



Emperor penguin optimization algorithm- and bacterial foraging optimization algorithm-based novel feature selection approach for glaucoma classification from fundus images

Law Kumar Singh¹ · Munish Khanna² · Hitendra Garg¹ · Rekha Singh³

Accepted: 2 May 2023 / Published online: 27 May 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Feature selection is an important component of the machine learning domain, which selects the ideal subset of characteristics relative to the target data by omitting irrelevant data. For a given number of features, there are 2^n possible feature subsets, making it challenging to select the optimal set of features from a dataset via conventional feature selection approaches. We opted to investigate glaucoma infection since the number of individuals with this disease is rising quickly around the world. The goal of this study is to use the feature set (features derived from fundus images of benchmark datasets) to classify images into two classes (infected and normal) and to select the fewest features (feature selection) to achieve the best performance on various efficiency measuring metrics. In light of this, the paper implements and recommends a metaheuristics-based technique for feature selection based on emperor penguin optimization, bacterial foraging optimization, and proposes their hybrid algorithm. From the retinal fundus benchmark images, a total of 36 features were extracted. The proposed technique for selecting features minimizes the number of features while improving classification accuracy. Six machine learning classifiers classify on the basis of a smaller subset of features provided by these three optimization techniques. In addition to the execution time, eight statistically based performance metrics are calculated. The hybrid optimization technique combined with random forest achieves the highest accuracy, up to 0.95410. Because the proposed medical decision support system is effective and ensures trustworthy decision-making for glaucoma screening, it might be utilized by medical practitioners as a second opinion tool, as well as assist overworked expert ophthalmologists and prevent individuals from losing their eyesight.

Keywords Feature selection · Emperor penguin optimization · Bacterial foraging optimization · Hybrid algorithm · Glaucoma prediction

✉ Law Kumar Singh
lawkumarcs1@gmail.com

Munish Khanna
munishkhanna.official@rocketmail.com

Hitendra Garg
hitendra.garg@gmail.com

Rekha Singh
singh.rekha70@gmail.com

¹ Department of Computer Engineering and Applications, G.L.A University, Mathura 281406, India

² Department of Computer Science and Engineering, Hindustan College of Science and Technology, Mathura 281122, India

³ Department of Physics, Uttar Pradesh Rajarshi Tandon Open University, Prayagraj, Uttar Pradesh, India

1 Introduction

A significant amount of high-dimensional data is being produced through the widespread use of social media and sensors. There are several features (characteristics) in the dataset; thus, feature extraction is essential in order to identify and extract the dataset's most usable data. Regression, classification, and clustering become less feasible due to the high-dimensional data's enormously increased space and temporal complexity. The high-dimensional dataset also includes numerous characteristics, some of which are superfluous or unimportant. A classifier's performance suffers as a result of duplicated and irrelevant features. Therefore, feature selection (FS) techniques that identify the optimal subset of features from

large-dimensional datasets are often utilized to address this problem (Mafarja and Mirjalili 2017). FS is essential in order to identify and extract the dataset's most usable data. By reducing redundant and noisy data from the high-dimensional dataset, FS enhances the learning accuracy as well as the clarity of the results (Saraswat and Arya 2014). Using feature selection, the machine learning algorithm can be trained more quickly and easily. Additionally, it reduces the over-fitting issue and complexity of the classifier, making it simpler to understand (Wei et al. 2017). Thus, a proper mining technique must be used to fetch the necessary features from the dataset. To decrease the dimensionality of the feature space, several meta-heuristic techniques have been created and applied in the past as well. FS is an important stage in the design and development of data mining and machine learning algorithms. It facilitates faster calculation and increases classification accuracy. The performance of the classifiers may be hampered by feature spaces that include a significant number of duplicate or irrelevant characteristics. In order to boost the effectiveness of the classifiers, redundant

features are removed from the original set using feature selection techniques to choose the best subset of features.

In computer-assisted diagnosis (CAD), FS chooses the relevant characteristics that have a deeper influence on classification accuracy and discards the features with a lower impact since they would negatively affect the performance of the classification subsystem. The classification of glaucoma is getting more difficult due to the absence of feature selection in computer-aided diagnostics. This is due to the fact that it must produce findings with more precision and that there must be more characteristics available for analysis. The location, few unpredictable development patterns, large dataset, and growing number of factors involved make it difficult to classify glaucoma disorders. The FS subsystem works to eliminate these duplicate features and only chooses the best subsets of features out of all the features in the dataset. Therefore, feature selection will significantly improve the accuracy and efficiency of the CAD system.

