Revista Argentina de Clínica Psicológica
2023, Vol. XXXII, N°1, 05-12
DOI: 10.24205/03276716.2023.7002

A Parallel Biogeography Optimization-Based Feature Selection Architecture for Fundus Classification

C. K. Sarumathiy[a], Dr. K. Geetha[b], Dr.C. Rajan[c]

Abstract

Diabetic Retinopathy (DR) is one of most common eye disease suffered by people with diabetics. The DR can be detected and classified using fundus retina images by employing efficient feature selection techniques. One of the challenges in fundus image classification is the high dimensionality of the features extracted from the images. There are several techniques for classification and feature extraction that are suggested earlier for texture analysis. Feature selection is utilized to find the optimal feature subset. In this work, various metaheuristic methods like Biogeography-based Optimization (BBO) and Particle Swarm Optimization (PSO) are used to find the optimal features. These algorithms are based on the population that explores the space of a certain problem to identify optimal parameters. The Support Vector Machine (SVM) is a machine learning that works well for the problems of binary classification. The proposed methods optimize the feature selection and SVM parameters to enhance image classification.

Keywords: Biogeography-based Optimization (BBO), Particle Swarm Optimization (PSO), Support Vector Machine (SVM), parallel Biogeography-based Optimization and AdaBoost, feature selection, precision, recall

1. Introduction

Diabetes Mellitus is a serious health problem affecting millions of people across the world. Emerging Asian countries like India is seeing a rising prevalence of diabetics. Diabetic Retinopathy is largely asymptomatic in early stages due to the microvascular changes in the retina. Regular screening is required for timely diagnosis and management of DR. Fundus images is used to observe the retina lesions at high resolution for diagnosing DR and assess its severity. As high level of expertise, time and effort is required by the professional ophthalmologist, automated systems for detection of DR are being sought after (Tsiknakis et al., 2021).

The mechanisms of image recognition are challenging owing to the images having high dimensional space that can degrade the image classification performance in connection to accuracy and time taken to build the model. Thus, feature selection is employed for reducing the dimensionality of the feature set extracted from images which are done by the removal of redundant and irrelevant features. For image classification and image recognition, the algorithm for classification is employed for building the classifier, which learns from the feature space of the image dataset with a class attribute. The classifier is used for recognizing and predicting unlabelled images.

[a]*Department of Computer Science, Excel Engineering College,* Tamil Nadu, India. Email: cksarumathiy@gmail.com
[b]*Department of Computer Science, Excel Engineering College,* Tamil Nadu, India. Email: geetharajsri@gmail.com
[c]*Department of IT, K S Rangasamy College of Technology, Tamil Nadu, India. Email:* rajancsg@gmail.com

The wavelet transform-based algorithm of texture classification has various crucial traits. A wavelet transform can de-correlate data and achieve the same kind of goal as in the case of the linear transformation employed (Hiremath & Shivashankar, 2006). It provides information that is orientation-sensitive and essential in the analysis of texture. The complexity of computation is lower to a significant extent by considering the wavelet decomposition. It is understood that both context and position information for an image pixel is crucial for classification purposes.

Feature selection has a major role to play in various problems of pattern recognition, like the classification of images (Song & Tao, 2009). Since there are many features that are used for characterizing images, only some of them can be effective and efficient in classification. More features do not necessarily mean better performance of classification, and so feature selection is performed for choosing a relevant and compact feature subset to reduce feature space dimensionality, and this improves the accuracy of classification, thereby reducing the consumption of time. Based on various evaluation criteria, the methods of feature selection are grouped into two: the filter and the wrapper models. The former makes use of feature data characteristics and is computationally more efficient. Selecting optimal features is an NP-hard problem, and metaheuristic algorithms are widely used for solving it. In this work, we used Particle Swarm Optimization (PSO) and Biogeography-based Optimization (BBO)

(Jaafari et al., 2019) for optimizing feature selection. PSO is a method that is used for training classifiers within a short period of time. In the PSO algorithm, every iteration is known for training the new and weak classifier, and once this is run many times, an optimal feature subset is found to be a stronger classifier. BBO is stochastic and based on the population used for problem-solving. The study of the geographical distribution of certain biological organisms is known as Biogeography. This simulates the mutation and migration of the species in different habitats (and islands).

