon	**mean**	**6.72E-01**	4.28E+00	1.56E+04	1.34E+00	1.03E+04	4.66E+05	3.68E+01	1.51E+05	1.49E+05	1.41E+05	1.14E+04
	Min	**6.67E-01**	6.83E-01	3.04E+00	6.82E-01	3.09E+03	1.78E+05	**6.67E-01**	6.43E+04	8.87E+04	9.74E+04	5.64E+03
	Max	**6.93E-01**	8.52E+01	1.17E+05	9.12E+00	2.01E+04	7.32E+05	4.98E+02	3.19E+05	2.34E+05	1.90E+05	1.58E+04
	Std	5.77E-03	1.53E+01	4.02E+04	1.59E+00	4.11E+03	1.43E+05	1.01E+02	6.53E+04	3.88E+04	2.52E+04	2.85E+03
Griewank	**mean**	4.64E-07	3.00E-01	6.82E+00	3.66E-01	**0.00E+00**	2.93E+02	2.14E+00	1.34E+02	1.68E+02	1.57E+02	4.52E+01
	Min	6.21E-11	4.60E-02	2.98E-01	2.38E-02	**0.00E+00**	2.11E+02	1.01E-07	9.06E+01	1.30E+02	1.15E+02	3.53E+01
	Max	3.62E-06	1.14E+00	9.09E+01	1.18E+00	**0.00E+00**	3.88E+02	1.28E+01	1.77E+02	1.96E+02	1.89E+02	5.36E+01
	Std	7.95E-07	3.31E-01	2.28E+01	3.99E-01	**0.00E+00**	3.90E+01	3.65E+00	2.03E+01	1.74E+01	1.58E+01	4.75E+00
Rothyp	**mean**	**3.66E-03**	1.88E-01	1.87E+04	2.40E-02	2.86E+05	1.98E+05	1.45E+03	8.97E+04	1.11E+05	1.00E+05	2.92E+04
	Min	1.28E-03	1.96E-02	2.00E+00	5.35E-04	1.95E+05	1.45E+05	**2.94E-07**	5.78E+04	8.07E+04	7.56E+04	2.13E+04
	Max	**1.28E-02**	2.41E+00	1.69E+05	2.45E-01	3.38E+05	2.45E+05	1.75E+04	1.24E+05	1.38E+05	1.18E+05	3.46E+04
	Std	2.20E-03	4.31E-01	3.34E+04	5.56E-02	3.63E+04	2.81E+04	3.54E+03	1.76E+04	1.31E+04	1.26E+04	3.92E+03
Rosenbrock	**mean**	**3.03E+01**	5.81E+01	4.62E+04	4.69E+01	4.21E+03	3.92E+05	8.01E+02	1.13E+05	1.28E+05	7.73E+04	8.21E+03
	Min	**2.09E-02**	9.46E+00	4.28E+01	4.87E-01	1.73E+02	2.26E+05	2.17E+01	3.01E+04	6.14E+04	4.48E+04	5.24E+03
	Max	**8.05E+01**	1.84E+02	2.00E+05	1.36E+02	1.62E+04	5.68E+05	2.01E+04	2.19E+05	1.86E+05	1.04E+05	1.08E+04
	Std	2.79E+01	4.48E+01	5.51E+04	3.87E+01	3.57E+03	1.03E+05	3.65E+03	4.30E+04	2.88E+04	1.42E+04	1.36E+03
Sphere	**mean**	**2.45E-03**	4.41E-02	1.75E+00	1.17E-01	5.60E-02	8.69E+01	2.60E-01	3.98E+01	4.92E+01	4.72E+01	1.25E+01
	Min	9.38E-04	7.94E-03	1.43E-03	7.91E-03	**0.00E+00**	6.09E+01	1.35E-09	2.12E+01	3.57E+01	3.68E+01	9.40E+00
	Max	**4.37E-03**	4.55E-01	2.62E+01	4.24E-01	4.09E-01	1.11E+02	4.05E+00	5.08E+01	6.07E+01	5.37E+01	1.66E+01
	Std	9.45E-04	8.00E-02	6.65E+00	1.04E-01	1.31E-01	1.28E+01	7.53E-01	6.20E+00	6.27E+00	4.52E+00	1.74E+00
Sumpow	**mean**	**2.08E-06**	1.77E-05	3.17E-06	5.04E-05	8.27E-04	1.23E-01	2.34E-05	1.46E-02	1.01E-02	7.02E-03	2.07E-04
	Min	1.57E-07	1.37E-06	3.14E-07	4.70E-06	**0.00E+00**	9.89E-03	1.67E-36	1.34E-03	5.80E-04	1.75E-03	6.69E-05
	Max	**1.02E-05**	7.71E-05	1.14E-05	1.36E-04	1.68E-02	4.68E-01	5.74E-04	3.67E-02	3.25E-02	2.32E-02	6.77E-04
	Std	2.40E-06	1.99E-05	2.84E-06	5.31E-05	3.18E-03	9.95E-02	1.06E-04	8.67E-03	6.32E-03	4.09E-03	1.25E-04
Sumsqu	**mean**	**2.36E-03**	6.72E-02	1.23E+02	1.50E-01	1.26E-01	1.16E+03	8.30E-01	5.61E+02	6.77E+02	6.29E+02	1.88E+02
	Min	7.95E-04	7.16E-03	3.16E-02	1.14E-03	**0.00E+00**	8.22E+02	2.14E-09	2.90E+02	4.89E+02	4.63E+02	1.47E+02
	Max	**4.66E-03**	2.41E-01	8.65E+02	1.17E+00	6.40E-01	1.57E+03	7.94E+00	8.01E+02	8.45E+02	7.66E+02	2.29E+02
	Std	1.15E-03	5.78E-02	1.74E+02	2.73E-01	1.37E-01	2.07E+02	1.76E+00	1.13E+02	8.62E+01	7.61E+01	2.03E+01
Powell	**mean**	**1.21E-02**	2.59E+00	7.77E+02	8.77E-02	8.56E-01	4.60E+03	6.26E+00	1.71E+03	1.75E+03	1.42E+03	2.89E+02
	Min	3.18E-03	1.15E-02	3.28E+00	1.30E-02	1.83E-01	2.59E+03	**2.93E-03**	6.31E+02	8.10E+02	1.00E+03	1.67E+02
	Max	**8.52E-02**	7.52E+01	4.70E+03	5.07E-01	2.27E+00	7.27E+03	4.11E+01	3.10E+03	2.25E+03	1.98E+03	4.19E+02
	Std	1.48E-02	1.37E+01	1.25E+03	1.12E-01	5.23E-01	1.24E+03	1.02E+01	6.06E+02	3.40E+02	2.84E+02	5.49E+01