Glaucoma, also known as "silent theft of sight," was most likely identified as a disease in the early seventeenth

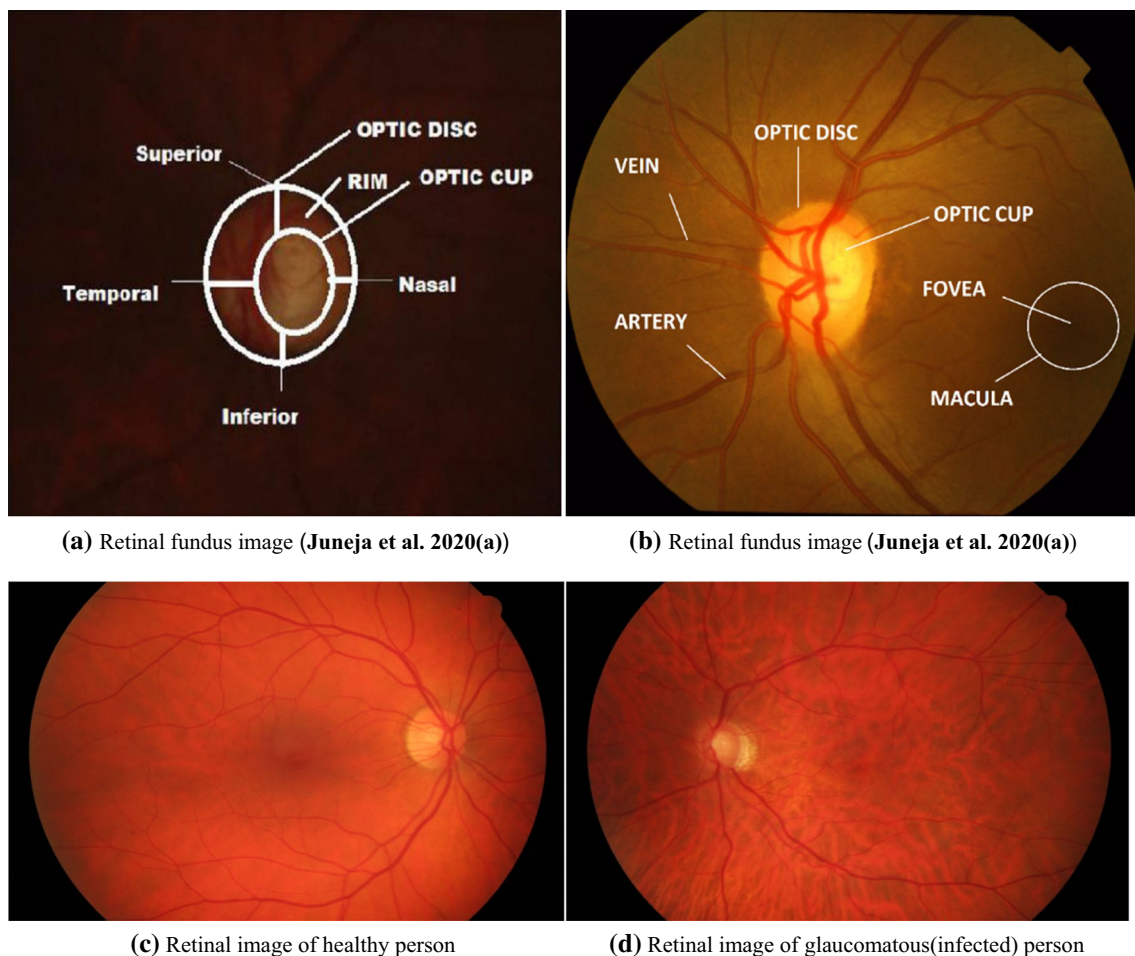


Fig. 1 Retinal fundus images

century when the Greek word *glaukEoma*, which means obscurity of the lens or cataract and denotes ignorance of the condition, was coined. Glaucoma is the second most common cause of blindness worldwide. After cataracts, it is the second most frequent cause of permanent blindness, and if it goes undetected, it might overtake cataracts as the most prevalent cause. Globally, glaucoma affects around 60 million individuals, and by 2020, that figure is predicted to increase to 79.6 million (Singh and Khanna 2022a). The World Health Organization (WHO) reports that glaucoma is one of the main causes of blindness, which has already impacted more than 60 million people worldwide and might reach 80 million very soon. Additionally, 12 million Indians are thought to be affected. An abrupt visual disruption, excruciating eye discomfort, impaired vision, inflamed eyes, and haloes surrounding lights are some of its typical symptoms. People over the age of 40 are often affected by this condition. The brain visualizes the outside world as a consequence of receiving visual information from the retina through the optic nerve. The optic nerve is damaged as a result of a rise in intraocular pressure, which causes the disease to start. Sudden visual disruption, excruciating eye discomfort, hazy vision, inflamed eyes, and haloes surrounding lights are some of its typical symptoms (Singh et al. 2022b). The optic nerve head's structure, the thickness of the retinal nerve fiber layer, the ganglion cell, and the inner plexiform layers all alter as a consequence of the malfunctioning and loss of ganglion cells. Glaucoma may permanently damage the optic nerve if it is not treated, which would result in blindness. As a

result, early detection of glaucoma in its early stages significantly reduces the risk of permanent blindness.