Classification refers to the process that groups and collects objects, ideas, or data into several groups in which every member has a similar trait. For purposes of classification, the classes may not be contested before the examination of data and are referred to as supervised learning. The Support Vector Machine (SVM) is used for the classification of data, both linear and nonlinear (Zhang, 2021). The SVM having nonlinear mapping functions, converts original data into that of higher dimensions which is done at the time data is not separated. The Ada Boost is a process used in empirical studies that have been gaining attention from the community of machine learning recently. The rest of the paper presents the literature survey, methodology, results, and conclusion.

2. Literature Survey

Hiremath and Shivashankar (2006) proposed an algorithm for feature extraction with the decomposed images of the wavelet along with its complementary image for the classification of texture. These features will be constructed using various combinations of images and offer better strategies of discrimination for the classification of textures and also for enhancing the rate of classification.

Zhou (2015) proposed a technique for feature selection that combined the reliefF with the SVM-RFE algorithm. There was an integration of the weight vector from reliefF into the SVM-RFE method. Here, the reliefF will filter out several noisy features in the initial stage itself, and after this, another ranking criterion that is based on the VM-RFE method will be applied to get its final feature subset. An SVM classifier will calculate the accuracy of the classification. The experimental results proved that the proposed relief-SVM-RFE achieves a significant level of improvement in image classification.

Takruri, Mahmoud, and Al-Jumaily (2019) introduced a new automated system used for the detection of skin cancer (melanoma) from the histopathological images that were sampled with microscopic slides in skin biopsy. This system is hybrid in nature that is based on PSO-SVM. These features were extracted using the grayscale image-based histogram along with the co-occurrence matrix along with the wavelet coefficient energy that is a result of the decomposition of wavelet packets. A PSO-SVM system can choose an ideal feature set along with the best values used for the parameters of the SVM (C and γ) for optimizing the SVM classifier's performance. This was tested based on real datasets that were procured from the Southern Pathology Laboratory. The findings of the evaluation showed an accuracy of 87.13, and the sensitivity and specificity were equivalent to the results from dermatologists.

A Biogeography-based algorithm for choosing optimal feature sets from data was proposed by Shahid, Javed, and Zafar (2017). After this, the techniques of Naïve Bayes with Support Vector Machine were employed in order to perform product review classification. This technique was applied to the other problems in the classification that have large feature sets. Panchal et al. (2009) focussed on satellite image classification using the BBO optimizer. The BBO algorithm originally did not permit the inbuilt clustering property needed at the time of image classification. Thus, modifications were proposed to its original algorithm. The modified algorithm was employed for satellite image classification. The results proved higher accuracy extracted effectively.

A new and improved Ada Boost algorithm that was based on the search for optimization within the search space was proposed by Mohammadpour, Ghorbanian, and Mozaffari (2016). Using large-scale data will require more time in comparing samples to identify an Ada Boost algorithm threshold while making use of a decision stump to be a weak classifier. The PSO was used to develop and choose the best feature within the sample space for the reduction of time. The experiments demonstrated that the application of PSO to a decision stump could consume the Ada Boost time. The result of this was using evolutionary algorithms with large-scale issues to identify the best solution and to increase performance.

