Intelligent Skin Cancer Detection Using Enhanced Particle Swarm Optimization

Teck Yan Tan[1], Li Zhang[1], Siew Chin Neoh[2] and Chee Peng Lim[3]

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

[2]School of Computer Science
Nottingham University (Malaysia Campus)
Malaysia

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

Email: **{teck.tan; li.zhang}@northumbria.ac.uk; u_jane80@yahoo.co.uk; chee.lim@deakin.edu.au**

Abstract.
In this research, we undertake intelligent skin cancer diagnosis based on dermoscopic images using a variant of the Particle Swarm Optimization (PSO) algorithm for feature optimization. Since the identification of the most significant discriminative characteristics of the benign and malignant skin lesions plays an important role in robust skin cancer detection, the proposed PSO algorithm is employed for feature optimization. It incorporates not only subswarms, local and global food and enemy signals, attraction and flee operations, and mutation-based local exploitation, but also diverse matrix representations to mitigate premature convergence of the original PSO algorithm. Specifically, two remote swarm leaders, which show similar fitness but low position proximity, are used to lead the subswarm-based search and to enable the exploration of more distinctive search regions. Modified velocity updating strategies are also proposed to enable the particles to follow multiple swarm leaders and avoid the local and global worst individuals, partially (i.e. in randomly selected sub-dimensions) and fully (in every dimension), with an attempt to search for global optima. Probability distribution and dynamic matrix representations are used to diversify the search process. Evaluated with multiple skin lesion and UCI databases and diverse unimodal and multimodal benchmark functions, the proposed PSO variant shows a superior performance over those of other advanced and classical search methods for identifying discriminative features that facilitate benign and malignant lesion classification as well as for solving diverse optimization problems with different landscapes. The Wilcoxon rank sum test is adopted to further ascertain superiority of the proposed algorithm over other methods statistically.

Keywords: Skin Cancer Detection, Feature Selection, Evolutionary Algorithm, and Dermoscopic Images

1. INTRODUCTION

Melanoma is an aggressive type of skin cancer that can spread to other organs. Automatic and early diagnosis of melanoma is essential for administering effective treatment and increasing survival chances. Since medical skin cancer diagnosis employs the Asymmetry, Border, Colour, Diameter and Enlargement (ABCDE) guideline, the extraction and identification of such discriminative and significant morphological characteristics play a crucial role in attaining accurate diagnosis rates. However, it is still a challenging task for the retrieval of such distinguishing attributes, owing to the fine-grained variability in the appearance of benign and cancerous skin lesions [1].

This research aims to deal with the above challenge by proposing an enhanced Particle Swarm Optimization (PSO) algorithm for discriminative feature selection in skin cancer diagnosis using dermoscopic images. The main motivations of this research are as follows. Evolutionary algorithms possess powerful search capabilities, and have been widely used for solving various feature selection challenges. Owing to the comparatively simple underlying concepts and relatively few user-defined parameters, PSO has been widely studied for feature selection tasks. Since PSO has a weak exploration capability, and its search process is likely to be trapped in local optima when dealing with multimodal or complex optimization problems, new PSO variants with superior explorative capabilities are required. Therefore, we propose an enhanced PSO model in this research.

Specifically, our proposed PSO model incorporates the subswarm concept, food and enemy signals, attraction and flee operations, mutation-based local exploitation, and diverse matrix representations to mitigate premature convergence of the original PSO algorithm. It shows a great superiority over other methods for the identification of the most significant characteristics of benign and malignant lesion images to facilitate subsequent skin cancer classification. The proposed skin cancer detection system consists of five key stages, i.e. pre-processing, skin lesion segmentation, feature extraction, PSO-based feature selection and classification. The overall system architecture is illustrated in Figure 1.

The key contributions of this research, which focus on PSO-based feature selection, are as follows.

- The proposed PSO model first of all divides the original population into two subswarms. Then, two swarm leaders with competitive fitness scores but low position proximity are identified. Each leader leads one subswarm-based search for discriminative lesion feature selection. Since the subswarm-based search is more likely to explore distinctive search regions owing to the low position correlation between the two leaders, it reduces the probability of being trapped in local optima and increases the chances of finding the global optimum.
- A number of attraction and enemy signals are used for velocity updating in each subswarm. The proposed PSO model enables each particle to follow the leader (i.e. food attraction) and avoid unpromising solutions (i.e. enemies) partially (i.e. in randomly selected sub-dimensions) as well as fully (i.e. in each dimension) to diversify the search process. Three random walks, i.e. Gaussian, Cauchy, and Levy distributions, are used to further enhance the best subswarm solution and to increase exploitation. A dynamic matrix representation of the swarm is also utilized during the search process to increase search diversity. The proposed algorithm shows great efficiency in optimal feature selection for melanoma classification, as well as solving unimodal and multimodal benchmark problems in comparison with other search methods. It is also among the top performers for skin cancer detection in comparison with related research studies reported in the literature.

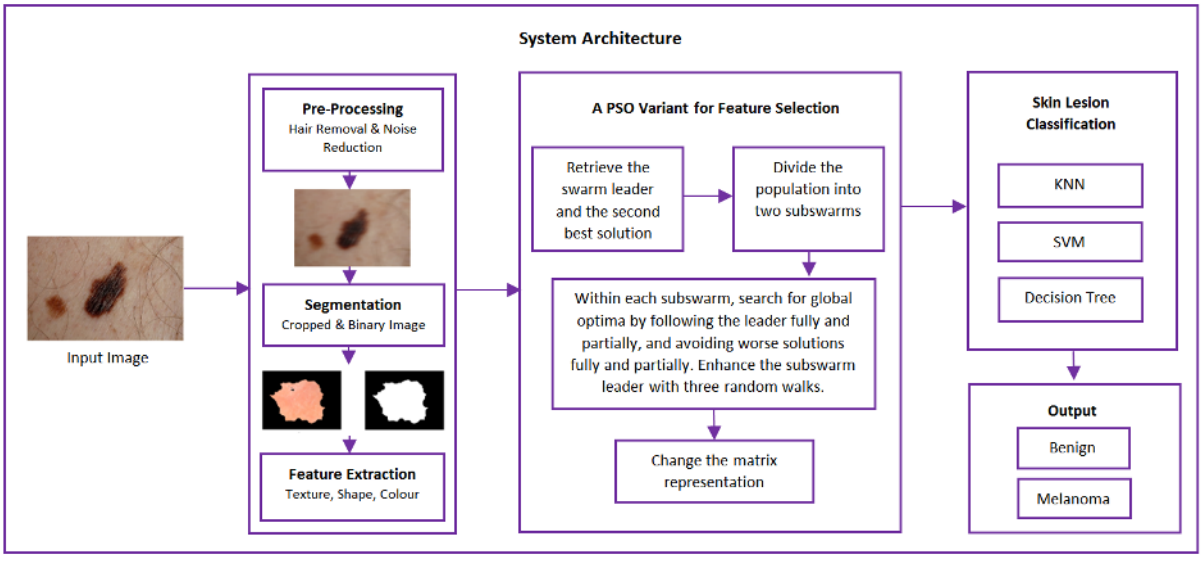

Figure 1 The system architecture for skin cancer detection

The paper is organised as follows. Section 2 presents the related studies on skin cancer detection, enhanced PSO variants and diverse evolutionary algorithm-based feature selection techniques. Section 3 presents the proposed PSO model with mutation-based local exploitation as well as attraction and flee based global exploration. A detailed evaluation of the proposed algorithm and other classical methods using skin lesion data sets and other

benchmark problems is presented in Section 4. Concluding remarks and suggestions for further research are provided in Section 5.

2. RELATED WORK

In this section, we discuss the related work on computerized skin cancer diagnosis, diverse variants of the PSO algorithm, and evolutionary algorithm-based feature selection methods.

2.1 Skin Cancer Detection

Ain et al. [2] proposed a skin cancer detection system using Genetic Programming (GP) based feature selection from dermoscopic images. Their work extracted both high-level domain specific features recommended by the dermatologists and low-level Local Binary Pattern (LBP) features. GP was used to identify the most significant features from the raw feature set to support subsequent benign and malignant cancer detection. Abuzaghleh et al. [3] proposed an intelligent system for early detection and prevention of malignant lesions such as melanoma. Their method consisted of a prevention component and a detection component. The former integrated a novel equation to identity the risk of skin burn and generated alerts when necessary, while the latter comprised several key stages for the classification of different lesion types, including noise removal, lesion segmentation, feature extraction and classification. Evaluated with the PH2 Dermoscopy image database from Pedro Hispano Hospital, the proposed method showed impressive performance for the classification of benign, atypical, and melanoma cases. Xie et al. [4] proposed a neural network ensemble classifier for melanoma classification using dermoscopy images. Their work first of all segmented the lesions using a self-generating neural network. The feature extraction process extracted border irregularities from both complete and incomplete lesions. The proposed meta-ensemble classifier combined three ensemble models with different network topologies and base model types. The first ensemble model was composed of a set of networks of the same type and structure, while the second and third ensemble models were built with different types of networks and networks with diverse topologies, respectively. Their work showed impressive performance when evaluated with two dermoscopy databases with images from xanthous and caucasian races.

Esteva et al. [1] performed skin cancer classification using deep neural networks. Their work employed an end-to-end convolutional neural network (CNN), where pixels and disease types were used as the training inputs. The CNN model was trained with a total of 129,450 clinical images with 2,032 different diseases. Evaluated against board-certified dermatologists on two binary classification tasks using clinical images, i.e., benign versus malignant (i.e. the deadliest cancer) and benign versus keratinocyte carcinomas (i.e. common cancers)), their CNN model achieved a comparable performance with those of dermatologists. Shimizu et al. [5] developed a digital diagnosis system for the classification of four skin lesion conditions including melanoma, nevus, basal cell carcinoma (BCC), and seborrheic keratosis (SK). In their work, lesions were categorized into two general categories, i.e. melanocytic skin lesions (MSLs) and nonmelanocytic skin lesions (NoMSLs). Melanoma and nevus belong to MSLs while BCC and SK belong to NoMSLs. The detection of NoMSLs has been rarely addressed in the literature, despite their high occurrences. Their work was therefore dedicated to the identification of both MSLs and NoMSLs. It extracted a total of 828 features representing colour, sub-region, and texture information. A layered classification method was used to conduct the four-class lesion classification. Their classification model firstly categorized MSL and NoMSL using a binary classifier. Another two binary classifiers were subsequently used to distinguish melanoma from nevus and BCC from SK, respectively. Tan et al. [6] developed an intelligent decision support system for melanoma detection. Their work consisted of several key stages and extracted a high-dimensional feature vector integrating shape, colour, and texture information owing to their high correlation with clinical characteristics associated with skin cancer identification. The Genetic Algorithm (GA) was subsequently used to identify the most contributing factors for diagnosis. The work showed performance improvements in comparison with those from other related work reported in the literature. Doukas et al. [7] conducted automatic skin lesion assessment based on cloud and smartphone platforms. Their work employed the Active Shape Model to extract texture, shape, and size features. A set of classifiers, including Neural Network (NN), Support Vector Machine (SVM), Decision Tree (DT), was used to conduct the classification of different lesion types such as melanoma, dysplastic and common (benign) nevus. There are also many other developments on automatic skin cancer diagnosis in the literature. As an example, Barata et al. [8] employed a probabilistic Correspondence LDA algorithm to extract medically inspired colour information for skin lesion classification using the EDRA database, whereas Glaister et al. [9] and Glaister et al. [10] proposed a multistage illumination modelling algorithm for illumination correction and a texture-based skin lesion segmentation model, respectively.

2.2 PSO Variants

Many PSO variants have been proposed to overcome premature convergence of the original PSO model. Lynn and Suganthan [11] proposed an ensemble of PSO algorithms. Their ensemble model utilized a pool of PSO variants including inertia weight PSO, comprehensive learning PSO (CLPSO), and distance-based locally informed PSO (LIPS). The method divided the swarm into two subpopulations, i.e. one large and one small subswarms. CLPSO was applied to the small subpopulation to maintain swarm diversity whereas a self-adaptive probabilistic selection scheme was used to identify the best PSO algorithm from the pool in each iteration using the large subswarm. In their method, if a PSO variant was able to achieve performance improvements within a fixed learning period, it was stored in a success memory, otherwise it was recorded in the failure memory. The PSO strategies with higher success rates were more likely to be selected to guide the search of the large subswarm. Gou et al. [12] developed a PSO variant based on individual difference evolution. Motivated by the social and psychological models, their work treated each particle as a virtual human, and proposed an emotional status fitness indicator for each particle. It not only determined a specific evolutionary mechanism for each particle based on its emotional status and current fitness, but also utilized a re-starting mechanism to increase swarm diversity. Optimal parameter settings were also explored in their work. Dash et al. [13] developed two hybrid evolutionary models, known as CSPSO and ICSPSO, respectively. The search strategies of PSO and Cuckoo Search (CS) were integrated to increase local exploitation and global exploration of the original PSO algorithm. Both proposed methods enabled each particle to conduct position updating using the original PSO mechanism. A new nest was then generated using the Levy flight operator in CS from the position vector of a current particle if this particle was ranked within the best 4% of the population. In ICSPSO, the mutation and crossover operations of Differential Evolution (DE) were integrated with CSPSO to further enhance the global best solution in each iteration. Both algorithms were used for the design of linear phase multi-band stop filters and the search of the desired impulse responses of the filters.