An image of the eye taken with a specialized fundus camera is called a fundus image (Fig. 1). A fundus camera, which comprises a microscope connected with a flash-enabled camera to capture the picture of the internal surface/back of an eye, is used to obtain the retinal fundus image. The fundus region's structural features include the fovea, macula, optic disc, and optic cup. It is a noninvasive method that has been shown to be a useful instrument for assessing a person's ocular health. The optic disc shows up as a bright yellowish area in colored fundus imaging, and it may be further split into two components, the optic cup (inner portion) and the neuroretinal rim (outer boundary). The optic nerve cup or the enlargement of the optic cup may be used to detect glaucoma in an eye. The cup-to-disc ratio (CDR), which is the ratio of the diameter of the optic cup to the diameter of the optic disc, is the best diagnostic for glaucoma identification. According to physicians, an eye with a CDR value of 0.65 or higher is classified as having glaucoma (Juneja et al. 2020a). However, the procedures used by the doctors are exceedingly laborious, time-consuming, and ineffective since they require manually separating the discs and cups from individual photographs. An experienced grader can measure, record, and ultimately determine whether a patient has glaucoma in roughly 8 min per eye on average. Additionally, glaucoma is also diagnosed by ophthalmologists using a variety of thorough retinal tests, including ophthalmoscopy, tonometry, perimetry, gonioscopy, and pachymetry. Ophthalmoscopy is the evaluation of the color and form of the optic

Table 1 List of extracted features

S. No.	Features	S. No.	Features	S. No.	Features
1	CDR (cup–disc ratio)	13	mat0_avg	25	DWT (discrete wavelet transform)
2	GLCM0 (grey-level co-occurrence matrix)	14	mat45_avg	26	FOS (First-Order Statistical)
3	GLCM45	15	mat90_avg	27	Cum2est
4	GLCM90	16	mat135_avg	28	Coherence
5	GLCM135	17	GLRM 0 (Grey Level Run Length Matrix)	29	Bicoherence
6	SRE (Short-Run Emphasis)	18	GLRM 45	30	Spectrum
7	LRE (Long-Run Emphasis)	19	GLRM 90	31	Energy
8	GLU (Grey-Level Uniformity)	20	GLRM 135	32	Homogeneity
9	RLU (Run Level Uniformity)	21	HOS (Higher-Order Spectral)	33	Correlation
10	RPC (Rational Polynomial Coefficient)	22	NRR (Neuro-Retinal Rim)	34	Contrast
11	mat0	23	DDLS (Disc Damage Likelihood Scale)	35	Dissimilarity
12	mat45	24	HOC (Higher Order Cumulant)	36	Entropy