3. Methodology

A. *Dataset*

Fundus images captured using fundus cameras are used for evaluation. The images were collected through EyePACS (Picture Archiving and Communication System) clinics and collated as dataset and presented in Kaggle. High-resolution retinal images acquired from diverse conditions form the dataset. The dataset is annotated by expert ophthalmologists. Every image will be allocated the following DR grades from a 0 to 4 scale: 0 = No DR, 1 = Mild, 2 = Moderate, 3 = Severe, and 4 = Proliferative DR(PDR). The Kaggle training set will be used to pick a subset having 6,230 distinct images (Asiri et al., 2019).

B. *Wavelet-based Texture feature*

A Wavelet transform processes mathematical signals used for solving complex issues. It has a great performance of analysis with multi-resolution traits making it worthy of image processing. For a multi-resolution and localization analysis of any signal, a wavelet transform may be utilized. Furthermore, it may be appropriate for complex images in edge detection (Hu, Jiang, & Bo, 2009). A wavelet transform has been defined to be a linear operation wherein a signal will decompose into mechanisms that are at altered scales. The Wavelet functions ψ (t) will be well-defined within a space which has measurable functions and are absolute and square integrals as in equation (1):

$$\int_{-\infty}^{+\infty} |\psi(t)|\ dt < \infty$$

$$\int_{-\infty}^{+\infty} |\psi(t)|^2\ dt < \infty \qquad (1)$$

The conditions of zero mean and square norm one are gratified:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0$$

$$\int_{-\infty}^{+\infty} |\psi(t)|^2\ dt = 1 \qquad (2)$$

The wavelet transform of a function $f(t) \in L^2(R)$ at scale a and position τ is represented in equation (3):

$$Wf(a,\tau) = \frac{1}{\sqrt{2}}\int_{-\infty}^{+\infty} f(t)\Psi^*(\frac{t-\tau}{a})dt \qquad (3)$$

Here, an asterisk * will signify a complex conjugation. Equation (3) refers to the analysis of a signal which is f(t), that gets convolved using dilated or stretched copies in the mother ψ (t). For a < 1, the wavelet will contract with the transform stretching data to the finer particulars of the f(t). For a > 1, the same wavelet will enlarge and further transform, giving a rough view of its signal. The roles of a wavelet transform found in continuous time on the functions and in the discrete-time on the vectors are employed. These wavelet coefficients originate by means of assessing the integral ones. In discrete time, all coefficients will originate through passing the vector (x(n), n integer) by means of a two-filter bank of low-pass and high-pass.

The primary benefit of the wavelet analysis was that it permitted the usage of long-term intervals in which precise and low-frequency information was obtained, and shorter high-frequency information was sought. Thus, wavelet analysis was proficient in showing the different characteristics of data that were missed by other techniques of image analysis that discontinue in self-similarity and higher derivatives. The wavelets could also compress and de-noise images without degrading the original images.

C. *Particle Swarm Optimization with Support Vector Machine (PSO-SVM)*

A stochastic technique of optimization based on population mimicking the swarm movement and stimulated by the social behaviour of fish or birds is known as PSO. This works with a population of candidate solutions termed particles (Tharwat & Hassanien, 2019). Every particle Xik will be moved within the search space in accordance with (2) and (3) that is guided by the best-known position $pibestk$ found within the search space. There is also the whole swarm's best position known as $gbestk$ with its own velocity Vik (that is bound with a maximum value $Vmaxk$). This process was iterative. For every iteration Xik, ik, $pibestk$, and $gbestk$ will be updated, and when the improved positions are found, they will guide the swarm movements.

$$v_i^{t+1} = w * v_i^t + c_1 * rand * (pbest - x_i^t) + c_2 * rand * (g\,best - x_i^t) \quad x_i^{t+1} = x_i^t + v_i^{t+1} \qquad (4)$$

Here, the current generation k, personal and social factors of learning $c1$and $c2$ are positive constants, and random numbers are $r1$ and $r2$ from an interval [0, 1].