Table 12 The Wilcoxon rank sum test results of the benchmark functions

	PSO	MFO	CPSO	ALPSO	GPSO	DNLPSO	ELPSO	CFA1	CFA2	VSSFA
Ackley	1.41E-04	1.11E-06	2.60E-08	5.86E-06	3.02E-11	3.77E-04	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Dixon	1.56E-08	3.02E-11	2.15E-10	3.02E-11	3.02E-11	1.36E-07	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Griewank	3.02E-11	3.02E-11	3.02E-11	1.21E-12	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Rothyp	3.02E-11	3.02E-11	1.54E-01	3.02E-11	3.02E-11	5.27E-05	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Rosenbrock	4.86E-03	1.21E-10	1.12E-02	3.02E-11	3.02E-11	2.78E-07	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Sphere	3.02E-11	2.61E-02	3.02E-11	4.28E-06	3.02E-11	1.02E-01	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Sumpow	1.11E-06	3.02E-11	6.12E-10	2.48E-08	3.02E-11	5.53E-08	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Sumsqu	3.34E-11	3.02E-11	9.53E-07	6.71E-05	3.02E-11	3.99E-04	3.02E-11	3.02E-11	3.02E-11	3.02E-11
Powell	4.62E-10	3.02E-11	9.92E-11	3.02E-11	3.02E-11	4.44E-07	3.02E-11	3.02E-11	3.02E-11	3.02E-11

We conduct the Wilcoxon rank sum test to further indicate the model efficiency. Table 12 shows the statistical test results. The proposed PSO model depicts statistically significant superiority over all the baseline methods in nearly all the test cases. The exceptions are the Rotated Hyper-Ellipsoid and Sphere functions, where CPSO and

DNLPSO show similar distributions to those of the proposed model, respectively, over a set of 30 runs. ALPSO also shows a statistically better performance than the proposed model for the Griewank landscape.