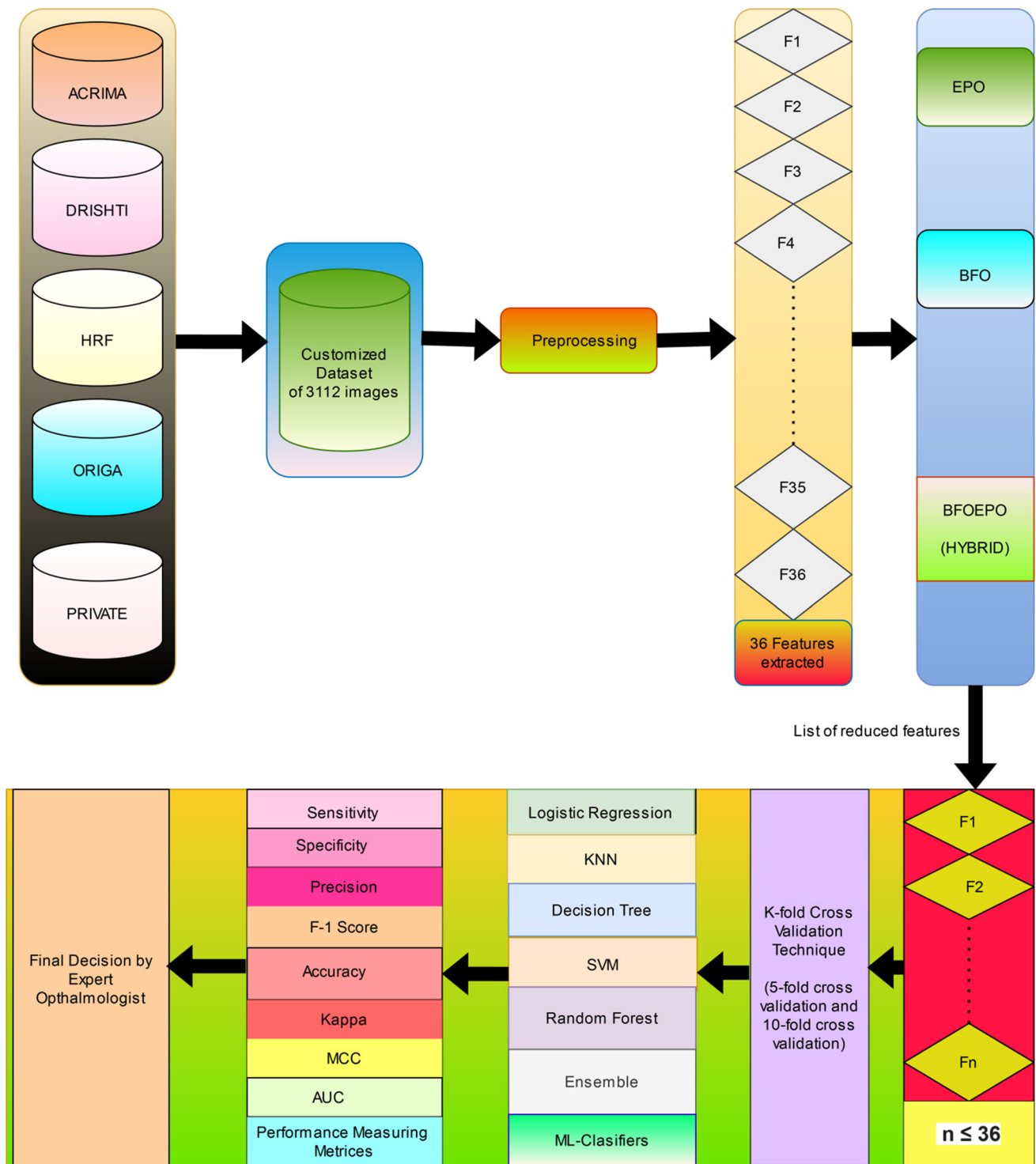


Fig. 2 Framework of the suggested study for glaucoma identification

nerve, while tonometry is the measurement of internal eye pressure. Perimetry is the examination of the visual field; gonioscopy is the measurement of the angle between the iris and cornea; and pachymetry is the measurement of corneal thickness. However, each of these methods necessitates human labor and takes time (Juneja et al.

2020b). All of these methods require human labor, take time, and might result in biased judgments from various experts (vulnerable to human error). In order to get beyond the restrictions of classical approaches, an automated glaucoma diagnostic system is required. Nowadays, computers are used in medical imaging as a noninvasive