The SVM makes use of kernel tricks for connecting a training sample space to the feature space of a high dimension to identify an optimal separator hyperplane. RBF (Radial Basis Function) kernel having gamma parameters will be used. For control of the model's complexity and their training errors, a regulatory parameter C is employed. to select the correct gamma with the C values, the overfitting problem is solved. Another low parameter C value will make a smooth decision, and the high C goals will classify the training samples correctly. The primary advantage of a linear SVM was the speed of execution that is fast without any parameters. They are intolerant of the relative sizes of training examples of both classes. For the learning algorithms, in case there are some more examples of a single class, the algorithm will classify this class with more examples, thus reducing the rate of error. As the SVMs do not attempt to bring down the rate of errors but try separating the high dimensional patterns, the SVM results will be insensitive to the numbers in every class.

The SVM parameters for optimally choosing the PSO-SVM system are (Takruri et al., 2019):

- Initialization: Every particle has been defined to be an array representing the C and γ parameters, and each cell corresponds to each feature. These feature cells consist of weights as 0 and 1 that signify the actual importance of the features. For the initial population, the PSO particles are created using random numbers that range between 0 and 1.
- The Selection of features for a particle: For every iteration, the features having weights above a particular threshold (0.5) are sent to the SVM to calculate the accuracy that is based on the selected features.
- Fitness function: For the computation of fitness (or accuracy) for every particle, the SVM will be loaded with the chosen features of the particle with C and γ parameters. The SVM will be trained using a training set, and its performance will be verified based on the validation set and its feature set that is taken to the particle's fitness. This is the same case for each particle.

Once the accuracy of the particles is evaluated, the accuracy that is the best will be taken as $gbestk$. The best one in the history of all particles is taken to be the $pibestk$ for that particle. The particles of subsequent populations will be generated based on equations (1) and (2). The first population for $gbestk$ and $pibestk$ are clearly the same. This process will stop as soon as the maximum iteration is met. The particle that corresponds to the $pibestk$ consists of the optimal features and parameters; thus, it will be used for optimizing the SVM and choosing the relevant features.

D. *Proposed Parallel Biogeography-based optimization with SVM optimization*

The BBO is based on the study of the geographical dissemination of various biological organisms. The BBO algorithm works by simulating the migration and mutation of the species in various habitats (Simon, 2008). This algorithm shares some characteristics with GA and PSO. The BBO ideally begins the search by the initialization of habitats (that are equivalent to the chromosomes in the GA) that are composed of other species. If it is assumed there are some T habitats, they are grouped from the best to the worst, and the species count St that is in the habitat t will be initialized as per equation (5):

$$S_t = T - t - 1 \tag{5}$$

The rate of immigration λSt and emigration µSt for habitat t with the St species was initialized in accordance with equations (6 and 7) (Zhu, 2010):

$$\lambda_{St} = I \times \left(1 - \frac{S_t}{T}\right) \tag{6}$$

$$\mu_{St} = E \times \left(\frac{S_t}{T}\right) \tag{7}$$

Wherein I and E denote the rate of both immigration and emigration constants. At the time of evolution, both of these will be from the habitats that are taken into consideration. If p_{St}^{g} represents a probability of habitat t consists of species S at generation g, then the initial values for the p_{St}^{g} are given as per (8):

$$p_{St}^{g} = 1.0/T \tag{8}$$

For the remaining challenging issues, the BBO can need significant time for computation. If parallelism is added to the BBO, there is a possibility for more efficiency. For parallelization of the evolutionary algorithm, the following types can be used: at the objective function evaluation level, the population level, and the elementary level. The nature of migration of the BBO will be suitable for multi-population of parallel computing. The research further explores the implementation of low levels in which all habitats interact inside one single population. For the proposed variant of BBO, the habitats were initialized randomly. The habitats were candidate solutions. The habitat consisted of variables that were coded using real numbers. Every habitat variable will represent a new dimension variable. At the time of evolution, the habitats and their variables keep improving on the basis of both the emigration and immigration of the species.