Overall, the empirical and statistical test results of OD segmentation, DME classification, and numerical optimization ascertain the superiority of the proposed PSO model over other classical and advanced search methods in solving diverse optimization problems. The proposed accelerated and refined super-ellipse search operations, random and mean leader enhanced actions, and spherical local exploitation for swarm leader enhancement, account for the robustness and efficiency of the proposed PSO model.

6. CONCLUSION

In this research, we have proposed an ensemble transfer learning model with a PSO-based hyper-parameter selection method based on Mask R-CNN for OD segmentation. The proposed PSO model employs diverse elliptical search operations, a modified PSO mechanism, random and mean leader enhanced actions, and a spherical random walk mechanism for swarm leader enhancement. It enables the particles to follow the personal and global best experiences using either the accelerated or refined search steps along the irregular elliptical flights to diversify the search process. Moreover, it is capable of further increasing search diversification by assigning a distinctive search action to each particle in any iteration. The proposed PSO model is evaluated using the Drions and Messidor public fundus image data sets and demonstrates statistically superior capabilities in optimal hyper-parameter identification with respect to deep networks. The resulting ensemble segmenters achieve better average IoU scores in OD segmentation than those from three hybrid ensemble clustering methods as well as the deep ensemble networks formulated with default and optimal settings identified by the original and other PSO variants. The improved performances are shown through statistical test results. Furthermore, the proposed model also shows statistical superiority over other state-of-the-art search methods in solving diverse unimodal and multimodal benchmark optimization functions and the detection of DME.

In future studies, we aim to conduct hyper-parameter tuning and evolving clustering and deep network generation using the proposed model for solving other medical imaging problems, e.g. dermoscopic lesion localization and segmentation, nuclei counting and segmentation, MRI brain tumour segmentation, and stroke lesion segmentation based on acute CT perfusion images. We will also study the application of the proposed PSO model to other complex computer vision tasks, such as evolving deep architecture generation for image description and visual question generation [82-85] for resource-constrained deployments.

REFERENCES

[1] J. De Fauw, J.R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin and G. van den Driessche. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), p.1342. 2018.

[2] J. Cheng, J. Liu, D.W.K. Wong, F. Yin, C. Cheung, M. Baskaran, T. Aung and T.Y. Wong. Automatic optic disc segmentation with peripapillary atrophy elimination, in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 6224–6227. 2011.

[3] S. Morales, V. Naranjo, J. Angulo and M. Alcañiz. Automatic detection of optic disc based on PCA and mathematical morphology. *IEEE Transactions on Medical Imaging*, 32(4), pp.786-796. 2013.

[4] M.P. Sarathi, M.K. Dutta, A. Singh and C.M. Travieso. Blood vessel inpainting based technique for efficient localization and segmentation of optic disc in digital fundus images. *Biomedical Signal Processing and Control*, 25, pp.108-117. 2016.

[5] F. Yin, J. Liu, S.H. Ong, Y. Sun, D.W. Wong, N.M. Tan, C. Cheung, M. Baskaran, T. Aung and T.Y. Wong. Model-based optic nerve head segmentation on retinal fundus images. In *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 2626–2629. 2011.

[6] A. Li, Z. Niu, J. Cheng, F. Yin, D.W.K. Wong, S. Yan and J. Liu. Learning supervised descent directions for optic disc segmentation. *Neurocomputing*, 275, pp.350-357. 2018.

[7] B. Dai, X. Wu and W. Bu. Optic disc segmentation based on variational model with multiple energies. *Pattern Recognition*, 64, pp.226-235. 2017.

[8] B. Dashtbozorg, A.M. Mendonça and A. Campilho. Optic disc segmentation using the sliding band filter. *Computers in Biology and Medicine*, 56, pp.1-12. 2015.

[9] L.Y. Xue, J.W. Lin, X.R. Cao, S.H. Zheng and L. Yu. Optic Disk Detection and Segmentation for Retinal Images Using Saliency Model Based on Clustering. *Journal of Computers*, 29(5), pp.66-79. 2018.