Fig. 3 Layout of the preprocessing of the images

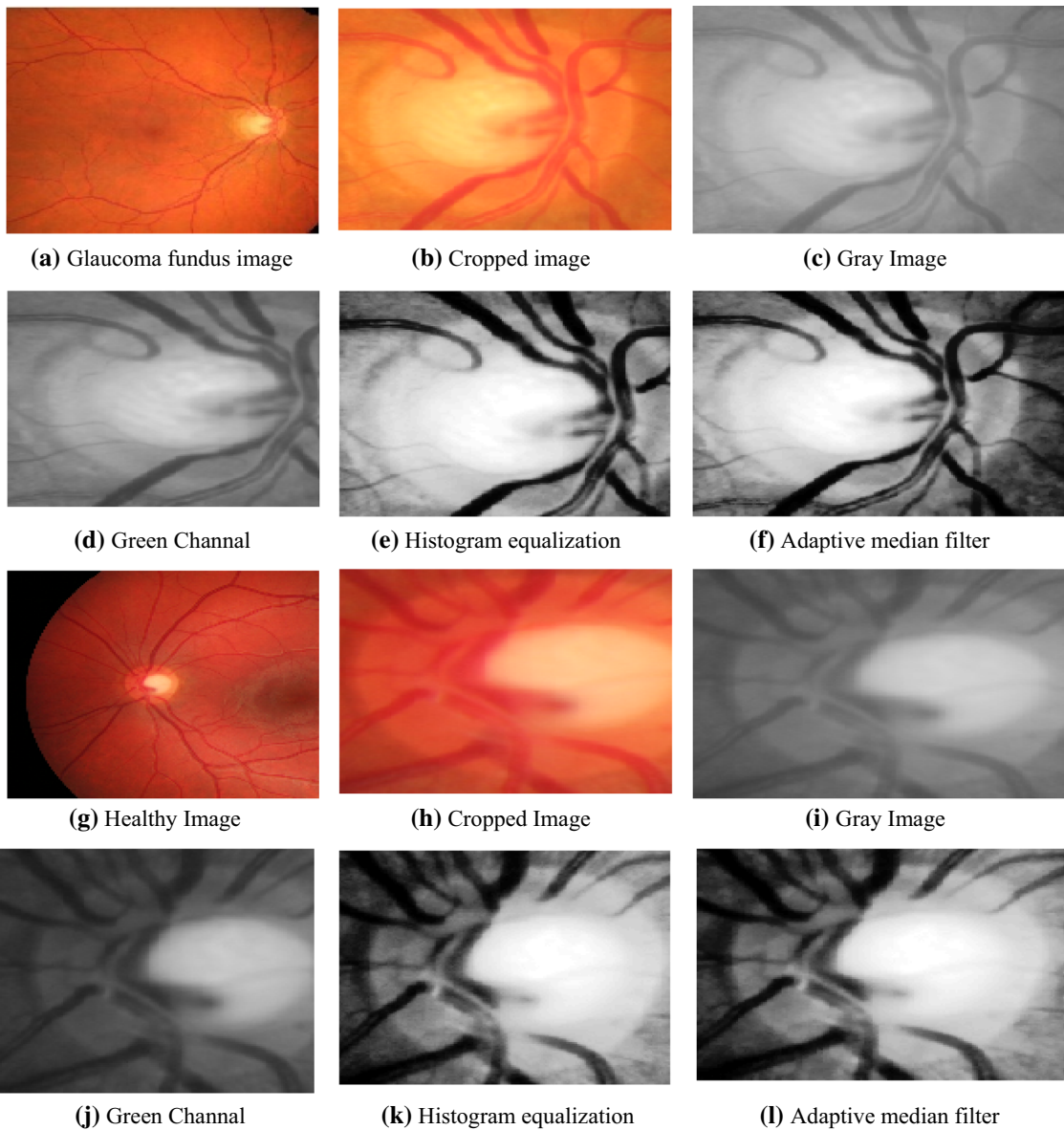
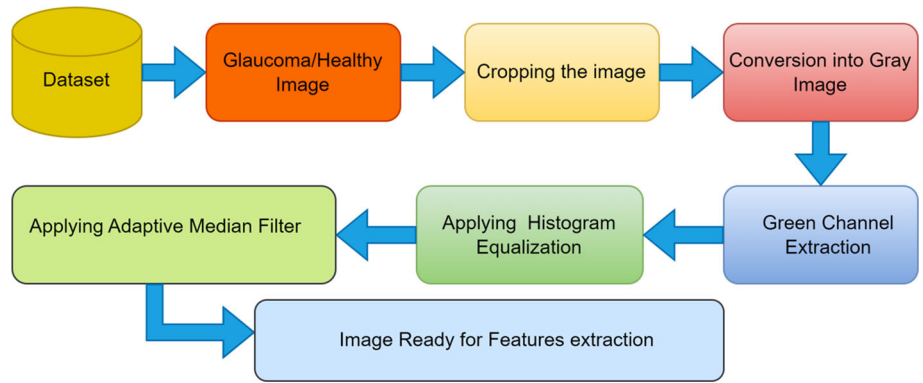