This method has been tailor-made for a new environment having T habitats (or threads) operating in a manner of a single interaction with multiple threads. Every thread will perform the operations of the BBO in constructing a new child solution. All habitats are sorted on the basis of costs to identify their best solution on completion of each generation. But a set of the elite habitat solutions are maintained without change to the subsequent generation. The Proposed Parallel Biogeography - based optimization, along with the SVM optimization, has been specified:

Generate the T habitats;
Compute both cost and constraints of the habitats;
Sort all habitats based on their cost;
Initialize the parameters of BBO;
Save current best habitats temporarily;
Initialize the species counts;
Compute rates of both immigration and emigration;
Compute Probability;
Update all habitats with rates of both immigration and emigration;
Evaluate all cost and constraints of the habitats;
Sort all habitats based on cost;
Mutate the worst half of habitats;
Compute both cost and constraints of the habitats;
Improve all habitats using Pattern Search;
Sort all habitats based on their cost;
Replace all worst few habitats along with the saved best habitats;
Sort all habitats on the basis of cost;
Classify the results obtained with the SVM classifier;
Collect the result of generation;
Iterate till maximum number of iterations is met.
Collect final result.

E. AdaBoost Classifier

The Ada Boost is a method of metaheuristics in machine learning which is an ensemble learning family that creates a classifier to be strong. The aim here was to improve the accuracy of a learning algorithm that was by a combination of such weak classifiers. The method has been iterative, and for every one of these, there was a weak classifier to reduce an average training error (Lopez-Garcia et al., 2019).

The Ada Boost will work in a T round, and for every such round, it trains another new weak learner. Once the training is complete, an algorithm will increase the weight of the samples that have been predicted to be misclassified and will decrease the weight of the sample that is correctly classified. This way, the samples are classified correctly, and their chances of decreasing to be reused in the subsequent iteration will also continue to exist. The Ada Boost further takes training samples that are $S = (x_1, y_1), \ldots, (x_n, y_n)$ of size N as input, in which every sample xi depicts a vector that has values for X, the domain space, and yi refers to a label of every sample xi belonging to label space Y. In the beginning, the weight is initialized in a uniform manner across the entire training set, and the weights of the examples are classified incorrectly. Thus, a weak learner focuses on all hard samples. The Ada boost will have a training example weight vector, and every iteration updates the weights of samples with a weight function as per (9):

$$W_{t+1,i} = W_{t,i}\beta_t^{1-bi} \qquad (9)$$

For the purpose of adjusting a probability distribution for the training samples, a calculation was made, and an error rate $\in_j$ for the classifier was made based on each of these samples.

$$\in_j = \sum_{i=1} W_{t,i}.b_i \quad (10)$$

H(x) was a final strong classifier constructed with every one of the weak classifiers h1, h2. In this, each one of them was weighted in accordance with the accuracy of the training phase. Thus, the focus of this was on the patterns or samples that were challenging to classify.

4. Results and Discussion

For experiments, Kaggle dataset with 6,230 fundus images wih 5 classes (No DR, Mild, Moderate, Severe and Proliferative DR) is considered for evaluation. Table 1 shows the simulation parameters of BBO. Table 2 shows the number of features selected. Figure 1 shows some of the sample images used in the investigation. Figures 2 to 5 show the recall, precision, F-Measure, and Misclassification Rate for PSO-Adaboost, PSO-SVM, Parallel BBO-AdaBoost, and Parallel BBO-SVM, respectively.