[10] N. Sharma and A. Verma. Segmentation and Detection of Optic Disc Using Kmeans Clustering. *International Journal of Scientific & Engineering Research*, 6(8), pp.237-240. 2015.

[11] G. Hamednejad and H. Pourghassem. Retinal optic disk segmentation and analysis in fundus images using dbscan clustering algorithm. In *Proceedings of the 23rd Iranian Conference on Biomedical Engineering (ICBME)*. pp.122-127. IEEE. 2016.

[12] N. Thakur and M. Juneja. Optic disc and optic cup segmentation from retinal images using hybrid approach. *Expert Systems with Applications*, 127, pp.308-322. 2019

[13] H. Xie, L. Zhang, C.P. Lim, Y. Yu, C. Liu, H. Liu and J. Walters. Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing*, 84, p.105763. 2019.

[14] 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.

[15] T.Y. Tan, L. Zhang and C.P. Lim. Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. *Knowledge-Based Systems*. 2019.

[16] K. He, G. Gkioxari, P. Dollár and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969. 2017.

[17] K. He, G. Gkioxari, P. Dollár and R. Girshick. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.2018.2844175. 2018.

[18] Z.Q. Zhao, P. Zheng, S.T. Xu and X. Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*. Volume 30, Issue 11. pp.3212-3232. 2019.

[19] L. Jiao, F. Zhang, F., Liu, S. Yang, L. Li, Z. Feng and R. Qu. A Survey of Deep Learning-based Object Detection. *IEEE Access*, 7, pp.128837-128868. 2019.

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

[21] Y.T. Tan, L. Zhang, C.P. Lim, B. Fielding, Y. Yu and E. Anderson. Evolving ensemble models for image segmentation using enhanced particle swarm optimization, *IEEE Access*. 7 (2019) 34004–34019.

[22] D. Zhang, W. Zhu, H. Zhao, F. Shi and X. Chen. Automatic localization and segmentation of optical disk based on faster R-CNN and level set in fundus image. In *Medical Imaging 2018: Image Processing* (Vol. 10574, p. 105741U). International Society for Optics and Photonics. 2018.

[23] Z.U. Rehman, S.S. Naqvi, T.M. Khan, M. Arsalan, M.A. Khan and M.A. Khalil. Multi-parametric optic disc segmentation using superpixel based feature classification. *Expert Systems with Applications*, 120, pp.461-473. 2019.

[24] S.E. Abed, S.A. Al-Roomi and M. Al-Shayeji. Effective optic disc detection method based on swarm intelligence techniques and novel pre-processing steps. *Applied Soft Computing*, 49, pp.146-163. 2016.

[25] L. Wang, H. Liu, Y. Lu, H. Chen, J. Zhang and J. Pu. A coarse-to-fine deep learning framework for optic disc segmentation in fundus images. *Biomedical Signal Processing and Control*, 51, pp.82-89. 2019.

[26] J.H. Tan, U.R. Acharya, S.V. Bhandary, K.C. Chua and S. Sivaprasad. Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science*, 20, pp.70-79. 2017.

[27] M.N. Zahoor and M.M. Fraz. Fast optic disc segmentation in retina using polar transform. *IEEE Access*, 5, pp.12293-12300. 2017.

[28] Y. Liu, D. Fu, Z. Huang and H. Tong. Optic disc segmentation in fundus images using adversarial training. *IET Image Processing*. 13(2), pp.375-381. 2019.

[29] F. Guo, H. Peng, B. Zou, R. Zhao and X. Liu. Localisation and segmentation of optic disc with the fractional-order Darwinian particle swarm optimisation algorithm. *IET Image Processing*. Volume 12, Issue 8, pp. 1303-1312. 2018.

[30] G. Lim, Y. Cheng, W. Hsu and M.L. Lee. Integrated optic disc and cup segmentation with deep learning. In *Proceedings of IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI*. pp.162-169. IEEE. 2015.

[31] J. Kennedy and R. Eberhart. Particle Swarm Optimization. In *Proceedings of IEEE Int. Conf. Neural Networks*, vol. 4, pp.1942-1948. 1995.

[32] M. He, M. Liu, R. Wang, X. Jiang, B. Liu and H. Zhou. Particle swarm optimization with damping factor and cooperative mechanism. *Applied Soft Computing*, 76, pp.45-52. 2019.