Fig. 4 a–f Preprocessing of Glaucoma image and g–l preprocessing of healthy image

imaging modality to predict abnormalities. Therefore, automated techniques based on computer-aided detection and diagnosis (CADE/CADx) systems are preferable for solving the identification issue. In order to reduce the computational complexity, a computer-aided diagnostic system may act as a supportive measure for the first screening of glaucoma for diagnosis purposes. Consequently, a computer-aided diagnostic (CAD) system that may serve as a second opinion for clinicians is required to save valuable time for ophthalmologists. It has the potential to reduce the likelihood of misclassification, reduce the burden on physicians, and improve inter- and intra-observability. CAD systems can use images of the retinal fundus to predict glaucoma, pull out different kinds of features, and classify retinal images as “normal” or “abnormal.”

A number of individuals experience the retinal illness of glaucoma each year. Due to increasing intraocular pressure that damages the eye’s optic nerve, glaucoma is the most common ocular condition that causes permanent blindness. An experienced ophthalmologist would often examine the dilated pupil in the eye to make a diagnosis of glaucoma progression. However, this method is time-consuming and difficult; thus, the problem may be handled using automation that would speed up diagnosis by using the idea of CAD (in the limelight of machine learning). The only way to stop it from progressing into irreversible blindness is through early identification. Treatment is often delayed by the need for onerous manual input from ophthalmologists, such as disc and cup size. Therefore, it is crucial to create quicker and more precise glaucoma diagnostic techniques. Automation offers a more practical and efficient answer to the glaucoma detection issue.

Any classification system that wants to get rid of extraneous or redundant features must first pick its best features. The classifier’s ability to predict outcomes may be diminished by a feature collection that contains a significant number of irrelevant or undesirable features. Because of this, the best feature selection is crucial. This lowers the cost of computation and increases the classifier’s predictive

ability. The most crucial phase of glaucoma detection system design is feature selection. The primary goal of the suggested technique is to shrink the feature space in order to boost the classification system’s performance. EPO, BFO, and their hybrid are the three algorithms used and employed in this study for the problem at hand. To the best of our knowledge, these algorithms have rarely been used in this way in the detection of any human disease. Six machine-learning classifiers that have been nominated for evaluation are used to assess the characteristics chosen by these three methods. Over fundus images from a different benchmark datasets, the suggested methodology’s effectiveness is evaluated. There have been 24 tests in all. In many cases, the feature space is reduced to less than 50%. This reduction may be as high as 86.111% (5 features are returned from the original 36 features without significantly compromising accuracy).

To the best of our knowledge, there is presently no report on glaucoma identification using EPO, BFO, and their hybrid. One of the objectives of our study is to close this gap. A hybrid algorithm that combines BFO and EPO for FS and classification in fundus images is implemented in order to strike a balance between global and local search, as well as between exploration and exploitation.