TABLE I. **SIMULATION PARAMETERS OF BBO**

Parameters	Values
Population size	100
Number of iterations	300
Maximum immigration rate	1
Maximum emigration rate	1
Maximum mutation rate	0.1

TABLE II. **NUMBER OF FEATURES SELECTED**

Technique	Number of features selected
PSO	87
BBO	76

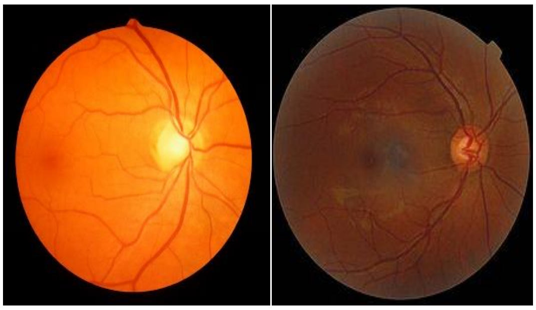

Fig. 1. **Sample Images used in the investigation**

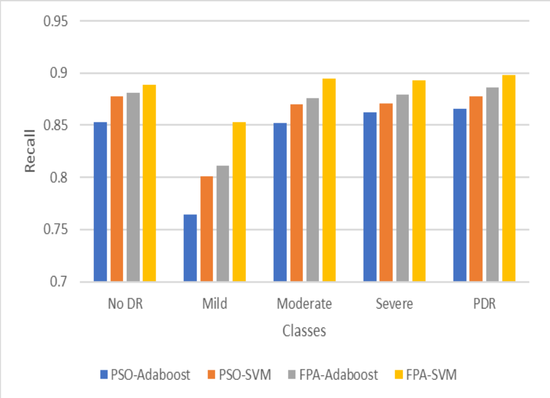

Fig. 2. **Recall for Proposed Parallel BBO-SVM**

Figure 2 shows that the recall of proposed parallel BBO-SVM performs better by 4.13%, 1.3%, and 0.93% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class No DR. The recall of proposed parallel BBO-SVM performs better by 4.81%, 2.79%, and 2.1% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class Moderate. The recall of proposed parallel BBO-SVM performs better by 3.7%, 2.4%, and 1.4% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class PDR.

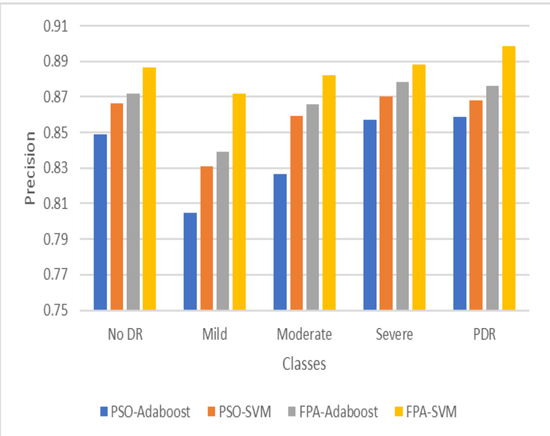

Fig. 3. **Precision for Proposed Parallel BBO-SVM**

Figure 3 shows that the precision of proposed parallel BBO-SVM performs better by 4.51%, 3.21%, and 2.4% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost respectively, for class No DR. The precision of the proposed parallel BBO-SVM performs better by 8.1%, 2.5%, and 1.7% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class Moderate. The precision of the proposed parallel BBO-SVM performs better by 5.45%, 4.6%, and 3.7% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class PDR.

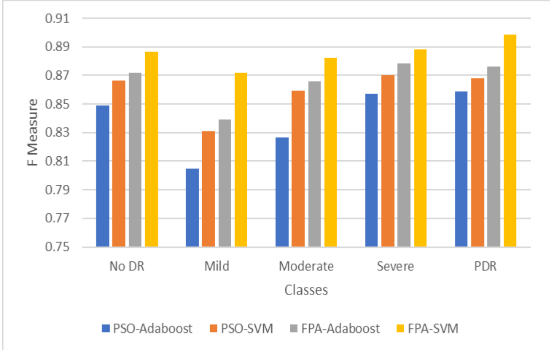

Fig. 4. **F-Measure for Proposed Parallel BBO-SVM**

Figure 4 shows that the F-Measure of the proposed parallel BBO-SVM performs better by 4.32%, 2.3%, and 1.7% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class No DR. The F-Measure of the proposed parallel BBO-SVM performs better by 6.5%, 2.63%, and 1.9% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class Moderate. The F-Measure of the proposed parallel BBO-SVM performs better by 4.6%, 3.5%, and 2.55% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively, for class PDR.