[33] G. Xu, Q. Cui, X. Shi, H. Ge, Z.H. Zhan, H.P. Lee, Y. Liang, R. Tai and C. Wu. Particle swarm optimization based on dimensional learning strategy. *Swarm and Evolutionary Computation*, 45, pp.33-51. 2019.

[34] Y. Chen, L. Li, H. Peng, J. Xiao and Q. Wu. Dynamic multi-swarm differential learning particle swarm optimizer. *Swarm and Evolutionary Computation*, 39, pp.209-221. 2018.

[35] M. Isiet and M. Gadala. Self-adapting control parameters in particle swarm optimization. *Applied Soft Computing*, p.105653. 2019.

[36] B. Fielding, T. Lawrence and L. Zhang. Evolving and Ensembling Deep CNN Architectures for Image Classification. *International Joint Conference on Neural Networks*, 2019.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

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

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

[43] L. Zhang, K. Mistry, C.P. Lim and S.C. Neoh. Feature selection using firefly optimization for classification and regression models, *Decision Support Systems*. 106 (2018) 64–85.

[44] T.Y. Tan, L. Zhang, S.C. Neoh and C.P. Lim. Intelligent skin cancer detection using enhanced particle swarm optimization, *Knowledge-Based Systems.* 158 (2018) 118–135.

[45] T.Y. Tan, L. Zhang and C.P. Lim. Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models. *Applied Soft Computing*, p.105725. 2019.

[46] J. Gielis. A generic geometric transformation that unifies a wide range of natural and abstract shapes. *American Journal of Botany*, 90(3), pp.333-338. 2003.

[47] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.580–587. 2014.

[48] R. Girshick. Fast R-CNN. In *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, pp.1440–1448. 2015.

[49] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp.1137–1149. 2017.

[50] J. Dai, K. He and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3150-3158. 2016.

[51] Y. Li, H. Qi, J. Dai, X. Ji and Y. Wei. Fully convolutional instance-aware semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2359-2367. 2017.

[52] E.J. Carmona, M. Rincón, J. García-Feijoo and J. M. Martínez-de-la-Casa. Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine*, Vol. 43(3), pp. 243-259. 2008.

[53] E. Decenciére, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton and J.-C. Klein. Feedback on a publicly distributed image database: the messidor database, *Image Analysis & Stereology*. 33 (3) (2014) 231–234.

[54] U. of Huelva. Expert System for Early Aautomated Detection of DR by Analysis of Digital Retinal Images Project Website. Accessed on 16th Dec 2020. URL http://www.uhu.es/retinopathy/muestras2.php

[55] A. Giachetti, L. Ballerini and E. Trucco. Accurate and reliable segmentation of the optic disc in digital fundus images. *Journal of Medical Imaging*, 1(2), p.024001. 2014.

[56] J. Cheng, J. Liu, Y. Xu, F. Yin, D.W.K. Wong, N.M. Tan, D. Tao, C.Y. Cheng, T. Aung and T.Y. Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, 32(6), pp.1019-1032. 2013.

[57] D. Marin, M.E. Gegundez-Arias, A. Suero and J.M. Bravo. Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. *Computer Methods and Programs in Biomedicine*, 118(2), pp.173-185. 2015.

[58] A. Aquino, M.E. Gegúndez-Arias and D. Marín. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imaging*, 29(11), pp.1860-1869. 2010.

[59] H. Yu, E.S. Barriga, C. Agurto, S. Echegaray, M.S. Pattichis, W. Bauman and P. Soliz. Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *IEEE Transactions on Information Technology in Biomedicine*, 16(4), pp.644-657. 2012.

[60] S. Roychowdhury, D.D. Koozekanani, S.N. Kuchinka and K.K. Parhi. Optic disc boundary and vessel origin segmentation of fundus images. *IEEE Journal of Biomedical and Health Informatics*, 20(6), pp.1562-1574. 2015.

[61] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of International Conference on Learning Representations*. 2015.

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

[63] 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.

[64] 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.

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

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

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

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

[69] A. Hernández-García and P. König. Further advantages of data augmentation on convolutional neural networks. In *Proceedings of International Conference on Artificial Neural Networks*. pp.95-103. Springer, Cham. 2018.