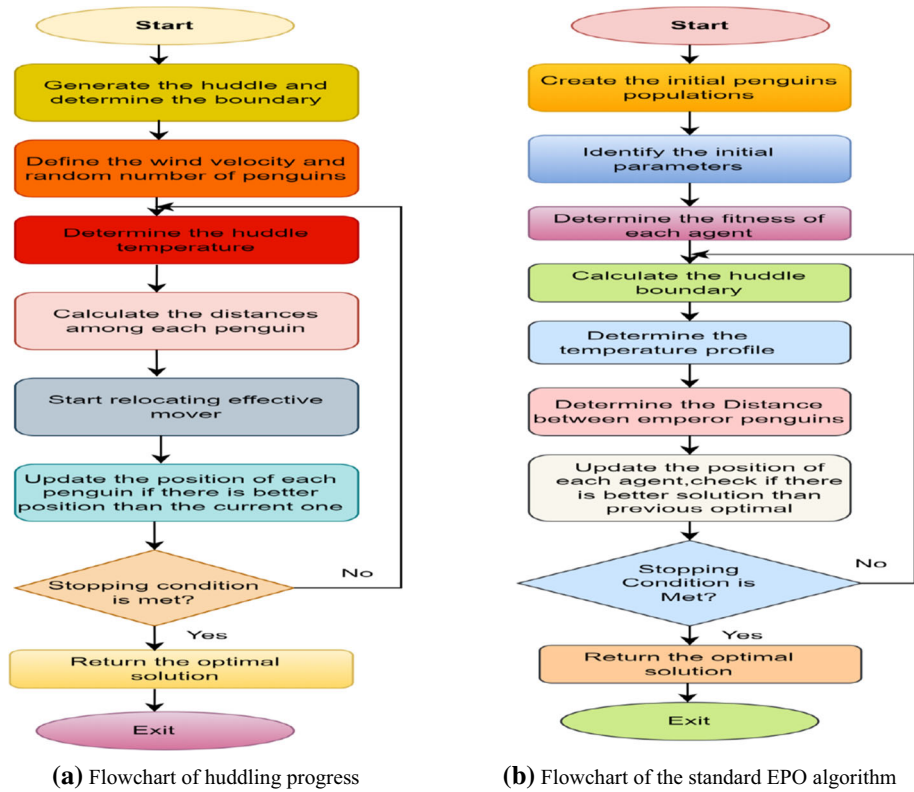
Our presented empirical study attempts to address all of these issues, beginning with the selection of the best, smallest, and most efficient features needed for the confirmation of this human eye-related, widespread disease, glaucoma, as a subject and ending with incredible performance and results. The primary contribution of this work may be summed up as follows:

- Through an extensive literature survey, it is observed that there is a lot of room for feature optimization for glaucoma identification. Taking into account the benefits of BFO and EPO, a BFOEPO hybrid FS strategy is proposed. This strategy could get rid of redundant and unnecessary features from the feature space, improving classification accuracy and lowering the cost of computing. Finally, we get the best subset of attributes

Table 2 List of parameters used in BFO algorithm

Parameter	Description	Parameter	Description
Q	Bacterial number	\aleph_d^j	The d th component of the \aleph^j
\aleph	A bacterium on the optimization domain	$\aleph^i(k, l, m)$	The j th bacterium in the k th chemotaxis l th elimination-dispersal procedure
\aleph^j	The j th bacterium position	$\Delta(j)$	A random value on $[-1, 1]$ for the j th bacterium
r_{attract} and x_{attract}	Coefficients of attraction–repulsion	$h1_{\text{repellant}}$ and $x_{\text{repellant}}$	Coefficients of repulsion–repulsion
$Bcc(\aleph, \aleph^i(k, l, m))$	Cell-to-cell communication value	Q_r	Number of population members, have sufficient nutrients

Fig. 5 a Flowchart of huddling progress **b** Flowchart of the standard EPO algorithm



(features) from the combination of different standard datasets. To the best of our knowledge, we are the frontrunners in applying these algorithms for glaucoma identification, thus filling a large research gap.

- When compared to recent state-of-the-art studies discussed in Table 17, our customized dataset is also one of the largest. Extensive experimentation has been conducted, which involves the implementation of 24 tests. In addition to demonstrating the applicability of the proposed method, a comprehensive study of a number of parameters is presented. Along with ROC

curves, eight efficiency measurement parameters of six machine learning classifiers have been computed to indicate their efficient performance. This fivefold and tenfold approach implementation effort also shows the time taken by different processes.

- We have also shown the extended results in the form of three tables depicting the best values generated for eight efficiency measurement parameters (using six machine learning classifiers) when the minimum number of features was selected (in the case of all three algorithms). This type of table is rarely observed in previous state-of-the-art studies.
- Through this study, we provide the best features to researchers; an efficient and rapid support system for ophthalmologists (on which they can rely); and a software-based tool for the human race to slow down human eye sight loss by ensuring early, efficient, and effective identification of this infection. The tool can be tweaked to work with mobile and wearable medical equipment, and it can be used in places where there aren't enough experienced medical practitioners.