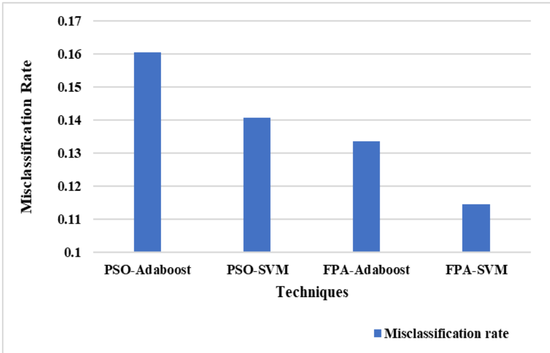

Fig. 5. **Misclassification Rate for Proposed Parallel BBO-SVM**

Figure 5 shows that the Misclassification Rate of proposed parallel BBO-SVM performs better by 33.4%, 20.53%, and 15.32% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively.

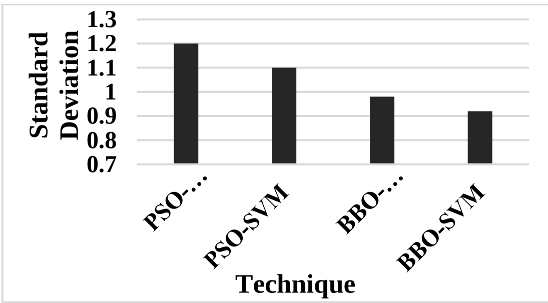

Fig. 6. **Standard Deviation for BBO-Adaboost**

Figure 6 shows that the Standard Deviation of proposed parallel BBO-SVM performs better by 26.42%, 17.82%, and 6.32% than PSO-AdaBoost, PSO-SVM, and Parallel BBO-AdaBoost, respectively.

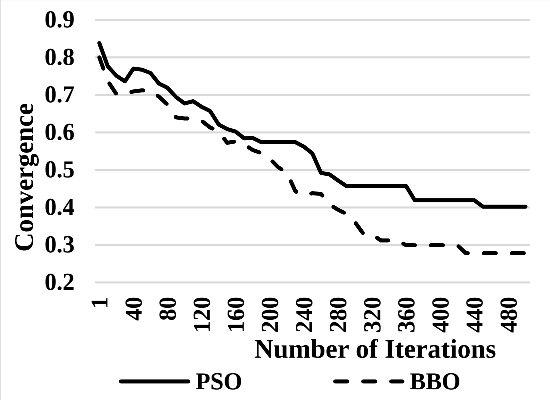

Fig. 7. **Convergence for BBO**

Figure 7 shows that the convergence of the proposed parallel BBO occurs at a number of iterations 430, whereas the convergence of PSO occurred at a number of iterations 450.

5. Conclusion

The classification of DR is crucial for early diagnosis and treatment. Fundus images are widely used for DR detection. In this work, an automated system for detection of classification of DR as No DR, Mild, Moderate, Severe and Proliferative DR is presented. The key challenge in image classification is the high dimensionality of the feature set. As feature selection is an NP-hard problem, optimization techniques are used for selecting optimal feature set. In this work, Biogeography Based Optimization (BBO) is used for the optimization, which is a technique of computational intelligence based on the swarm. The PSO is very promising with fast convergence and can be implemented easily as a machine learning algorithm. It is a swarm-based computational intelligence technique that has already been used for optimization in other problems. The evaluation of the proposed method demonstrates the efficiency in improving recall and precision.

6. Future Work

BBO can be combined with several other different evolutionary algorithms to improve the accuracy of classification.