Ouyang et al. [14] proposed another PSO variant, known as an improved global-best-guided PSO with learning operation (IGPSO), for solving engineering design optimization problems. In their work, the overall swarm was divided into three categories, i.e. current population, historical best population, and global best population. A specific search strategy was dedicated to each subpopulation. Besides using the original PSO operation to guide the search of the current population, a Gaussian distribution based local exploration mechanism was applied to the historical best subswarm to enable each personal best particle to learn from other more promising historical best particles independently. Stochastic learning and opposition based learning operations were also used to further enhance the global best solution and to overcome the local optima traps. IGPSO outperformed other PSO variants and classical search methods when evaluated using a total of 25 unimodal and multimodal benchmark optimization functions. Nasir et al. [15] proposed a PSO variant called the Dynamic Neighbourhood Learning Particle Swarm Optimizer (DNLPSO). It employed a neighbourhood-based learning strategy where the neighbourhood historical best information was used for velocity updating of the current particle. The neighbourhoods were also updated and reconstructed dynamically after a certain number of iterations to preserve swarm diversity. An Enhanced Leader PSO (ELPSO) was proposed by Jordehi [16]. A five-staged mutation mechanism was used to enhance the swarm leader and to overcome stagnation. On the other hand, an autonomous group PSO (AGPSO) model was developed by Mirjalili et al. [17]. Adaptively decreasing cognitive parameter and increasing social parameter were employed to balance between exploitation and exploration. Chen et al. [18] proposed a biogeography-based learning PSO (BLPSO) model to further enhance CLPSO. Both CLPSO and BLPSO performed velocity updating of each particle using the personal best solutions of different exemplar particles for different dimensions. In comparison with CLPSO where learning probability and tournament selection were used to generate exemplars for each particle, BLPSO employed the migration of biogeography-based optimization for exemplar generation. BLPSO achieved impressive performance for the evaluation of diverse challenging benchmark functions.

2.3 PSO-based and Other Feature Selection Methods

Diverse state-of-the-art PSO variants have been proposed in recent years for feature selection and dimensionality reduction. As an example, the original PSO model was employed for rough set-based feature selection in Wang et al. [19]. Their work outperformed GA-based and other deterministic rough set reduction methods. Overall, the attraction-based operation of the original PSO model was used in their work to guide rough set reduction. Comparatively, the proposed PSO model in this research incorporates not only attraction but also flee operations for discriminative feature selection. Zhang et al. [20] developed a binary Bare-bones PSO (BBPSO) algorithm for feature optimization. A reinforced memory mechanism was proposed to update the personal best solutions by avoiding gene degradation, while a uniform combination was used to increase swarm diversity. The 1-Nearest Neighbour classification model was employed to evaluate the classification performance of each particle. In their work, the original BBPSO operation, i.e. the Gaussian distribution based on personal and global best experiences, was utilized for position updating. Comparatively, the proposed PSO model in this research incorporates both good and poor signals to facilitate velocity updating and accelerate

convergence. Zhang et al. [21] proposed a multi-objective PSO for cost-based feature optimization to retrieve a Pareto front of non-dominated solutions (i.e. feature subsets) with high classification accuracy and low computational cost. Specifically, their work incorporated a multi-objective PSO with hybrid mutation as well as the crowding distance, external archive, and Pareto dominance relations. Firstly, the crowding distance integrated with the Pareto dominated comparison was used to update the external archive. Then, a binary tournament strategy with the crowding distance was employed for the identification of the global best solution while a domination-based mechanism was utilized for personal best updating. In their work, the original PSO operation led by both personal and global best experiences was used to guide the search process. Both re-initialization and jumping mutation operators were used to increase swarm diversity and improve the global search capability. Several UCI (University of California, Irvine) Machine Learning data sets were used for evaluation. Their model outperformed DE-based, NSGA-based and SPEA2-based multi-objective feature selection algorithms. As mentioned earlier, the velocity updating of each particle was conducted using the original PSO operation. Comparatively, in the proposed PSO model in this research, both the original PSO operation and an additional flee action are used for velocity updating.

Mistry et al. [22] incorporated micro-GA with PSO for discriminative facial feature selection and facial expression recognition. Their model embedded a small-population secondary swarm, sub-dimension based search, and a velocity updating mechanism with the consideration of average personal best experiences and Gaussian distribution enhanced global best solutions to lead the search process and attain the global optimum solution. Evaluated using CK+ and MMI databases, their proposed model outperformed other PSO variants and classical search methods significantly. The work could also be further improved by incorporating hybrid search mechanisms such as adaptive chaotic attraction or evading actions to increase search diversity. Ahila et al. [23] integrated a discrete-valued PSO with a continuous-valued PSO for selection of the input feature subset and identification of the optimal number of hidden nodes for the extreme learning machine (ELM). Their improved ELM model showed an impressive performance for power system disturbances classification. Since the work mainly relies on the original PSO operation to guide the search process, it is more likely to be trapped in local optima. Moradi and Gholampour [24] integrated PSO with a local search mechanism for feature optimization. The local search operation took the feature correlation information into account to increase swarm diversity. This led to a higher probability of selecting less correlated features in comparison with more correlated ones. Specifically, two operators, i.e. ‘Add’ and ‘Delete’, were proposed for the above operations. The former was used to incorporate dissimilar features into the particle while the latter was employed to discard similar ones from a particle. A subset size determination scheme was also used for the identification of discriminative features. A total of 13 benchmark data sets were used for evaluation. In comparison with popular filter-based feature selection methods such as information gain and fisher score, and wrapper-based feature selection models such as PSO, Simulated Annealing (SA), GA and several PSO variants, their hybrid PSO model showed statistically better results. Sheikhpour et al. [25] employed the original PSO model for feature selection and kernel bandwidth determination in kernel density estimation (KDE) based classifiers for breast cancer diagnosis. Several UCI breast cancer data sets were used for evaluation. Their work outperformed the GA-KDE model in breast cancer diagnosis.

Krisshna et al. [26] developed a Threshold-based binary PSO (ThBPSO) model for facial feature selection and face recognition. In ThBPSO, the original PSO model was performed multiple times (e.g. k times) to obtain multiple global best solutions. Subsequently, the significance of a feature was determined by the frequency that it was selected by the global best solutions. A threshold value ($1 \leq$ threshold $\leq k$) was used as the criterion for the selection of a specific feature dimension. Evaluated with seven benchmark face recognition data sets, their work showed a competitive performance. Despite that ThBPSO conducted multiple runs of the original PSO model for retrieving the final optimal output, its search process purely relied on the original PSO operation, therefore having a high probability of premature convergence. Chang [27] proposed a PSO variant with multiple sub-populations, denoted as MS-PSO, for solving multimodal optimization problems. Specifically, MS-PSO divided the original swarm into several sub-populations. In each subswarm, a modified PSO operation was used for velocity updating, where the subswarm leader was employed to replace the original global best solution to guide the search process. Evaluated with unimodal and multimodal artificial landscapes, MS-PSO showed an impressive performance in terms of finding the global minimum solution with a fast convergence speed. MS-PSO was employed in [22] for facial feature selection and facial expression recognition. However when its attraction-based search process converges prematurely, there is no alternative strategy to overcome stagnation. A modified PSO model, which incorporated PSO with the GA and probability distributions, known as GM-PSO, was developed by Zhang et al. [28] for static and dynamic bodily expression feature optimization. After performing several rounds of the PSO operations, GM-PSO split the original swarm into two subswarms randomly. GA and Gaussain as well as Cauchy and Levy distributions were used to improve the subswarm leaders, respectively. Since GM-PSO depends on the original PSO operation to guide the search of the primary

swarm, it is prone to local optimum traps. BBPSO variants (denoted as BBPSOV) were developed in Srisukkham et al. [29] for discriminative feature selection for leukaemia diagnosis using microscopic images. Two position updating strategies, i.e. a chaotic accelerated attraction action and an additional enemy avoidance operation, were proposed to increase search diversity of the original BBPSO model. The mean of local and global promising solutions and the mean of the personal and global worst experiences were used to lead the attraction and enemy avoidance actions, respectively. Their work showed impressive performance for the retrieval of distinctive nucleus and cytoplasm attributes for leukaemia diagnosis. As described earlier, since their search process was guided by the mean of good and poor signals respectively (as compared with multiple individual promising and weak signals with different weights in the proposed model in this research), their model shows less discriminative capabilities, and tends to select a larger set of features as compared with those of this research. Their BBPSO model also sometimes converges slower when solving complex optimization problems (e.g. mathematical artificial landscapes) in comparison with our proposed method and other search methods.

Moreover, Shang et al. [30] conducted sentiment classification using a modified binary PSO (BPSO) model for feature selection. Known as F-BPSO, a mutation rate and a fitness proportionate selection strategy were employed for velocity updating and mitigating premature convergence of the original BPSO model. Specifically, F-BPSO divided a set of involved particles (i.e. the current particle, the personal best and global best solutions) into two groups based on their binary values (i.e. 0 or 1) of the d-th dimension. Then, the average fitness value of all particles in each group was calculated for the respective dimension. The group with a higher fitness value was used to set the position value of the d-th dimension. In this way, the position value of each particle was determined bit by bit. As a result, their model was able to focus more on each single feature dimension in comparison with other wrapper-based methods for feature selection. An enhanced F-BPSO model, known as FS-BPSO, was also proposed. In order to make the 'voting' mechanism of F-BPSO more robust, the previous set of particles was extended by adding the second personal best and the second global best solutions. Instead of using the average fitness value in fitness proportionate selection, the summation of all fitness values for each group was applied. Evaluated using two UCI benchmark data sets, FS-BPSO outperformed F-BPSO and BPSO significantly for feature optimization. A Genetic PSO (GPSO) model was proposed by Chen et al. [31] for feature selection in remotely sensed imagery object change detection. The genetic crossover operation was used to diversify the population in each iteration of GPSO. On the other hand, different meta-heuristic methods such as the firefly algorithm were employed for feature selection [32] and classifier ensemble reduction [33] with respect to classification and regression problems.

Besides the abovementioned evolutionary algorithm-based feature selection methods, there are other non-evolutionary algorithm-based feature selection methods proposed in the literature in recent years. Shang et al. [34] developed a subspace learning-based graph regularized feature selection method. Their model employed the graph theory to preserve the geometric structure information of the feature manifold. It also sustained the sparsity of the feature selection matrix by using an $\boldsymbol{L}_{2,1}$-norm sparse constraint. Their model outperformed several unsupervised discriminant feature selection methods based on 12 biological and digital image benchmark data sets. Shang et al. [35] proposed a self-representation-based dual-graph regularized feature selection clustering (DFSC) model. Unlike other clustering-based unsupervised feature selection methods, DFSC exploited the local geometrical structure of both the data space and feature space. Shang et al. [36] developed global discriminative-based nonnegative and spectral clustering models to preserve both the global geometrical and discriminative information of the data. Yang et al. [37] employed a Coupled Compressed Sensing inspired Sparse Spatial-Spectral Least Square SVM (CCS4-LSSVM) for hyperspectral image classification. CCS4-LSSVM was able to enhance the classification accuracy of the original Least Square SVM (LSSVM) and to minimize the influence of noisy pixels by integrating spectral and adaptively extracted spatial information. It outperformed Spatial-Spectral SVM and Spatial-Spectral LSSVM for the evaluation of several hyperspectral image data sets.

The differences between some of the abovementioned PSO-based feature selection methods and this research are summarized in Table 1.

Table 1 Summary of differences between related PSO-based feature selection methods and this research

Studies	Multiple leaders	Subswarm division	Modified attraction-based velocity updating	Other velocity updating mechanisms	Leader enhancement	Other strategies
Wang et al. [19]	No	No	No (the original PSO operation is	No	No	No

			used)			
Zhang et al. [20]	No	No	No (the original BBPSO operation is utilized)	No	A reinforced memory mechanism is used to update the personal best solutions.	A uniform combination is used to increase swarm diversity.
Mistry et al. [22]	No	A small secondary swarm is used.	Yes (the average personal best experience and Gaussian distribution enhanced global best solutions are used for velocity updating)	No	Gaussian distribution enhanced global best solutions	Sub-dimension based search
Ahila et al. [23]	No	No	No (the original PSO operation is used)	No	No	No
Moradi and Gholampour [24]	No	No	No (the original PSO operation is used)	No	No	"Add" and "Delete" operators are used to improve the local search of each particle.
Sheikhpour et al. [25]	No	No	No (the original PSO operation is used)	No	No	No
Krisshna et al. [26] (ThBPSO)	Yes (obtained by multi-runs of the original PSO model)	No	No (the original PSO operation is used)	No	No	No
Chang [27] (MS-PSO)	No	Yes	Yes (for each subswarm-based search, the subswarm leader is used to replace the original global best solution.)	No	No	No
Zhang et al. [28] (GM-PSO)	No	Yes	No (the original PSO operation is used)	No	Yes (GA and probability distributions are used to enhance subswarm leaders respectively)	No
Srisukkham et al. [29] (BBPSOV)	No	No	Yes (the search is guided by the average of the local and global optimal signals.)	The mean of the local and global worst indicators is used to lead the evading action.	No	No
Shang et al. [30]	Besides the personal and global best solutions, the second personal best and the second global best solutions are considered.	No	Yes (the position value of each particle is determined bit by bit based on a fitness proportionate selection strategy.)	No	No	No
Nasir et al. [15] (DNLPSO)	No	No	Yes (its own p_{best} or p_{best} of other particles in the swarm is used for velocity updating.)	No	No	No
Chen et al. [31] (GPSO)	No	No	No (the original PSO operation is used)	No	No	The crossover operation is used to diversify the population in each iteration.
Jordehi [16] (ELPSO)	No	No	No (the original PSO operation is	No	A five-staged mutation	No

			used)		mechanism is used to enhance the swarm leader.	
Mirjalili et al. [17] (AGPSO)	No	No	Yes (Adaptively decreasing cognitive parameter and increasing social parameter are used.)	No	No	No
This research	Multiple optimal and unpromising signals are used to guide the search process.	Yes	Yes (for each subswarm-based search, the subswarm leader is used to replace the original global best solution. The updated PSO operation is conducted in not only each dimension, but also any randomly selected sub-dimensions.)	The search avoids personal and global worst experiences in every dimension and any randomly selected sub-dimensions.	Gaussian, Cauchy, and Levy distributions are used to further enhance the best subswarm solutions.	A dynamic matrix representation of the swarm is also used.