[70] L. Giancardo, F. Meriaudeau, T.P. Karnowski, Y. Li, S. Garg, K.W. Tobin Jr. and E. Chaum. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical image analysis*, 16(1), pp.216-226. 2012.

[71] S.T. Lim, W.M.D.W. Zaki, A. Hussain, S.L. Lim and S. Kusalavan. Automatic classification of diabetic macular edema in digital fundus images. In *Proceedings of 2011 IEEE Colloquium on Humanities, Science and Engineering*. pp. 265-269. IEEE. 2011.

[72] M.U. Akram, M. Akhtar and M.Y. Javed. An automated system for the grading of diabetic maculopathy in fundus images. In *Proceedings of International Conference on Neural Information Processing*. pp.36-43. Springer, Berlin, Heidelberg. 2012.

[73] K.S. Sreejini and V.K. Govindan. Severity grading of DME from retina images: A combination of PSO and FCM with Bayes classifier. *International Journal of Computer Applications*. 81(16), pp.11-17. 2014.

[74] M.R.K. Mookiah, U.R. Acharya, V. Chandran, R.J. Martis, J.H. Tan, J.E.W. Koh, C.K. Chua, L. Tong and A. Laude. Application of higher-order spectra for automated grading of diabetic maculopathy. *Medical & biological engineering & computing*, *53*(12), pp.1319-1331. 2015.

[75] F. Ren, P. Cao, D. Zhao and C. Wan. Diabetic macular edema grading in retinal images using vector quantization and semi-supervised learning. *Technology and Health Care*, 26(S1), pp.389-397. 2018.

[76] B. Al-Bander, W. Al-Nuaimy, M.A. Al-Taee, B.M. Williams and Y. Zheng. Diabetic macular edema grading based on deep neural networks. In *Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop (OMIA)*, held in conjunction with MICCAI 2016, Athens, Greece, October 21, pp.121–128. 2016.

[77] J. Sahlsten, J. Jaskari, J. Kivinen, L. Turunen, E. Jaanio, K. Hietala and K. Kaski. Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading. *Scientific Reports* 9, Article number 10750, 2019. doi:10.1038/s41598-019-47181-w

[78] X. Li, X. Hu, L. Yu, L. Zhu, C.W. Fu and P.A. Heng. CANet: Cross-disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading. *IEEE Transactions on Medical Imaging*. 2019. DOI: 10.1109/TMI.2019.2951844

[79] Q. Chen, Y. Peng, T. Keenan, S. Dharssi, E. Agro, W. Wong, E. Chew and Z. Lu. A multi-task deep learning model for the classification of age-related macular degeneration. *AMIA Summits on Translational Science Proceedings*, vol. 2019, p.505, 2019.

[80] 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.

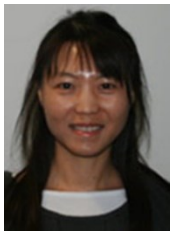
[81] X.S. Yang. *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, UK. 2008.

[82] P. Kinghorn, L. Zhang and L. Shao. Deep learning based image description generation. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. pp.919-926. IEEE. 2017.

[83] T. Lawrence and L. Zhang. IoTNet: An Efficient and Accurate Convolutional Neural Network for IoT Devices. *Sensors*. 19(24), p.5541. 2019.

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

[85] P. Kinghorn, L. Zhang and L. Shao. A Hierarchical and Regional Deep Learning Architecture for Image Description Generation. *Pattern Recognition Letters.* 119 (2019) 77-85.



Li Zhang is currently an Associate Professor (Reader) in Computer Science in Northumbria University, UK. She holds expertise in deep learning, machine learning, intelligent robotics and evolutionary computation. She gained her PhD and postdoctoral experience from University of Birmingham previously. Dr Zhang is an Associate Editor for Decision Support Systems.

Chee Peng Lim received the Ph.D. degree in intelligent systems from the University of Sheffield, Sheffield, U.K., in 1997. He is currently Professor of Complex Systems with the Institute for Intelligent Systems Research and Innovation, Deakin University, Australia. His research interests include computational intelligence-based systems for data analytics, condition monitoring, optimization, and decision support.