Acknowledgment

Compliance with Ethical Standards

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Conflict of Interest: Author A declares that she has no conflict of interest. Author B declares that she has no conflict of interest.

References

[1] Asiri, N., Hussain, M., Al Adel, F., & Alzaidi, N. (2019). Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artificial intelligence in medicine, 99*, 101701. https://doi.org/10.1016/j.artmed.2019.07.009

[2] Hiremath, P., & Shivashankar, S. (2006). Wavelet based features for texture classification. *GVIP journal, 6*(3), 55-58. https://www.academia.edu/download/76813265/wavelet_based_texture_features_a_new_method_for_sub_band.pdf

[3] Hu, C., Jiang, L.-J., & Bo, J. (2009). Wavelet transform and morphology image segmentation algorism for blood cell. In *2009 4th IEEE Conference on Industrial Electronics and Applications* (pp. 542-545). IEEE. https://doi.org/10.1109/ICIEA.2009.5138265

[4] Jaafari, A., Panahi, M., Pham, B. T., Shahabi, H., Bui, D. T., Rezaie, F., & Lee, S. (2019). Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena, 175*, 430-445. https://doi.org/10.1016/j.catena.2018.12.033

[5] Lopez-Garcia, P., Masegosa, A. D., Osaba, E., Onieva, E., & Perallos, A. (2019). Ensemble classification for imbalanced data based on feature space partitioning and hybrid metaheuristics. *Applied Intelligence, 49*(8), 2807-2822. https://doi.org/10.1007/s10489-019-01423-6

[6] Mohammadpour, M., Ghorbanian, M., & Mozaffari, S. (2016). AdaBoost performance improvement using PSO algorithm. In *2016 Eighth international conference on information and knowledge technology (IKT)* (pp. 273-275). IEEE. https://doi.org/10.1109/IKT.2016.7777785

[7] Panchal, V., Singh, P., Kaur, N., & Kundra, H. (2009). Biogeography based satellite image classification. *arXiv preprint arXiv:0912.1009*. https://doi.org/10.48550/arXiv.0912.1009

[8] Shahid, R., Javed, S. T., & Zafar, K. (2017). Feature selection based classification of sentiment analysis using biogeography optimization algorithm. In *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICIEECT.2017.7916549
[9] Simon, D. (2008). Biogeography-based optimization. *IEEE transactions on evolutionary computation, 12*(6), 702-713. https://doi.org/10.1109/TEVC.2008.919004
[10] Song, D., & Tao, D. (2009). Biologically inspired feature manifold for scene classification. *IEEE Transactions on Image Processing, 19*(1), 174-184. https://doi.org/10.1109/TIP.2009.2032939
[11] Takruri, M., Mahmoud, M. K. A., & Al-Jumaily, A. (2019). PSO-SVM hybrid system for melanoma detection from histo-pathological images. *International Journal of Electrical and Computer Engineering, 9*(4), 2941-2949. https://doi.org/10.11591/ijece.v9i4.pp2941-2949
[12] Tharwat, A., & Hassanien, A. E. (2019). Quantum-behaved particle swarm optimization for parameter optimization of support vector machine. *Journal of Classification, 36*, 576-598. https://doi.org/10.1007/s00357-018-9299-1
[13] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D. I., & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in biology and medicine, 135*, 104599. https://doi.org/10.1016/j.compbiomed.2021.104599
[14] Zhang, D. (2021). Support Vector Machine. In D. Zhang (Ed.), *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval* (pp. 201-228). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-69251-3_8
[15] Zhou, X. (2015). Feature selection for image classification based on a new ranking criterion. *Journal of Computer and Communications, 3*(03), 54739. https://doi.org/10.4236/jcc.2015.33013
[16] Zhu, W. (2010). Parallel biogeography-based optimization with GPU acceleration for nonlinear optimization. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 44090, pp. 315-323). https://doi.org/10.1115/DETC2010-28102