As indicated in Table 1, most of the related methods employ either the original PSO velocity updating operation [16, 19, 23-26, 28, 31] or the original BBPSO position updating action [20] to lead the search process. Mistry et al. [22], Nasir et al. [15], Mirjalili et al. [17] and Chang [27] employed modified velocity updating strategies of the original PSO model. The above original and modified search operations are led by personal best experiences and global best solutions. When such attraction driven search mechanisms stagnate or converge prematurely, there is no alternative mechanism to drive the search out of the local optima traps. In other words, there is no effective strategy to avoid poor solutions while moving towards the optimal ones.

In order to deal with the abovementioned challenge, the proposed PSO model employs a two-tier strategy, i.e. an updated PSO operation and an additional flee action for velocity updating. In comparison with [29], the updated attraction-based PSO and additional flee operations in the proposed model are guided by multiple swarm leaders as well as the local and global worst individuals, respectively. Moreover, subswarm-based search, dynamic matrix representation, and enhancement of the subswarm leaders using probability distributions are adopted to improve search diversity and avoid stagnation. A summary of the distinctive aspects of our research is as follows.

1. Two remote leaders with competitive fitness scores but low position proximity are identified to guide the subswarm-based search. Since the subswarm-based search is more likely to explore distinctive search regions owing to the low position correlation between the two leaders, it reduces the probability of being trapped in local optima and increases the probability of finding the global optimum solution.

2. In comparison with the abovementioned studies, both the updated attraction-based PSO operation and the additional evading action are used for velocity updating in the proposed model. These attraction and flee operations are guided by a number of attraction and enemy signals, respectively. Moreover, different from [29], both operations are conducted in several randomly selected sub-dimensions as well as in every dimension, in order to improve search diversity and avoid local optima traps.

3. Three random walks, i.e. Gaussian, Cauchy, and Levy distributions, are used to further enhance the subswarm leader and increase exploitation. A dynamic matrix representation of the swarm is also utilized during the search process to increase search diversity.

In essence, the updated attraction-based PSO operation and the additional flee action work in a collaborative manner to avoid premature convergence and attain global optimality. When the modified PSO operation cannot find a fitter solution, the flee action is able to drive the search towards the optimal search space by avoiding unpromising regions. On the other hand, when the flee action stagnates, the updated PSO operation is capable of guiding the swarm particles to move towards promising regions by following both local and global optimal solutions.

3. THE PROPOSED SKIN CANCER DETECTION SYSTEM

We propose an intelligent system for benign and malignant skin lesion classification. The proposed system consists of five key stages, i.e. pre-processing, skin lesion segmentation, feature extraction, PSO-based feature optimization and classification. Each key stage, especially the feature selection process, is explained comprehensively, as follows.

3.1 Feature Extraction

First of all, pre-processing techniques are applied to lesion images for noise filtering, image segmentation, and grayscale conversion. Specifically, to remove 'salt and pepper' noise, median filtering is performed by diminishing the effects of thin hairs and air bubbles. A ground truth motivated image segmentation algorithm is used to separate the lesion from the skin. Subsequently, grayscale conversion transforms the original RGB images into grayscale.

Several feature extraction methods are used to extract shape, colour, and texture features from the separated lesion regions. According to the ABCDE guideline for clinical skin cancer diagnosis, three categories of features play very important roles in distinguishing between benign and malignant lesions, namely (1) shape features such as asymmetry, border irregularity, and compactness; (2) colour features such as relative chromaticity, and differences in lightness and colour; (3) Gray Level Run Length Matrix (GLRLM) based texture features. As such, these shape, colour, and texture features are extracted in this research.

Specifically, owing to the significance of asymmetry to benign/malignant lesion classification, an asymmetry index is generated by identifying the areas comprising the inner and outer of lesion. A border irregularity index is also produced to identify small irregularities in the edges. Other morphological features include compactness, radial variance, perimeter, solidity, roundness, extent, equivalent diameter, form factor and difference of left & right. Besides the abovementioned morphological features, the RGB colour space is employed for feature extraction. Colour features such as relative chromaticity and ratio of red, green and blue, and factors exhibited with respect to the lesion's tone are extracted. The following colour features are also retrieved, i.e. variance, entropy, skewness, correlation, Principal Component Analysis (PCA) variance, mean of image darkness, variance of image darkness, mean & standard for both lesion and skin, and average colour of red, green and blue. Moreover, four orientations (0, 45, 90, and 135) of the GLRLM-based texture features are retrieved, whereby each level embeds 11 different emphases. These 11 statistics include Short Run Emphasis, Long Run Emphasis, Gray-Level Nonuniformity, Run Length Nonuniformity, Run Percentage, Low Gray-Level Run Emphasis, High Gray-Level Run Emphasis, Short Run Low Gray-Level Emphasis, Short Run High Gray-Level Emphasis, Long Run Low Gray-Level Emphasis, and Long Run High Gray-Level Emphasis. Two more texture features pertaining to Tamura Coarseness Indexes for both skin and lesion are also obtained. Overall, we extract a total of 146 features with 13 morphological, 87 colour, and 46 textural features for representing the lesion region.

Since the features are not equally important for the identification of lesion types, we propose a PSO model for feature optimization and dimensionality reduction. The aim is to identify the most discriminative features and remove redundant ones. The proposed feature optimization algorithm is discussed in detail, as follows.

3.2 The Proposed PSO Model for Feature Selection

In this research, a new PSO model is proposed to identity the most significant distinguishing characteristics of benign and malignant skin lesions. It incorporates partial and full attraction and flee operations, mutation-based local exploitation, and diverse matrix representations to mitigate premature convergence of the original PSO model. To diversify the search process, the proposed model enables each particle to follow multiple swarm leaders and avoid local and global worst individuals partially as well as fully to search for the global optimum solution. The proposed algorithm shows great superiority of selecting features discriminatively and attaining the global optimum solution. The original and the proposed PSO models are explained in detail in the following sub-sections.

3.2.1 The Original PSO Algorithm

PSO [38] is one of the classical swarm intelligence-based algorithms. It employs personal and global best experiences of the swarm to guide the search process. The following equations are used for position and velocity updating of each particle.

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

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

where $x_{id}^{t}$and x_{id}^{t+1} represent the positions of the *i*-th particle during the *t*-th and *t*+1-th iterations in the *d*-th dimension, respectively while v_{id}^{t} and v_{id}^{t+1} denote the velocities of the *i*-th particle during the *t*-th and *t*+1-th iterations, respectively. *w* is an inertia weight that determines the influence of the previous velocity over iterations. The two acceleration constants are represented by c_1 and c_2, while r_1 and r_2 denote random vectors with each element within the range of [0, 1]. In addition, p_{id} and p_{gd} represent the personal best experience, p_{best} and the global best experience, g_{best} in the *d*-th dimension, respectively. PSO shows superior capabilities of dealing with diverse optimization problems. However, since its search process relies on a single global best solution, it is more likely to be trapped in local optima.

3.2.2 The Proposed Search Strategies

In this research, we propose a PSO variant with mutation-based local exploitation and attraction and evading driven global exploration along with the subswarm concept to mitigate the local optimum traps. The overall algorithm is illustrated in Algorithm 1. First of all, a swarm of 50 particles is initialized. Then, the particles are ranked based on their fitness scores. We archive the global best leader and three worst solutions in the best-memory and worst-memory structures, respectively. A second swarm leader which has a competitive fitness but low correlation to the primary swarm leader is retrieved. The overall swarm is subsequently divided into two subpopulations. The two swarm leaders are used to lead the search process in each subswarm, respectively.

In each subswarm, if an individual particle is one of the previously identified archived worst particles in the memory, the original velocity updating strategy of PSO is applied. In other words, it follows the subswarm leader fully for position updating in each dimension, as defined in Equation (1). Otherwise, the algorithm performs the following four mechanisms, and eventually uses the operation that leads to the most optimal solution to guide the search process, i.e.,
(1) each particle follows the leader fully in every dimension as defined in the original PSO algorithm;
(2) it randomly selects some sub-dimensions and follows the leader partially in those dimensions;
(3) it avoids personal and global worst experiences in every dimension;
(4) it randomly selects some sub-dimensions and avoids personal and global worst individuals partially in those dimensions.

These four search mechanisms increase diversification of the search process and enable the proposed PSO model to explore distinctive search regions and to reduce the likelihood of being trapped in local optima. During the search process, each particle performs each of the four mechanisms for position updating. Then, the best offspring stems from the four actions is employed for velocity updating. After the subswarm-based search process iterates for a number of iterations, a new subswarm leader is retrieved in each subpopulation. The new subswarm leader is further enhanced using three probability distributions, i.e. Gaussian, Cauchy, and Levy distributions. If the offspring solution generated by any of these random walk strategies has a better fitness score than that of the current parent subswarm leader, this promising offspring is used to replace the parent subswarm leader. Finally, the two subpopulations are merged, and the best solution among the two subswarm leaders is regarded as the new global best solution. Another set of three worst solutions is also retrieved and used to update the worst archive.

However, if the second swarm leader, which possesses a promising fitness score but embeds a low correlation to the global best solution, cannot be retrieved at the beginning stage of the search process in a certain iteration, the search operations described above are conducted purely using the primary swarm, instead of two subswarms.

Subsequently, the matrix representation of the swarm is dynamically adjusted to a new form by switching the rows and columns of the original matrix (e.g. changing a particle matrix representation from 5×29 to 29×5) and the overall search process is repeated using this new matrix representation. The algorithm iterates until the termination criteria are reached.

	Algorithm 1: The Proposed PSO Algorithm
1	**Start**
2	Initialize a population (e.g. 50 particles);
3	Evaluate the population to identify the initial best leader, g_{best};
4	Archive the best leader, g_{best}, in the best-memory;
5	**While** (Stopping criterion is not satisfied)
6	{
7	**For** each particle x_i in the population **do**

8 {
9 **If** (|fitness(x_i) − fitness(g_{best})|<Threshold_fitness) && (correlation between x_i and g_{best} < 0)
10 Select the candidate, x_i, as the second leader;
11 **End If**
12 }**End For**
13 Identify three worst solutions in the population and store them in the worst-memory;
14 **If** (there are two leaders)
15 {
16 Randomly separate the population into two groups with each group led by one leader; //subswarm 1 led by g_{best} and subswarm 2 led by the second best
17 **While** (!stagnation detected)
18 { **For** each group **do**
19 { **For** each particle x_i in the group **do**
20 {
21 **If** (x_i is one of the worst solutions in the worst-memory)
22 Follow the leader fully (i.e. in every dimension);
23 **Else If** // $x_i \neq$ leader or one of the worst solutions
24 { Run the following steps and use the strategy which leads to the best solution for position updating;
25 1. Partially avoid the worst solutions by randomly selecting some sub-dimensions and moving away from the worst experiences in those sub-dimensions;
26 2. Partially follow the leader by randomly selecting some sub-dimensions and following the leader in those sub-dimensions;
27 3. Fully avoid the worst solutions in every dimension;
28 4. Fully follow the leader in every dimension;
29 } **End If**
30 Evaluate x_i at the new position and update personal best of x_i;
31 }**End For**
32 Evaluate the particles in the subswarm and update the subswarm leader;
33 Conduct long jumps of the current subswarm leader using Levy flights/Gaussian/Cauchy distributions;
34 Update the current subswarm leader if any new solution developed by the long jumps has a better fitness;
35 }**End For**
36 Compare two subswarm leaders and store the best subswarm leader;
37 } **Until** (stagnation detected, e.g. convergence to the same/similar fitness for 2-3 times);
38 Combine the sub-swarms and update the best leader, g_{best}, in the best-memory with the best subswarm leader;
39 }
40 **Else** //there is just one leader
41 { Conduct the single group optimization using the loop from lines 17-37;
42 Update the best leader, g_{best}, in the best-memory;
43 }
44 **End If**
45 Use g_{best} to replace the worst particle in the overall population;
46 }**Until** (Stagnate 2 times)
47 Change the matrix representation;
48 **While** (Overall termination criteria are not achieved)
49 {
50 Repeat lines 5-47;
51 } **Until** (the matrix has been changed 3 times and no more improvement can be found);
52 Return the most optimal solution;
53 **End**

The following fitness function is used to evaluate each particle, which consists of two criteria, i.e. classification accuracy and the number of selected features.

$$fitness(x_i) = w_a * Performance_{x_i} + w_f * (number\_of\_selected\_features)^{-1} \quad (3)$$

where w_a and w_f represent two pre-defined weights for classification performance and the number of selected features, respectively, with $w_a + w_f = 1$. In this research, since we emphasize more on the classification performance, a higher weight is assigned to w_a as compared with that of w_f. Imbalanced class instances are

observed in our experiments owing to the limitations of benign and malignant lesion cases. To deal with such imbalanced problems, the geometric mean (GM) is used as the performance indicator. This GM indicator is frequently used for the evaluation of imbalanced classification problems in machine learning [29, 39]. In this research, it is employed for fitness evaluation at the training stage, as well as calculation of the final classification results at the test stage.

Moreover, for the fitness score calculation using Equation (3), we transform the continuous value of each element in each particle to a binary value (i.e. 0 or 1) to indicate the selection (i.e. '1') or non-selection (i.e. '0') of a feature dimension. Then, the resulting binary string representing the selected feature subset is utilized for fitness evaluation. The binary string obtained from the global best solution is regarded as the most optimal feature subset, and is subsequently used for evaluation using the test set. To obtain subtle movements and avoid premature convergence, a continuous value of each element is used for each particle during the search process, and it is converted into a binary value only for fitness evaluation. We describe each key proposed mechanism in detail, as follows.

3.2.2.1 Selection of the Second Swarm Leader

After initialization of the initial swarm, each particle is ranked based on its fitness score. The global best solution is identified. To diversify the search process and avoid stagnation, a second swarm leader with a comparable fitness score but a low position proximity to that of the global best solution is retrieved. The MATLAB function, $corr2$, defined in Equation (4) is used to determine the correlation between two particles. The particle, which has a competitive fitness score but with the least correlation to the swarm leader, is selected as the second leader.

$$corr2(A,B) = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)\,(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \tag{4}$$

where $corr2(A,B)$ denotes the correlation between two particles, A and B, and $\bar{A} = mean2(A)$ and $\bar{B} = mean2(B)$ (provided by MATLAB) are used to compute the mean of the values in each matrix, respectively. m and n denote the row and column indexes of the particles. This $corr2$ function returns a value in the range of [-1, 1], with '1' indicating the two particles are exactly the same and '-1' indicating the two particles are distinctive entirely in positions.

After identifying the global best solution and the second swarm leader, we divide the overall swarm into two subswarms. The search process of each subswarm is guided by each leader. Since the two leaders are remote in positions, i.e. a low correlation in position, it is more likely that they lead the subswarm-based search process to explore distinctive regions, in an attempt to avoid stagnation.

3.2.2.2 Attraction and Evading Mechanisms

The proposed PSO model employs diverse attraction and evading search strategies in the primary or subswarm-based search to increase search diversity. Besides following the (sub)swarm leader fully in each dimension as in the original PSO algorithm for velocity updating, the proposed PSO model enables the particles to follow the leader in any randomly selected sub-dimensions to avoid stagnation. Specifically, the algorithm randomly selects some sub-dimensions, e.g. a row/column/diagonal, then each particle follows the leader in these sub-dimensions using Equations (1) and (2).

Besides the attraction action, the proposed PSO model avoids the local and global worst experiences partially and fully to diversify the search process. In each iteration, we employ the personal worst experience and the mean of three global worst solutions as the enemy signals. Similar to the attraction action, the model randomly selects a row/column/diagonal in the matrix. The evading mechanism defined in Equations (5)-(7) is subsequently used to enable the current particle to flee away from the personal and global less optimal experiences in each dimension or in randomly selected sub-dimensions.

$$v_{id}^{t+1} = w * v_{id}^{t} - c_1 * r_1 * (p_{iw} - x_{id}^{t}) - c_2 * r_2 * \left(p_{gw} - x_{id}^{t}\right) \tag{5}$$

$$p_{gw} = \frac{w_1 + w_2 + w_3}{3} \tag{6}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{7}$$

where p_{iw} indicates the personal worst experience, while p_{gw} denotes the mean solution of three global worst individuals, i.e. w_1, w_2, and w_3, in each iteration.

In the original PSO model, the search process is purely guided by the swarm leader. When the attraction driven search mechanism stagnates, this results in premature convergence of the original PSO model. On the contrary, the proposed evading action is complementary to the original attraction operation. In the proposed PSO algorithm, when the attraction mechanism guided by the subswarm leader becomes stagnant, the evading operation pushes the subswarm particles away from less optimal regions to overcome stagnation.

In the proposed PSO model, the settings of the acceleration constants, $c_1$and c_2, and the inertia weight, w for both the attraction and evading operations mainly inherit from those of the original PSO model. Nevertheless, further parameter tuning is conducted based on trial-and-error in our experiments. The detailed parameter settings of the proposed algorithm are provided in Section 4.1.

3.2.2.3 Local Exploitation Using Probability Distributions

Three continuous probability distributions, i.e. Gaussian, Cauchy and Levy, are used to further increase local exploitation of each subswarm leader. The following equation defines the exploitation mechanism.

$$gbest'_d = gbest_d + \left(X_d^{max} - X_d^{min}\right) * \delta \tag{8}$$

where δ represents either the Gaussian, Cauchy, or Levy distribution; X_d^{max} and X_d^{min} denote the upper and lower boundaries in the d^{th} dimension, respectively. We apply these three random walk distributions consecutively. If the offspring g'_{best} generated by any of the three distributions has a better fitness score than that of the current parent subswarm leader, g_{best}, g'_{best} is used to replace g_{best}. Otherwise, the subswarm leader remains intact. This search strategy increases exploitation of the search space, therefore, it is more likely to reach global optimality.

Finally, the matrix representation of the overall swarm is dynamically adjusted and evolved to increase search diversity. The search processes discussed in Sections 3.2.2.1-3.2.2.3 are conducted again under the new matrix representation. In addition, during the search process, when the algorithm cannot identify a second swarm leader that has a similar fitness score but with the least in correlation to the swarm leader, the search processes presented in Section 3.2.2.2 are conducted using a primary swarm led by a single leader, instead of two subswarm leaders. The following parameter setting recommended by empirical studies in [22, 28] and further modified by our experimental fine-tuning is utilized in this research, i.e. inertia weight=0.99, acceleration constants c_1=c_2=2.5, population=50, and maximum generation=500.

4. EVALUATION

To evaluate the proposed PSO variant, we implement several classical search methods for comparison, i.e., PSO [38], Bat Algorithm (BA) [40], Harmony Search (HS) [41], GA [42], Dragonfly Algorithm (DA) [43], Flower Pollination Algorithm (FPA) [44], Moth-Flame Optimization (MFO) [45], Artificial Bee Colony (ABC) [46], Cultural Algorithm (CA) [47], and BBPSO [48]. Several advanced PSO variants are implemented for comparison, including DNLPSO [15], ELPSO [16], AGPSO [17], ThBPSO [26], MS-PSO [27], GM-PSO [28], BBPSOV [29], and GPSO [31]. In addition, non-evolutionary algorithm-based feature selection methods such as ReliefF [49] and Minimum Redundancy and Maximum Relevance (mRMR) [50] are employed for comparison.

This research employs three dermoscopic image databases, i.e. Dermofit Image Library [51], PH2 [52], and Dermnet [53]. The Dermofit Image Library [51] has a total of 1,300 skin lesion images from ten classes including melanomas, SK and BCC. We extract 127 benign and 76 melanoma images from this database for evaluation of the proposed PSO model. The PH2 database [52] has a total of 200 images with 80 common nevi (benign), 80 atypical nevi, and 40 melanoma cases. We employ 80 and 40 images for the benign and melanoma cases, respectively, from this database for evaluation. The Dermnet data set [53] has a total of 152 images with 45 Clark Nevi (benign) and 107 malignant melanoma cases. We employ all the malignant images (i.e. 107) from this data set in our experimental studies. Therefore, a total of 207 benign (127 (Dermofit Image Library) + 80 (PH2)) and 223 (76 (Dermofit Image Library) +40 (PH2) +107 (Dermnet)) malignant melanoma images extracted from the three databases are used in the experiments. Table 2 shows the detailed data set information for training and test purposes.

As shown in Table 2, we utilize 302 (151 benign and 151 melanoma) and 128 (56 benign and 72 melanoma) images for training and test, respectively. The training set consists of 95(Dermofit Image Library)+56(PH2) benign images, and 53(Dermofit Image Library)+28(PH2)+70(Dermnet) malignant images. The test set is composed of 32(Dermofit Image Library)+24(PH2) benign images, and 23(Dermofit Image Library)+12(PH2)+37(Dermnet) malignant images. Two classifiers, i.e. the K-Nearest Neighbour (KNN) and SVM, are used for the skin lesion classification, owing to their popularity and stable performance in skin cancer and other classification tasks [6, 7].

Table 2 Training and test sets for melanoma classification

Data set	Benign	Melanoma
Dermofit Image Library	127 (95 training + 32 test)	76 (53 training + 23 test)
PH2	80 (56 training + 24 test)	40 (28 training + 12 test)
Dermnet	-	107 (70 training + 37 test)

Moreover, most of the images are presented in 8-bit RGB with various resolutions. As most images incorporate 720 pixels to the greatest extent, additional images that do not fulfil this requirement are edited to fit the required size. For each image, the lesion is presented in the centre of the image with non-lesion skin observable at the corners of the image. Each algorithm is executed 30 runs in each experiment. The average GM scores over 30 runs are computed for comparison.

4.1 Parameter Settings

Table 3 shows the parameter settings of the proposed PSO model and other search methods used in the experimental studies. The parameter setting of the proposed PSO model mainly inherits that of the original PSO model. Further parameter fine-tuning is conducted based on trial-and-error. As indicated in Table 3, an inertia weight of 0.99 is employed by the proposed PSO model to signify the impact of the previous velocity over iterations. Similar to those in the original PSO algorithm, large acceleration constants ($c_1 = c_2 = 2.5$) are applied to accelerate convergence. The parameter settings of other search methods are used in accordance with recommendation in their original studies.

Table 3 Parameter settings of each algorithm

Algorithms	Parameters
PSO [38]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1=c_2=2
BA [40]	loudness=0.5, pulse rate=0.5
HS [41]	bandwidth=0.2, harmony memory accepting rate=0.95, pitch-adjusting rate=0.3
GA [42]	crossover probability = 0.7, mutation probability = 0.3
DA [43]	separation factor=0.1, alignment factor=0.1, cohesion factor=0.7, food factor=1, enemy factor=1, and inertia weight=0.9 – m × ((0.9-0.4)/maxi_iterations), where m and maxi_iterations represent the current and maximum iteration numbers, respectively.
FPA [44]	switch/proximity probability=0.8
MFO [45]	Use adaptive parameter settings
ABC [46]	limit=0.6×dimension×population
CA [47]	probability of the knowledge source=0.35, number of accepted individuals=probability of the knowledge source × population
BBPSO [48]	No parameter setting required
DNLPSO [15]	c_1=c_2=1.49445, refreshing gap=3, regrouping period=5, inertia weight=0.9 – (0.9 – 0.4) × (k – 1)/(max_gen – 1), where k and max_gen represent the current and and maximum iteration numbers, respectively.
ELPSO [16]	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)/(max_gen – 1), where k and max_gen represent the current and and maximum iteration numbers, respectively.
AGPSO [17]	maximum velocity=0.6, inertia weight=0.9, adaptive decreasing $c_1$and increasing c_2 over generations
ThBPSO [26]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1=c_2=2
MS-PSO [27]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1=c_2=2
GM-PSO [28]	maximum velocity = 0.6, inertia weight = 0.5, acceleration constants c_1=c_2=1.5, standard deviation of Gaussian distribution=1, scaling factor of Cauchy distribution = 2, crossover probability = 0.6, mutation probability = 0.05
BBPSOV [29]	Logistic map used as the search parameter
GPSO [31]	maximum velocity=0.6, inertia weight=0.9, acceleration constants c_1=2.6, c_2=1.5 crossover probability = 0.7, mutation probability = 0.3
Proposed PSO	maximum velocity=0.6, inertia weight=0.99, acceleration constants c_1=c_2=2.5

4.2 Evaluation Using the Combined Skin Lesion Data Set

First of all, we employ the combined skin lesion data set extracted from the above three databases to evaluate the proposed PSO algorithm. A training set of 302 images and a test set of 128 images are used. The proposed PSO model is compared with 10 classical search methods and 8 advanced PSO variants, as well as two non-evolutionary algorithm-based feature selection methods, i.e., ReliefF and mRMR.

Many popular well-known optimization packages (such as Black-Box Optimization Benchmarking [54]) have employed the number of function evaluations as the performance criterion. Therefore the number of function evaluations is also used as the criterion in this research for performance comparison. Specifically, to enable a fair comparison, we employ the following maximum number of function evaluations, i.e. population size (50) × maximum number of iterations (500), in all search methods. In other words, although there are multiple search strategies embedded in the proposed PSO model as well as other PSO variants, they all employ approximately the same number of function evaluations during the training stage, i.e. population size (50) × maximum number of iterations (500), as that of PSO and other classical methods. As an example, a comparatively smaller number of iterations (e.g. 100) is used in the proposed PSO model in comparison with that (e.g. 500) used by other classical methods during the feature selection process owing to the fact that the proposed model utilizes an additional number of function evaluations for both velocity updating and the enhancement of the subswarm leaders in each iteration. In this way, the proposed model uses approximately the same number of function evaluations, i.e. population size (50) × maximum number of iterations (500), as those of all classical search methods and advanced PSO variants including DNLPSO, BBPSOV, AGPSO, MS-PSO, and GPSO. Similarly, although ELPSO, GM-PSO and ThBPSO conduct additional numbers of function evaluations for swarm leader enhancement, the setting of population size (50) × maximum number of iterations (500) is also used as the maximum number of function evaluations in these PSO variants for discriminative feature selection. Such a test strategy is applied to all subsequent experiments to ensure a fair comparison.

Table 4 shows the average classification performances over 30 runs for each method in combination with KNN and SVM classifiers for both 10-fold and hold-out validations. The minimum, maximum, standard deviation, and the mean GM scores over 30 runs are presented. The best result in each row is highlighted in bold. As indicated in Table 4, the best results are obtained when the SVM classifier is applied to each method. Integrated with SVM, the proposed PSO model achieves the highest average GM scores of 94.23% and 94.51% for the two-class skin lesion classification using 10-fold and hold-out validations, respectively. It outperforms 20 other algorithms consistently.

Table 4 Average classification results of each algorithm over 30 runs for the combined skin lesion data set

			Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
KNN	10 Fold	Mean	**0.9231**	0.9009	0.9035	0.8992	0.8967	0.8861	0.8998	0.8968	0.8958	0.8980	0.9037
		Std	**0.0146**	0.0217	0.0226	0.0209	0.0212	0.0321	0.0162	0.0287	0.0221	0.0225	0.0238
		Min	**0.8986**	0.8525	0.8495	0.8508	0.8528	0.8266	0.8620	0.8385	0.8528	0.8598	0.8429
		Max	0.9434	0.9354	**0.9532**	0.9287	0.9401	0.9380	0.9348	0.9462	0.9368	0.9415	0.9521
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9262**	0.9075	0.9099	0.9072	0.9038	0.8932	0.9076	0.9042	0.9021	0.9035	0.9085
		Std	**0.0141**	0.0187	0.0191	0.0193	0.0191	0.0283	0.0144	0.0266	0.0199	0.0207	0.0223
		Min	**0.8979**	0.8681	0.8716	0.8594	0.8681	0.8389	0.8805	0.8445	0.8627	0.8654	0.8468
		Max	0.9465	0.9373	**0.9555**	0.9346	0.9435	0.9346	0.9346	0.9465	0.9346	0.9435	**0.9555**
		Rank Sum		+	+	+	+	+	+	+	+	+	+
SVM	10 Fold	Mean	**0.9423**	0.9232	0.9208	0.9296	0.9270	0.8818	0.9161	0.9047	0.9235	0.9208	0.9268
		Std	**0.0162**	0.0206	0.0231	0.0198	0.0211	0.0337	0.0295	0.0286	0.0273	0.0169	0.0253
		Min	**0.9158**	0.8641	0.8614	0.8803	0.8894	0.8098	0.8536	0.8422	0.8539	0.8964	0.8589
		Max	0.9702	0.9626	0.9598	0.9577	0.9683	0.9342	0.9586	0.9563	0.9655	0.9532	**0.9789**
		Rank Sum		+	+	**0**	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9451**	0.9284	0.9255	0.9348	0.9314	0.8875	0.9204	0.9102	0.9281	0.9244	0.9307
		Std	**0.0146**	0.0192	0.0230	0.0166	0.0181	0.0335	0.0281	0.0261	0.0243	0.0169	0.0227
		Min	**0.9225**	0.8716	0.8716	0.8955	0.8979	0.8027	0.8536	0.8506	0.8746	0.8979	0.8746
		Max	0.9698	0.9674	0.9582	0.9582	0.9698	0.9435	0.9605	0.9555	0.9765	0.9582	**0.9791**
		Rank Sum		+	+	+	+	+	+	+	+	+	+

			Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO	ReliefF	mRMR
KNN	10 Fold	Mean	**0.9231**	0.8985	0.9018	0.9022	0.9009	0.9021	0.8959	0.9054	0.8846	0.9059	0.8791
		Std	**0.0146**	0.0251	0.0162	0.0154	0.0281	0.0180	0.0190	0.0152	0.0273	-	-
		Min	**0.8986**	0.8431	0.8624	0.8720	0.8394	0.8704	0.8475	0.8772	0.8216	-	-
		Max	0.9434	0.9296	0.9302	0.9334	**0.9448**	0.9314	0.9286	0.9415	0.9293	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9262**	0.9057	0.9078	0.9093	0.9068	0.9075	0.9017	0.9111	0.8917	0.9103	0.8853
		Std	0.0141	0.0219	0.0159	0.0141	0.0254	0.0164	0.0166	**0.0134**	0.0241	-	-
		Min	**0.8979**	0.8561	0.8716	0.8768	0.8561	0.8836	0.8594	0.8853	0.8327	-	-
		Max	**0.9465**	0.9373	0.9435	0.9346	**0.9465**	0.9346	0.9313	0.9435	0.9302	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+

SVM	10 Fold	**Mean**	**0.9423**	0.9208	0.9178	0.9114	0.9216	0.9210	0.9139	0.9169	0.8805	0.9395	0.8746
		Std	**0.0162**	0.0270	0.0233	0.0295	0.0232	0.0286	0.0303	0.0250	0.0395	-	-
		Min	**0.9158**	0.8597	0.8659	0.8446	0.8742	0.8671	0.8232	0.8723	0.8023	-	-
		Max	0.9702	0.9551	0.9493	0.9568	0.9598	**0.9711**	0.9655	0.9626	0.9521	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	**Mean**	**0.9451**	0.9253	0.9224	0.9163	0.9250	0.9246	0.9190	0.9224	0.8881	0.9435	0.8814
		Std	**0.0146**	0.0250	0.0229	0.0279	0.0219	0.0254	0.0264	0.0225	0.0363	-	-
		Min	**0.9225**	0.8654	0.8716	0.8561	0.8790	0.8805	0.8561	0.8885	0.8259	-	-
		Max	**0.9698**	0.9582	0.9555	0.9582	0.9645	0.9674	0.9674	0.9674	0.9582	-	-
		Rank Sum		+	+	+	+	+	+	+	+	**0**	+

Table 5 The training computational cost and the average number of features selected by each optimization algorithm

	Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
Average no. of selected features	49.33	66.67	63.63	69.70	61.10	53.43	60.77	61.20	68.23	64.37	59.83
Training cost (seconds)	4554.03	4420.33	4684.56	4429.17	4397.66	4278.71	4565.42	4766.10	4607.55	4612.62	4617.27

	Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO
Average no. of selected features	49.33	65.83	67.67	63.50	61.10	61.63	56.73	64.43	38.53
Training cost (seconds)	4554.03	4552.60	4979.95	4674.78	4670.05	4657.17	4590.90	4750.04	4741.40

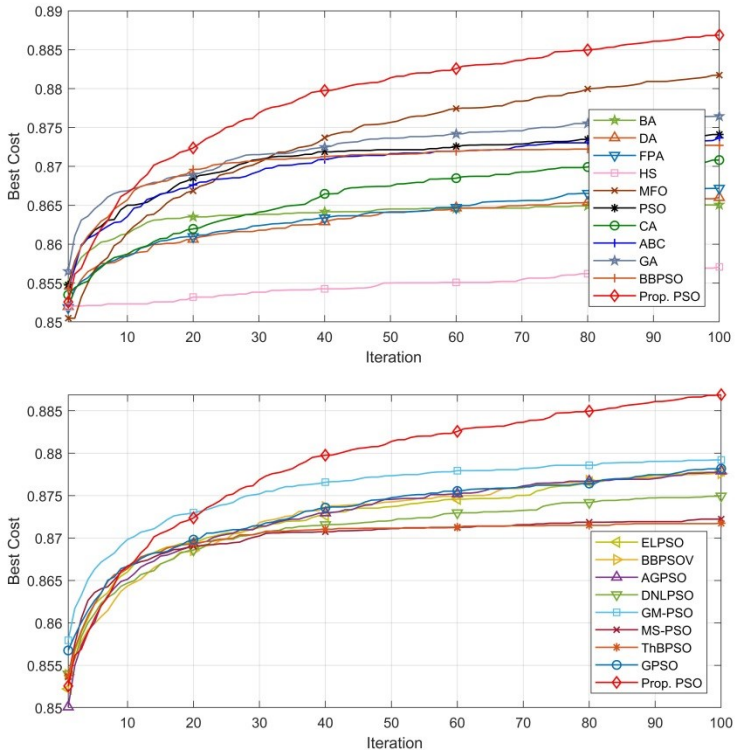

Figure 2 Average convergence curve for each method over 30 runs in combination with the SVM classifier

As indicated in Table 5, ThBPSO selects the smallest feature subset, i.e. 38.53 features on average over 30 runs. This is followed by our proposed model, with an average feature size of 49.33, while all other methods retrieve comparatively larger feature subsets. Although ThBPSO identifies the smallest feature subset based on the frequency of each feature recommended by multiple global best solutions, owing to the removal of some distinctive characteristics, ThBPSO obtains lower mean GM scores in comparison with those of the proposed model. In addition, we employ the corresponding top-ranked 60 salient features identified by ReliefF and mRMR respectively for performance comparison owing to their superior classification performances.

In this experiment, the MATLAB parallel computing toolbox is used to increase the computational efficiency and reduce the time consumption of each feature selection task. A multicore workstation with 36 core Intel Xeon processors and 256GB RAM is used. We spread 30 experimental runs and the 10-fold cross validation into multiple treats. The feature selection training cost of each method is shown in Table 5. As illustrated in Table 5, the training cost of the proposed PSO model is comparable with (or slightly lower than) those of most

classical methods and some advanced PSO variants under approximately the same number of function evaluations.

To further indicate superiority of the proposed model, the statistical Wilcoxon rank sum test [55] is conducted. It is a non-parametric test that determines whether two distributions (i.e. solutions) have a statistically equal median. The rank sum test returns a *p*-value which indicates the rejection of the null hypothesis of an equal median, or otherwise, at the default 5% significance level. We further transform the *p*-value results into the statistical outcomes shown in Table 4. Specifically, the last row of each set of the results in Table 4 shows the statistical comparison results between the proposed model and other methods, where symbols such as '+' and '-' are used to indicate the proposed model is statistically significantly better or worse than the compared methods. Symbol '0' is also used to indicate that the proposed model and other methods have the same result distributions. Results indicate that the proposed model is statistically significantly better than other methods in nearly all test cases. The exception is for DA and ReliefF, which show similar result distributions as those of the proposed PSO model when integrated with the SVM classifier under 10-fold and hold-out validations, respectively.

Figure 2 illustrates the average convergence curve over 30 runs for each method in combination with the SVM classifier. The proposed PSO model outperforms other methods, and depicts the fastest convergence rate, which is followed by those of MFO and GM-PSO.

To evaluate the efficiency of each search strategy in the proposed PSO model, the following four additional experiments are conducted by incrementally adding different employed mechanisms. The four tested components are as follows, i.e. (1) the subswarm-based search led by two remote leaders + the attraction operation by following the subswarm leader partially and fully, (2) the setting of (1) + the evading action by avoiding personal and global worst experiences partially and fully, (3) the setting of (2) + long jumps of each subswarm leader using Levy flights/Gaussian/Cauchy distributions, and (4) the setting of (3) + changing the matrix representation, i.e. the full version. Table 6 shows the average GM scores over 30 runs for each test component and the original PSO model under both 10-fold and hold-out validations.

Table 6 Evaluation of different proposed mechanisms using the combined skin lesion data set

			PSO	(1) Subswarms + attraction	(2) Subswarms + attraction + evading	(3) Subswarms + attraction + evading + leader enhancement	(4) Full version
KNN	10 Fold	Mean	0.8958	0.9202	0.9215	0.9221	**0.9231**
		Std	0.0221	**0.0128**	0.0144	0.0159	0.0146
		Min	0.8528	0.8981	0.8983	**0.8990**	0.8986
		Max	0.9368	0.9421	0.9428	0.9431	**0.9434**
	Hold Out	Mean	0.9021	0.9217	0.9257	0.9250	**0.9262**
		Std	0.0199	**0.0118**	0.0141	0.0142	0.0141
		Min	0.8627	0.8946	0.8965	0.8946	**0.8979**
		Max	0.9346	0.9435	0.9462	0.9455	**0.9465**
SVM	10 Fold	Mean	0.9235	0.9350	0.9402	0.9411	**0.9423**
		Std	0.0273	0.0175	0.0167	**0.0154**	0.0162
		Min	0.8539	0.9091	0.9114	**0.9159**	0.9158
		Max	0.9655	0.9604	0.9683	**0.9711**	0.9702
	Hold Out	Mean	0.9281	0.9386	0.9405	0.9422	**0.9451**
		Std	0.0243	0.0160	0.0160	**0.0137**	0.0146
		Min	0.8746	0.9164	0.9209	**0.9255**	0.9225
		Max	**0.9765**	0.9613	0.9689	0.9674	0.9698

As illustrated in Table 6, the experiments using different proposed mechanisms in configurations 1-4 show performance improvements in comparison with that of the original PSO model. Each proposed mechanism is able to improve the performance incrementally, while the best mean GM results are achieved by the full version of the proposed algorithm in configuration 4. In short, the above experiments indicate the competence of each proposed strategy in improving the performance of the proposed PSO model.

4.3 Evaluation Using the PH2 Skin Lesion Data Set

For comparison with other related research studies on skin cancer detection, we conduct another evaluation dedicated to the PH2 data set. The PH2 database has a total of 200 images with 80 common nevi (benign), 80 atypical nevi, and 40 melanoma cases. We conduct 3-class skin lesion classification in this experiment. For the hold-out validation, we use 70% and 30% of the images from each class for training and test, respectively. A total of 30 trials are performed for each feature selection algorithm. The KNN and SVM models are used as the underlying classifiers for evaluating the PH2 data set, owing to their popularity in related research studies [6, 56]. The maximum number of function evaluations, i.e. population size (50) × maximum number of iterations

(500), is applied to all the methods in this experiment. Table 7 shows the experimental and the statistical test results. The size of selected feature subsets for each method is presented in Table 8.

As illustrated in Table 7, the empirical and statistical test results indicate the statistical superiority of the proposed model over other methods for the 3-class lesion classification for all test cases. Referring to Table 8, ThBPSO identifies a smaller average subset of features, i.e. 48.53, than that of our proposed model (i.e. an average subset of 50.23 features). However, the proposed model shows a better discriminative power with better mean GM scores. The computational cost of the proposed model is comparable with those of all other classical methods and advanced PSO variants. In this experiment, two sets of 60 top-ranked discriminative features identified by ReliefF and mRMR, respectively, are employed for comparison, owing to their impressive classification performances.

Table 7 Average classification results of each algorithm over 30 runs using the PH2 data set

			Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
KNN	10 Fold	Mean	**0.9416**	0.8945	0.8863	0.8845	0.8870	0.9200	0.9100	0.9342	0.8843	0.9095	0.9050
		Std	**0.0551**	0.0619	0.0778	0.0712	0.0856	0.0684	0.0623	0.0598	0.0756	0.0574	0.0563
		Min	**0.8000**	0.7316	0.6808	0.7524	0.6633	0.7524	0.7414	**0.8000**	0.7101	0.7816	0.7500
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9604**	0.9187	0.9161	0.9209	0.9247	0.9364	0.9277	0.9524	0.9217	0.9290	0.9256
		Std	**0.0267**	0.0378	0.0422	0.0418	0.0411	0.0426	0.0418	0.0305	0.0381	0.0336	0.0417
		Min	**0.9057**	0.8432	0.7906	0.8432	0.8539	0.8615	0.8432	0.8839	0.8323	0.8750	0.8101
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
		Rank Sum		+	+	+	+	+	+	+	+	+	+
SVM	10 Fold	Mean	**0.9523**	0.9111	0.9094	0.9166	0.9432	0.9221	0.9258	0.9316	0.9093	0.9238	0.9340
		Std	0.0534	0.0699	0.0630	0.0725	**0.0462**	0.0624	0.0707	0.0709	0.0709	0.0646	0.0629
		Min	0.8121	0.7121	0.7707	0.7121	**0.8214**	0.8000	0.7000	0.7000	0.7707	0.7414	0.7414
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9645**	0.9456	0.9374	0.9417	0.9614	0.9485	0.9487	0.9562	0.9468	0.9574	0.9583
		Std	0.0316	0.0369	0.0358	0.0456	**0.0263**	0.0331	0.0376	0.0329	0.0386	0.0304	0.0358
		Min	0.8660	0.8478	0.8660	0.8292	**0.8714**	0.8660	0.8660	0.8660	0.8660	0.8660	0.8660
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
		Rank Sum		+	+	+	+	+	+	+	+	+	+

			Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO	ReliefF	mRMR
KNN	10 Fold	Mean	**0.9416**	0.9057	0.9178	0.9255	0.9089	0.8991	0.8987	0.9123	0.8801	0.8998	0.8811
		Std	0.0551	0.0693	0.0738	0.0587	0.0796	0.0744	0.0687	0.0639	**0.0303**	-	-
		Min	0.8000	0.7157	0.7157	0.7406	0.6678	0.7524	0.7571	0.7633	**0.8103**	-	-
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9604**	0.9327	0.9400	0.9425	0.9361	0.9272	0.9356	0.9373	0.8869	0.9287	0.8920
		Std	**0.0267**	0.0408	0.0423	0.0370	0.0483	0.0399	0.0424	0.0383	0.0270	-	-
		Min	**0.9057**	0.8323	0.8539	0.8229	0.7873	0.8432	0.8229	0.8539	0.8353	-	-
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
SVM	10 Fold	Mean	**0.9523**	0.9109	0.9266	0.9244	0.9219	0.9105	0.9196	0.9001	0.9004	0.9292	0.9001
		Std	0.0534	0.0833	0.0614	0.0702	0.0678	0.0627	0.0599	0.0804	**0.0271**	-	-
		Min	0.8121	0.7414	0.7707	0.7524	0.7121	0.7707	0.7707	0.7121	**0.8126**	-	-
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold Out	Mean	**0.9645**	0.9378	0.9558	0.9519	0.9428	0.9414	0.9434	0.9387	0.9054	0.9481	0.9210
		Std	0.0316	0.0428	0.0331	0.0322	0.0378	0.0365	0.0404	0.0468	**0.0264**	-	-
		Min	0.8660	0.8292	0.8660	**0.8824**	0.8292	0.8660	0.8660	0.8292	0.8561	-	-
		Max	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+

Table 9 shows the comparison between our work and other related research reported in the literature for skin cancer detection using the PH2 data set. Since each work utilizes either different classification methods or different training and test data sets, Table 9 is an approximate indication of the performance comparison among different methods. According to Table 9, the proposed model is one of the top performers for the PH2 data set. The proposed attraction and evading strategies embedded in the subswarm-based search and dynamic matrix representations account for the superiority of the proposed PSO model.

Table 8 The training computational cost and the average number of features selected by each optimization algorithm

	Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
Average no. of selected features	50.23	68.60	64.67	69.60	64.83	62.20	65.33	57.70	66.63	66.63	61.37
Training cost (seconds)	3897.92	3808.29	3931.76	3806.97	3798.13	3782.29	3927.84	3900.98	3876.96	3897.18	3922.39

	Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO
Average no. of selected features	50.23	64.60	67.80	61.37	64.67	65.50	58.90	63.00	48.53
Training cost (seconds)	3897.92	3855.17	3962.14	3875.86	3877.59	3874.13	3887.29	3940.21	4039.71

Table 9 Performance comparison between the proposed system and related research using the PH2 data set

Studies	Methodology	Classes	Evaluation strategy	Recognition rate (%)
Abuzaghleh et al. [56]	2-D Fast Fourier Transform + 2-D Discrete Cosine Transform + KNN	3	25% for testing	63.33
Abuzaghleh et al. [3]	2-D Fast Fourier Transform + 2-D Discrete Cosine Transform + Complexity + Colour + Pigment Network + Lesion Shape, Orientation, Margin, Intensity Pattern + A Two-Level Classifier	3	25% for testing	96.50
Barata et al. [57]	Lesion or Pigment Ratio Features + Boosting Algorithm	3	10-Fold	86.20
Barata et al. [58]	Colour Scale Invariant Feature Transform + Bag-of-Features	3	10-Fold	87.00
Barata et al. [59]	Colour Constancy + Bag-of-Features Framework + K-Means + SVM	3	10-Fold	84.30
Vasconcelos et al. [60]	Colour Features + SVM	3	10-Fold	81.38
Soumya et al. [61]	Colour Correlogram + Segmentation based Fractal Texture Analysis + Bayes Classifier	3	10-Fold	91.50
Pennisi et al. [62]	Artefact Removal, Skin Detection, Lesion Segmentation & Binary Mask + AdaBoost	3	N/A	93.60
This research	Shape, Colour & GLRLM Features + Prop. PSO-based Feature Selection + SVM	3	10-Fold & Hold-Out (30% for testing)	**95.23 (10-Fold)** **96.45 (Hold-Out)**

4.4 Evaluation Using the UCI Spambase Data Set

The proposed PSO model is also applied to other large-scale problems for feature optimization. To evaluate the discriminative capabilities of the proposed model, the UCI spambase data set [63] is used for evaluation. The spambase data set has 4601 instances, 57 attributes, and 2 classes (spam or non-spam). All the instances are used in the experiment with 80%:20% as the training and test split. A total of 30 runs are performed for each method. The following experimental settings are used, i.e. dimension=57, population=20, and iteration=200. The maximum number of function evaluations, i.e. population size (20) × maximum number of iterations (200), is applied to all methods.

Table 10 shows the detailed average GM scores over 30 runs for each method and the Wilcoxon rank sum test results, respectively. The empirical and statistical results in Table 10 indicate the statistical superiority of the proposed PSO model over all other methods, except for GM-PSO, which shows similar result distributions to those of the proposed PSO model when integrated with the KNN classifier. Moreover, two sets of 50 top-ranked salient features identified by ReliefF and mRMR, respectively, are employed for performance comparison owing to their competitive classification performances.

As shown in Table 11, the proposed model selects a comparatively smaller set of features (i.e. 33.73) with comparable computational efficiency, than those of majority of PSO variants at the training stage. Although MS-PSO and ThBPSO retrieve the smallest feature subsets, i.e. 31.93 and 32.98 features, respectively, our proposed model yields a better trade-off between computational efficiency and discriminative capability. The empirical results indicate the superiority of the proposed PSO model over other methods for solving discriminative feature selection for this spambase data set.

Table 10 Average classification results of each algorithm over 30 runs for the UCI spambase data set

			Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
KNN	**10-fold**	**Mean**	**0.8982**	0.8645	0.8799	0.8727	0.8724	0.8916	0.8801	0.8809	0.8757	0.8807	0.8762
		Std	0.0117	0.0120	0.0079	0.0116	0.0105	**0.0067**	0.0071	**0.0067**	0.0113	0.0078	0.0136
		Min	0.8580	0.8477	0.8575	0.8483	0.8496	**0.8819**	0.8649	0.8674	0.8540	0.8642	0.8397
		Max	**0.9138**	0.8898	0.8912	0.8901	0.9027	0.9039	0.8909	0.8925	0.8936	0.8928	0.8946
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold-out	**Mean**	**0.8987**	0.8652	0.8806	0.8734	0.8731	0.8922	0.8808	0.8817	0.8765	0.8816	0.8769
		Std	0.0117	0.0119	0.0080	0.0116	0.0104	0.0068	0.0070	**0.0067**	0.0113	0.0075	0.0135
		Min	0.8585	0.8489	0.8579	0.8496	0.8506	**0.8821**	0.8656	0.8678	0.8550	0.8652	0.8409
		Max	**0.9144**	0.8900	0.8921	0.8911	0.9031	0.9050	0.8913	0.8928	0.8940	0.8931	0.8951
		Rank Sum		+	+	+	+	+	+	+	+	+	+

			Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO	ReliefF	mRMR
SVM	10-fold	Mean	**0.9031**	0.8810	0.8930	0.8832	0.8883	0.8986	0.8966	0.9009	0.8911	0.9002	0.8887
		Std	0.0044	0.0158	0.0074	0.0118	0.0122	0.0075	0.0052	**0.0030**	0.0088	0.0045	0.0104
		Min	**0.8959**	0.8296	0.8773	0.8559	0.8537	0.8788	0.8832	0.8955	0.8697	0.8907	0.8631
		Max	0.9098	0.9085	0.9061	0.9041	0.9130	**0.9160**	0.9082	0.9110	0.9058	0.9103	0.9022
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold-out	Mean	**0.9036**	0.8817	0.8936	0.8838	0.8889	0.8991	0.8971	0.9014	0.8919	0.9008	0.8893
		Std	0.0044	0.0157	0.0072	0.0117	0.0120	0.0075	0.0052	**0.0030**	0.0088	0.0045	0.0104
		Min	**0.8962**	0.8308	0.8785	0.8569	0.8548	0.8793	0.8836	**0.8962**	0.8710	0.8911	0.8640
		Max	0.9102	0.9090	0.9067	0.9046	0.9138	**0.9162**	0.9083	0.9116	0.9062	0.9108	0.9025
		Rank Sum		+	+	+	+	+	+	+	+	+	+
KNN	10-fold	Mean	**0.8982**	0.8711	0.8726	0.8817	0.8758	0.8820	0.8669	0.8767	0.8714	0.8507	0.8725
		Std	0.0117	0.0084	0.0091	**0.0075**	0.0100	0.0077	0.0122	0.0088	0.0119	-	-
		Min	0.8580	0.8584	0.8566	**0.8687**	0.8473	0.8651	0.8393	0.8549	0.8506	-	-
		Max	**0.9138**	0.8994	0.8925	0.9006	0.8954	0.8946	0.8907	0.8966	0.8951	-	-
		Rank Sum		+	+	+	+	**0**	+	+	+	+	+
	Hold-out	Mean	**0.8987**	0.8718	0.8733	0.8824	0.8764	0.8827	0.8678	0.8774	0.8721	0.8515	0.8732
		Std	0.0117	0.0084	0.0090	**0.0075**	0.0099	0.0076	0.0121	0.0088	0.0118	-	-
		Min	0.8585	0.8586	0.8579	**0.8690**	0.8480	0.8661	0.8406	0.8556	0.8512	-	-
		Max	**0.9144**	0.9001	0.8930	0.9012	0.8959	0.8950	0.8913	0.8967	0.8963	-	-
		Rank Sum		+	+	+	+	**0**	+	+	+	+	+
SVM	10-fold	Mean	**0.9031**	0.8850	0.8939	0.8976	0.8948	0.8989	0.8792	0.8919	0.8679	0.8864	0.8960
		Std	**0.0044**	0.0100	0.0071	0.0061	0.0068	0.0054	0.0095	0.0068	0.0159	-	-
		Min	**0.8959**	0.8539	0.8760	0.8834	0.8773	0.8892	0.8588	0.8742	0.8211	-	-
		Max	0.9098	0.9047	0.9060	**0.9118**	0.9069	0.9086	0.8943	0.9078	0.8918	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+
	Hold-out	Mean	**0.9036**	0.8856	0.8945	0.8982	0.8954	0.8995	0.8799	0.8925	0.8686	0.8871	0.8966
		Std	**0.0044**	0.0100	0.0071	0.0061	0.0067	0.0053	0.0093	0.0067	0.0158	-	-
		Min	**0.8962**	0.8542	0.8765	0.8842	0.8779	0.8905	0.8602	0.8753	0.8221	-	-
		Max	0.9102	0.9055	0.9066	**0.9123**	0.9075	0.9089	0.8945	0.9085	0.8924	-	-
		Rank Sum		+	+	+	+	+	+	+	+	+	+

Table 11 The training computational cost and the average number of features selected by each optimization algorithm

	Prop. PSO	BA	BBPSO	DA	FPA	GA	HS	MFO	PSO	ABC	CA
Average no. of selected features	33.73	36.43	43.90	36.10	38.80	38.70	45.13	51.43	41.60	50.97	39.40
Training cost (seconds)	6916.25	6891.09	7005.57	6811.01	6858.62	6793.35	6963.88	6995.37	6946.64	6947.75	6934.96

	Prop. PSO	DNLPSO	ELPSO	BBPSOV	AGPSO	GM-PSO	MS-PSO	GPSO	ThBPSO
Average no. of selected features	33.73	43.73	47.57	47.33	45.20	48.20	31.93	43.70	32.98
Training cost (seconds)	6916.25	6947.66	7037.60	6899.99	6891.06	6957.73	6983.78	7026.52	7136.04

4.5 Evaluation Using Benchmark Functions

We evaluate the proposed PSO model using different optimization tasks, i.e. unimodal and multimodal benchmark functions, to further ascertain its efficiency. A set of eight standard benchmark functions is selected, owing to their multi-modalities and varied difficulties. They are summarized in Table 12. These benchmark functions have been widely used for evaluating swarm intelligence algorithms, e.g. in [33, 43, 45, and 64].

Table 12 Unimodal and multimodal benchmark functions

	Function	Definition	Range	Global minima
F1	Dixon-Price	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{d} i\left(2x_i^2 - x_{i-1}\right)^2$	[-10, 10]	0
F2	Sphere	$f(x) = \sum_{i=1}^{d} x_i^2$	[-5.12, 5.12]	0
F3	Rotated Hyper-Ellipsoid	$f(x) = \sum_{i=1}^{d} \sum_{j=1}^{i} x_j^2$	[-65.536, 65.536]	0
F4	Sum Squares	$f(x) = \sum_{i=1}^{d} i x_i^2$	[-5.12, 5.12]	0

F5	Sum of Different Powers	$f(x) = \sum_{i=1}^{d} \lvert x_i \rvert^{i+1}$	[-1, 1]	0
F6	Ackley	$f(x) = 20 + e - 20\, exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)\right)$	[-32, 32]	0
F7	Griewank	$f(x) = \sum_{i=1}^{d}\frac{x_i^2}{4000} - \prod_{i=1}^{d}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	[-600, 600]	0
F8	Powell	$f(x) = \sum_{i=1}^{d/4}[(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4]$	[-4, 5]	0

All benchmark functions can be grouped into two categories with respect to the number of local minima, i.e. unimodal and multimodal functions. F1-F5 and F6-F8 in Table 12 represent unimodal and multimodal test functions, respectively. The unimodal test functions have single global optimum (i.e. 0), whereas the multimodal test functions are characterised by multiple local minima. As indicated in [64], these unimodal and multimodal benchmark functions are challenging testbeds for evaluation of local exploitation and global exploration capabilities of search algorithms.

4.5.1 Comparison with Other Search Methods

We compare the proposed PSO model with other methods using the above mathematical landscapes. Except for ThBPSO which is dedicated to feature selection problems, all other PSO variants have been implemented for comparison. The following experimental settings are used, i.e. dimension=50, population size=50, and the maximum number of iterations=500. We apply the maximum number of function evaluations, i.e. population size (50) × the maximum number of iterations (500), in all search methods. Table 13 presents the evaluation and statistical test results for all benchmark functions. We employ the mean of the global minima over 30 runs for each algorithm as the main criterion for comparison, along with the minimum, maximum, and standard deviation results.

As shown in Table 13, the proposed PSO model achieves better mean global minima over 30 runs than those of 8 PSO variants and 10 classical methods for nearly all test functions, except for F6 (Ackley) where ELPSO obtains the best performance. The results indicate that the subswarm-based search with diverse attraction and flee mechanisms of the proposed PSO model leads to a highly efficient exploration capability of the search space, therefore reducing the probabilities of being trapped in local optima.

The Wilcoxon rank sum test results are provided to indicate the significance of the proposed model. As shown in Table 13, the statistical test results indicate that our proposed model achieves statistically significant improvements over other methods, except for F6 where the proposed model, DNLPSO, GM-PSO and AGPSO have similar medians. ELPSO also outperforms the proposed model statistically for F6. In short, the experiments using different mathematical landscapes further ascertain efficiency of the proposed model.

Table 13 Average evaluation results for all benchmark functions over 30 runs with *d*=50

		Prop. PSO	**BA**	**PSO**	**DA**	**HS**	**FPA**	**BBPSO**	**MFO**	**ABC**	**CA**	**GA**
F1 Dixon	**Mean**	**4.77E+00**	6.71E+04	2.30E+06	6.48E+03	7.91E+01	3.22E+04	2.63E+05	2.84E+05	3.19E+06	4.96E+03	4.12E+05
	Std	**3.16E+00**	5.39E+04	7.71E+05	6.75E+03	1.87E+01	1.27E+04	3.53E+05	3.90E+05	8.10E+05	4.51E+03	2.49E+05
	Min	**6.70E-01**	2.89E+03	1.28E+06	9.39E+01	4.54E+01	1.24E+04	7.86E+01	8.59E+02	1.67E+06	5.91E+02	1.02E+05
	Max	**9.82E+00**	2.29E+05	5.13E+06	2.42E+04	1.19E+02	6.55E+04	1.41E+06	1.57E+06	4.97E+06	1.86E+04	1.01E+06
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F2 Sphere	**Mean**	**5.21E-08**	1.43E+01	1.47E+02	9.91E+00	3.91E-01	1.83E+01	2.35E+01	1.88E+01	1.54E+02	1.01E+00	4.14E+01
	Std	**1.03E-07**	9.78E+00	2.59E+01	4.63E+00	4.28E-02	3.60E+00	2.57E+01	1.97E+01	2.35E+01	1.25E+00	1.29E+01
	Min	**9.30E-10**	3.54E-01	8.46E+01	3.73E+00	3.09E-01	1.19E+01	1.29E-02	4.86E-01	1.06E+02	7.09E-02	2.31E+01
	Max	**3.22E-07**	3.33E+01	1.80E+02	2.29E+01	4.61E-01	2.84E+01	1.05E+02	7.92E+01	1.94E+02	6.42E+00	8.40E+01
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F3 Rothyp	**Mean**	**1.88E-04**	2.75E+05	6.03E+05	2.16E+04	1.68E+03	6.83E+04	1.21E+05	1.11E+05	5.05E+05	4.85E+03	2.40E+05
	Std	**3.94E-04**	8.74E+04	1.29E+05	1.39E+04	4.51E+02	1.49E+04	1.13E+05	9.56E+04	7.24E+04	4.89E+03	4.88E+04

	Min	**3.84E-06**	1.08E+05	3.96E+05	3.43E+03	7.65E+02	4.07E+04	4.46E+03	9.17E+03	3.88E+05	5.01E+02	1.53E+05
	Max	**1.74E-03**	5.05E+05	8.79E+05	5.79E+04	2.73E+03	1.01E+05	3.48E+05	3.83E+05	6.54E+05	2.33E+04	3.87E+05
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F4 Sum2	**Mean**	**4.78E-06**	2.86E+03	1.48E+04	7.16E+02	1.52E+01	1.66E+03	2.32E+03	2.40E+03	1.18E+04	1.46E+02	4.09E+03
	Std	**8.15E-06**	9.49E+02	3.97E+03	6.90E+02	2.16E+00	3.90E+02	1.53E+03	2.35E+03	1.62E+03	1.91E+02	1.25E+03
	Min	**1.41E-07**	1.08E+03	7.65E+03	2.35E+01	1.21E+01	9.81E+02	3.02E+02	1.00E+02	8.60E+03	5.08E+00	2.44E+03
	Max	**3.11E-05**	4.99E+03	2.47E+04	3.27E+03	2.27E+01	2.39E+03	5.71E+03	9.02E+03	1.43E+04	1.01E+03	7.34E+03
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F5 Sumpow	**Mean**	**3.28E-18**	5.69E-08	3.43E-01	6.73E-06	4.78E-07	2.94E-04	7.92E-10	2.79E-06	7.59E-01	6.15E-05	2.39E-02
	Std	**1.01E-17**	2.95E-08	2.60E-01	1.08E-05	6.68E-07	2.71E-04	3.68E-09	4.67E-06	3.00E-01	1.38E-04	2.15E-02
	Min	**1.20E-22**	1.13E-08	2.36E-02	2.52E-17	1.08E-08	9.78E-06	1.27E-15	2.74E-08	1.87E-01	1.86E-07	2.92E-03
	Max	**5.33E-17**	1.17E-07	9.53E-01	4.87E-05	3.45E-06	9.12E-04	2.01E-08	1.70E-05	1.50E+00	6.89E-04	8.04E-02
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F6 Ackley	**Mean**	**3.27E+00**	1.57E+01	1.95E+01	1.01E+01	3.09E+01	1.38E+01	1.85E+01	1.85E+01	2.03E+01	1.10E+01	1.61E+01
	Std	8.23E-01	9.40E-01	5.71E-01	2.03E+00	3.97E-01	9.09E-01	1.41E+00	1.60E+00	**2.54E-01**	2.36E+00	8.56E-01
	Min	**2.08E+00**	1.37E+01	1.82E+01	5.58E+00	3.02E+01	1.18E+01	1.42E+01	1.31E+01	1.98E+01	5.46E+00	1.44E+01
	Max	**5.68E+00**	1.74E+01	2.05E+01	1.38E+01	4.08E+01	1.53E+01	2.01E+01	1.97E+01	2.08E+01	1.74E+01	1.81E+01
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F7 Griewank	**Mean**	**1.78E-02**	2.31E+02	5.28E+02	2.95E+01	3.06E+00	6.38E+01	7.27E+01	6.58E+01	5.59E+02	5.77E+00	2.56E+02
	Std	**2.43E-02**	6.55E+01	9.93E+01	1.73E+01	4.47E-01	1.30E+01	8.93E+01	6.33E+01	7.47E+01	5.48E+00	7.11E+01
	Min	**1.14E-06**	1.06E+02	3.09E+02	6.04E+00	2.08E+00	3.96E+01	1.04E+00	3.51E+00	3.73E+02	1.15E+00	1.45E+02
	Max	**8.50E-02**	4.14E+02	6.91E+02	7.76E+01	3.78E+00	9.35E+01	3.62E+02	1.87E+02	7.23E+02	2.83E+01	4.39E+02
	Rank Sum		+	+	+	+	+	+	+	+	+	+
F8 Powell	**Mean**	**5.79E-02**	2.43E+02	1.40E+04	3.58E+02	2.25E+01	3.98E+02	2.67E+03	2.98E+03	1.56E+04	1.49E+02	2.35E+03
	Std	**4.07E-02**	3.06E+02	4.03E+03	5.01E+02	8.61E+00	9.86E+01	1.97E+03	2.01E+03	3.57E+03	1.14E+02	1.10E+03
	Min	**1.85E-02**	1.67E+01	6.44E+03	3.04E+01	8.69E+00	2.04E+02	1.79E+02	1.76E+02	1.10E+04	1.99E+01	1.02E+03
	Max	**1.76E-01**	1.57E+03	2.24E+04	2.49E+03	4.77E+01	6.25E+02	6.93E+03	7.90E+03	2.35E+04	4.62E+02	6.55E+03
	Rank Sum		+	+	+	+	+	+	+	+	+	+

		Prop. PSO	**DNLPSO**	**ELPSO**	**BBPSOV**	**AGPSO**	**GM-PSO**	**MS-PSO**	**GPSO**
F1 Dixon	**Mean**	**4.77E+00**	1.68E+03	1.00E+01	6.20E+04	2.68E+01	1.42E+01	1.29E+05	2.32E+06
	Std	3.16E+00	8.09E+03	**0.00E+00**	6.68E+04	2.49E+01	6.17E+00	3.74E+04	5.32E+05
	Min	**6.70E-01**	6.82E-01	1.00E+01	3.45E+03	3.92E+00	6.54E+00	6.68E+04	1.57E+06
	Max	**9.82E+00**	4.45E+04	1.00E+01	3.45E+05	1.01E+02	3.33E+01	2.01E+05	3.79E+06
	Rank Sum		+	+	+	+	+	+	+
F2 Sphere	**Mean**	**5.21E-08**	1.89E+00	5.12E+00	1.77E+01	5.96E-03	1.10E-02	4.58E+01	1.72E+02
	Std	1.03E-07	6.02E+00	**1.81E-15**	8.31E+00	5.45E-03	6.37E-03	7.69E+00	2.10E+01
	Min	**9.30E-10**	5.18E-07	5.12E+00	5.80E+00	8.21E-04	1.28E-03	3.11E+01	1.27E+02
	Max	**3.22E-07**	2.43E+01	5.12E+00	4.85E+01	2.36E-02	2.52E-02	6.19E+01	2.19E+02
	Rank Sum		+	+	+	+	+	+	+
F3 Rothyp	**Mean**	**1.88E-04**	1.02E+04	6.55E+01	6.39E+04	3.18E+03	1.92E+01	1.66E+05	6.46E+05
	Std	3.94E-04	2.84E+04	**4.34E-14**	3.74E+04	9.14E+03	1.25E+01	3.10E+04	1.15E+05
	Min	**3.84E-06**	1.63E-03	6.55E+01	1.74E+04	1.81E+00	7.20E+00	1.22E+05	4.34E+05
	Max	**1.74E-03**	1.30E+05	6.55E+01	1.78E+05	3.86E+04	5.50E+01	2.28E+05	8.94E+05
	Rank Sum		+	+	+	+	+	+	+
F4 Sum2	**Mean**	**4.78E-06**	4.03E+02	1.00E+01	1.61E+03	1.16E+02	9.42E-01	3.72E+03	1.56E+04
	Std	8.15E-06	8.88E+02	**0.00E+00**	1.06E+03	2.19E+02	6.14E-01	7.13E+02	2.30E+03
	Min	**1.41E-07**	1.60E-03	1.00E+01	4.81E+02	9.03E-02	1.69E-01	2.46E+03	1.07E+04
	Max	**3.11E-05**	3.51E+03	1.00E+01	5.01E+03	9.00E+02	2.83E+00	4.86E+03	1.96E+04
	Rank Sum		+	+	+	+	+	+	+
F5 Sumpow	**Mean**	**3.28E-18**	9.90E-08	3.42E-08	4.45E-06	3.34E-14	7.40E-16	8.49E-05	4.24E-01
	Std	**1.01E-17**	3.18E-07	1.60E-07	7.78E-06	9.51E-14	2.86E-15	8.57E-05	3.32E-01
	Min	1.20E-22	**5.65E-37**	9.17E-19	6.77E-10	1.37E-18	9.37E-21	4.08E-06	2.22E-02
	Max	**5.33E-17**	1.58E-06	8.76E-07	3.00E-05	3.84E-13	1.55E-14	3.03E-04	1.28E+00
	Rank Sum		+	+	+	+	+	+	+
F6 Ackley	**Mean**	3.27E+00	6.89E+00	**2.37E+00**	1.25E+01	5.84E+00	2.81E+00	1.54E+01	1.99E+01
	Std	8.23E-01	2.74E+00	4.82E-01	1.95E+00	1.35E+00	5.00E-01	6.32E-01	**2.73E-01**
	Min	2.08E+00	2.81E+00	**1.32E+00**	7.41E+00	3.46E+00	1.76E+00	1.40E+01	1.93E+01
	Max	5.68E+00	1.44E+01	**3.36E+00**	1.55E+01	9.04E+00	3.62E+00	1.67E+01	2.05E+01
	Rank Sum		**0**	**-**	+	**0**	**0**	+	+
F7 Griewank	**Mean**	**1.78E-02**	7.97E+00	2.11E+01	5.71E+01	6.51E-01	3.45E-01	1.51E+02	6.01E+02
	Std	**2.43E-02**	2.38E+01	1.09E+02	2.52E+01	2.93E-01	1.31E-01	2.94E+01	6.90E+01
	Min	**1.14E-06**	2.17E-02	7.88E-01	1.90E+01	1.22E-01	1.13E-01	9.05E+01	4.57E+02
	Max	**8.50E-02**	1.30E+02	6.00E+02	1.21E+02	1.08E+00	6.02E-01	1.97E+02	7.76E+02
	Rank Sum		+	+	+	+	+	+	+
F8 Powell	**Mean**	**5.79E-02**	2.57E+01	5.00E+00	7.78E+02	3.62E+01	6.65E-01	1.51E+03	1.21E+04
	Std	4.07E-02	5.14E+01	**0.00E+00**	7.29E+02	5.09E+01	5.24E-01	3.87E+02	2.71E+03
	Min	**1.85E-02**	6.14E-02	5.00E+00	9.35E+01	4.76E-01	5.15E-02	7.92E+02	7.87E+03
	Max	**1.76E-01**	2.07E+02	5.00E+00	2.94E+03	1.79E+02	2.50E+00	2.32E+03	1.86E+04
	Rank Sum		+	+	+	+	+	+	+

5. CONCLUSIONS

In this research, we have described skin lesion classification using PSO-based feature optimization. The proposed PSO model integrates diverse alternative velocity updating strategies in the subswarms to enable a wider exploration of the search space. Two remote swarm leaders have been employed to lead the subswarm-based search to explore distinctive regions. Probability distributions and dynamic matrix representation have also been utilized to increase diversification. The proposed PSO model is capable of mitigating premature convergence of the original PSO model. Evaluated using several skin lesion and UCI data sets, the empirical

results indicate that the proposed search mechanisms account for the superior capabilities of the proposed model. The statistical Wilcoxon rank sum test results further ascertain the efficiency of the proposed PSO model over 10 other classical search methods and 8 advanced PSO variants for solving discriminative feature selection and mathematical optimization problems with different landscapes.

For further work, the proposed PSO model can be improved in several aspects. The acceleration constants currently are fixed throughout the search process. A strategy to adaptively decrease c_1 and increase c_2 learning parameters can be studied, which enables the search process to concentrate on the exploration of the search space in early iterations and converge towards the global optimum solution in subsequent iterations [17, 65]. Chaotic accelerated attraction and evading actions could also be considered. Currently, two subswarms are employed in the proposed model with an attempt to achieve the best trade-off between computational efficiency and swarm diversity. In this regard, we will incorporate diverse hybrid search operations in multiple (>2) subswarms to further evaluate the efficiency of the proposed model. Furthermore, we aim to use the proposed PSO model for other medical imaging tasks such as retinal disease and blood cancer detection using microscopic images. Deep and ensemble neural networks combined with clustering methods [66-70] will be employed to detect the arrival of any unseen new types of skin lesions. The optimal network topologies and hyper-parameters of such classification models will also be explored using the proposed PSO model.

ACKNOWLEDGEMENT

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

REFERENCES

[1] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature*. 542 (2017) 115–118.

[2] Q.U. Ain, B. Xue, H. Al-Sahaf, M. Zhang, Genetic Programming for Skin Cancer Detection in Dermoscopic Images, In *Proceedings of IEEE Congress on Evolutionary Computation (CEC)*, Spain. 2017, pp. 2420–2427.

[3] O. Abuzaghleh, B.D. Barkana, M. Faezipour, Noninvasive Real-Time Automated Skin Lesion Analysis System for Melanoma Early Detection and Prevention, *IEEE Journal of Translational Engineering in Health and Medicine*. Article Sequence Number 4300212 (3) (2015).

[4] F. Xie, H., Fan, Y. Li, Z. Jiang, R., Meng, A. Bovik, Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model, *IEEE Transactions on Medical Imaging*. 36 (3) (2017) 849–858.

[5] K. Shimizu, H. Iyatomi, M.E. Celebi, K.A. Norton, M. Tanaka, Four-Class Classification of Skin Lesions with Task Decomposition Strategy, *IEEE Transactions on Biomedical Engineering*. 62 (1) (2015) 274–283.

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

[7] C. Doukas, P. Stagkopoulos, C.T. Kiranoudis, I. Maglogiannis, Automated Skin Lesion Assessment using Mobile Technologies and Cloud Platforms, In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS),* USA. 2012, pp. 2444–2447.

[8] C. Barata, M.E. Celebi, J.S. Marques, J. Rozeira, Clinically inspired analysis of dermoscopy images using a generative model, *Computer Vision and Image Understanding*. 151 (2016) 124–137.

[9] J. Glaister, A. Wong, D.A. Clausi, Segmentation of Skin Lesions from Digital Images Using Joint Statistical Texture Distinctiveness, *IEEE Transactions on Biomedical Engineering*. 61 (4) (2014) 1220–1230.

[10] J. Glaister, R. Amelard, A. Wong, D.A. Clausi, MSIM: Multistage Illumination Modeling of Dermatological Photographs for Illumination-Corrected Skin Lesion Analysis, *IEEE Transactions on Biomedical Engineering*. 60 (7) (2013) 1873–1883.

[11] N. Lynn, P.N. Suganthan, Ensemble particle swarm optimizer, *Applied Soft Computing*. 55 (2017) 533–548.

[12] J. Gou, Y.X. Lei, W.P. Guo, C. Wang, Y.Q. Cai, W. Luo, A novel improved particle swarm optimization algorithm based on individual difference evolution, *Applied Soft Computing*. 57 (2017) 468–481.

[13] J. Dash, B. Dam, R. Swain, Optimal design of linear phase multi-band stop filters using improved cuckoo search particle swarm optimization, *Applied Soft Computing*. 52 (2017) 435–445.

[14] H.B. Ouyang, L.Q. Gao, S. Li, X.Y. Kong, Improved global-best-guided particle swarm optimization with learning operation for global optimization problems, *Applied Soft Computing*. 52 (2017) 987–1008.

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

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

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

[18] X. Chen, H. Tianfield, C. Mei, W. Du, G.Liu, Biogeography-based learning particle swarm optimization, *Soft Computing*. 21 (24) (2017) 7519–7541.

[19] X. Wang, J. Yang, X. Teng, W. Xia, R. Jensen, Feature selection based on rough sets and particle swarm optimization, *Pattern Recognition Letters*. 28 (2007) 459–471.

[20] Y. Zhang, D. Gong, Y. Hu, W. Zhang, Feature selection algorithm based on bare bones particle swarm optimization, *Neurocomputing*. 148 (2015) 150–157.

[21] Y. Zhang, D. Gong, J. Cheng, Multi-Objective Particle Swarm Optimization Approach for Cost-Based Feature Selection in Classification, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 14 (1) (2017) 64–75.

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

[23] R. Ahila,V. Sadasivam, K. Manimala, An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances, *Applied Soft Computing*, 32 (2015) 23–37.

[24] P. Moradi, M. Gholampour, A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy, *Applied Soft Computing*, 43 (2016) 117–130.

[25] R. Sheikhpour, M.A. Sarram, R. Sheikhpour, Particle swarm optimization for bandwidth determination and *feature selection* of kernel density estimation based classifiers in diagnosis of breast cancer, *Applied Soft Computing*, 40 (2016) 113–131.

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

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

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

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

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

[31] Q. Chen, Y. Chen, W. Jiang, Genetic Particle Swarm Optimization-Based Feature Selection for Very-High-Resolution Remotely Sensed Imagery Object Change Detection, *Sensors*. 16 (8) (2016) 1204.

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

[33] L. Zhang, W. Srisukkham, S.C. Neoh, C.P. Lim, D. Pandit, Classifier Ensemble Reduction Using a Modified Firefly Algorithm: An Empirical Evaluation, *Expert Systems with Applications*. 93 (2018) 395–422.

[34] R. Shang, W. Wang, R. Stolkin, L. Jiao, Subspace learning-based graph regularized feature selection, *Knowledge-Based Systems*. 112 (2016) 152–165.

[35] R. Shang, Z. Zhang, L. Jiao, C. Liu, Y. Li, Self-representation based dual-graph regularized feature selection clustering, *Neurocomputing*. 171 (2016) 1242–1253.

[36] R. Shang, Z. Zhang, L. Jiao, W. Wang, S. Yang, Global discriminative-based nonnegative spectral clustering, *Pattern Recognition*. 55 (2016) 172–182.

[37] L. Yang, S. Yang, S. Li, R. Zhang, F. Liu, L. Jiao, Coupled compressed sensing inspired sparse spatial-spectral LSSVM for hyperspectral image classification, *Knowledge-Based Systems*. 79 (2015) 80–89.

[38] J. Kennedy, R. Eberhart, Particle swarm optimization, In *Proceedings of IEEE International Conference on Neural Networks*, 4, 1995, pp. 1942–1948.

[39] C. Beyan, R. Fisher, Classifying imbalanced data sets using similarity based hierarchical decomposition, *Pattern Recognition*. 48 (2015) 1653–1672.

[40] X.S. Yang, A.H. Gandomi, Bat algorithm: a novel approach for global engineering optimization, *Engineering Computations*. 29 (5) (2012) 464–483.

[41] Z.W. Geem, J.H. Kim, G.V. Loganathan, A new heuristic optimization algorithm: harmony search, *Simulation.* 76 (2) (2001) 60–68.

[42] J.H. Holland, Adaptation in Natural and Artificial Systems, Ann Arbor, MI, USA: University of Michigan Press. 1975.

[43] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, *Neural Computing and Applications*. 27 (4) (2016) 1053–1073.

[44] X.S. Yang, Flower Pollination Algorithm for Global Optimization, in: Durand-Lose J., Jonoska N. (Eds.), Unconventional Computation and Natural Computation, UCNC 2012. *Lecture Notes in Computer Science*, Vol 7445. Springer, Berlin, Heidelberg, 2012, pp. 240–249.

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

[46] D. Karaboga, B. Basturk, On the performance of artificial bee colony (ABC) algorithm*, Applied Soft Computing*. 8 (1) (2008) 687–697.

[47] R.G. Reynolds, An Introduction to Cultural Algorithms, In *Proceedings of the Third Annual Conference on Evolutionary Programming*. World Scientific, River Edge, New Jersey, 1994, pp. 131–139.

[48] J. Kennedy, Bare bones particle swarms, In *Proceedings of the 2003 IEEE Swarm Intelligence Symposium*, 2003, pp. 80–87.

[49] I. Kononenko, Estimating attributes: analysis and extensions of Relief, In: L.De Raedt and F. Bergadano (eds.): *Machine Learning: ECML-94*. Springer Verlag. 1994, pp. 171–182.

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

[dataset] [51] L. Ballerini, R.B. Fisher, R.B. Aldridge, J. Rees, A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-melanoma Skin Lesions, in: M.E. Celebi, G. Schaefer (Eds.), *Color Medical Image Analysis*, *Lecture Notes in Computational Vision and Biomechanics* 6, 2013. pp. 63–86.

[dataset] [52] T. Mendonca, P.M. Ferreira, J.S. Marques, A.R. Marcal, J. Rozeira, PH2 - a dermoscopic image database for research and benchmarking, In *Proceedings of EMBC 2013 - 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Japan, 2013, pp. 5437–5440.

[dataset] [53] Dermnet Skin Disease Image Atlas. http://www.dermnet.com, 2017, (accessed 22 May 2017).

[54] N. Hansen, A. Auger, R. Ros, S. Finck, P. Pošík, Comparing Results of 31 Algorithms from the Black-Box Optimization Benchmarking BBOB-2009. In *GECCO'10 Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, USA, 2010, pp. 1689–1696.

[55] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, *Swarm and Evolutionary Computation*. 1 (2011) 3–18.

[56] O. Abuzaghleh, B.D. Barkana, M. Faezipour, SKINcure: A real time image analysis system to aid in the malignant melanoma prevention and early detection, In *Proceedings of IEEE Southwest Symposium on Image Analysis and* Interpretation (SSIAI). 2014, pp. 85–88.

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

[58] C. Barata, J.S. Marques, J. Rozeira, Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model, In *Proceedings of International Symposium on Visual Computing*. Springer, Berlin, Heidelberg. 2013, pp. 40–49.

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

[60] M.J.M. Vasconcelos, L. Rosado, M. Ferreira, A new color assessment methodology using cluster-based features for skin lesion analysis, In *Proceedings of 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 373–378.

[61] R.S. Soumya, S. Neethu, T.S. Niju, A. Renjini, R.P. Aneesh, Advanced earlier melanoma detection algorithm using colour correlogram, In *Proceedings of International Conference on Communication Systems and Networks (ComNet)*, 2016, pp. 190–194.

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

[dataset] [63] M. Lichman, UCI Machine Learning Repository. http://archive.ics.uci.edu/ml, Irvine, CA: University of California, 2013, (accessed on 13 Jan 2018).

[64] L.N. Zhang, L.Q. Liu, G.N. Yuan, Y.T. Dai, Modified Firefly Algorithm Using Randomized Mechanisms, In *Proceedings of IEEE Congress on Evolutionary Computation (CEC).* Canada. 2016, pp. 2255–2261.

[65] B. Xue, M. Zhang, W.N. Browne, Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach, *IEEE Transactions on Cybernetics*. 43 (6) (2013) 1656–1671.

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

[67] P. Kinghorn, L. Zhang, L. Shao, A Region-based Image Caption Generator with Refined Descriptions, *Neurocomputing*. 272 (2018) 416–424.

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

[69] Y. Zhang, L. Zhang, M.A. Hossain, Adaptive 3D facial action intensity estimation and emotion recognition, *Expert Systems with Applications*. 42 (2015) 1446–1464.

[70] P. Kinghorn, L. Zhang, L. Shao, A Hierarchical and Regional Deep Learning Architecture for Image Description Generation, *Pattern Recognition Letters*. (In Press).