This research study is organized as follows: The review of the literature is presented in Sect. 2. Section 3 describes the datasets and algorithms, and Sect. 4 describes the results. Section 5 is dedicated to analysis of our work along with findings and comparison with recent state-of-the-art studies. Limitations and future directions are suggested in

Table 3 Experiment settings

Size of the population	5, 10, 15, 20
No. of features	36
No. of samples	3112
Upper Bound	1
Lower Bound	0
Times of elimination	2
Times of reproduction	4
Maximum length of swim	5
Rate of elimination	0.25
Mutation rate	0.08
Selection method	Essential selection
Crossover type	One-site crossover

Table 4 Basic definition of the terminologies used

Term/Notation	Basic definition
True positive	A true positive is an outcome where the model correctly predicts the positive class
True negative	A true negative is an outcome where the model correctly predicts the negative class
False positive	A false positive is an outcome where the model incorrectly predicts the positive class
False negative	A false negative is an outcome where the model incorrectly predicts the negative class
Sensitivity(Recall)	Sensitivity in Machine Learning can be described as the metric used for evaluating a model's ability to predict the true positives of each available category
Specificity	Specificity itself can be described as the algorithm/model's ability to predict a true negative of each category available
Accuracy	Accuracy is actually the proportion of true results
Precision	It is defined as the ratio of correctly classified positive samples (True Positive) to a total number of classified positive samples
F1-score	The F1 score is defined as the harmonic mean of precision and recall
Matthews correlation coefficient (MCC)	The Matthews correlation coefficient (MCC) is a statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives)
Kappa score	Cohen's kappa is a metric often used to assess the agreement between two raters
Logistic Regression(LR)	Using a given collection of independent factors, it is used to predict the categorical dependent variable. It forecasts the value of a categorical dependent variable. It can be either Yes or No, 0 or 1, true or False, etc., and however, probability values between 0 and 1 are shown
Random Forest(RF)	It constructs decision trees from data samples, obtains the forecast from each, and then picks the optimal option by voting. It is an ensemble approach that is superior than a single decision tree since averaging the results reduces over-fitting
Decision Tree (DT)	It displays the predictions generated by a sequence of feature-based splits using a flowchart resembling a tree structure. It begins with a node at the root and concludes with a choice made by leaves
K-Nearest Neighbor (KNN)	It utilizes multiple-class data to predict the categorization of a new sample point. It is nonparametric because it makes no assumptions about the analyzed data; the model is derived from the data
Support vector machine (SVM)	A SVM model is essentially a representation of several classes in a hyperplane of multidimensional space. SVM will create the hyperplane in an iterative way so that the error may be reduced. The objective of SVM is to partition datasets into classes in order to identify the maximum marginal hyperplane (MMH)
Ensemble	It is a machine learning approach that integrates many base models into one best prediction model. Instead of creating a single model and hope it is the most accurate prediction possible, ensemble approaches take into account a multitude of models and average them to get a final model

Sect. 6. Finally, the conclusion of the paper is illustrated in Sect. 7.

2 Prior studies

The prime goal of FS is to improve the accuracy of a classifier by minimizing the irrelevant information. Da Silva et al. (2011) transformed the many-objective fitness function into single-objective fitness by integrating classification accuracy and feature quantity. Yang and Honavar (1998) introduced a single target fitness function in that maximizes accuracy and lowers costs. Winkler et al. (2011) have also used several fitness functions to improve classification accuracy. By using the best characteristics, the data categorization problem may be solved with the greatest amount of accuracy (Pandey and Kulhari 2018). Since there are 2^n alternative feature subsets for every n

attributes, choosing a meaningful feature subset from the dataset is a difficult job (Emary et al. 2016). Consequently, it might be challenging to identify the optimum feature subset. It will get much harder as more features are introduced. It is sometimes hard to locate the ideal set of traits via exhaustive or brute-force search (Xue et al. 2015). Heuristic search, complete search, random search, and metaheuristic-based approaches are some alternative FS methods (Dash and Liu 1997; Liu et al. 2014). Admired metaheuristic methods like genetic algorithm (GA), differential evolution (DE), particle swarm optimization (PSO), ant colony optimization (ACO), and cuckoo search (CS) are often employed to find the best feature subset. Chen and Hsiao (2008) coupled actual GA with a support vector machine (SVM) classifier to choose the pertinent features. Derrac et al. (2009) used three distinct populations to introduce a GA-based cooperative co-evolutionary method for FS. The focus has been on FS and instance

Table 5 Feature selection on population size 5, 10, 15 and 20 for k-fold cross-validation

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution Time
<i>Experiment 1: Fitness value taken on different iteration interval for population size 5(fivefold)</i>						
36	100	5	5	12, 18, 27, 31, 33	0.894	0.574
36	200	5	4	22, 31, 32, 36	1.014	1.108
36	300	5	4	1, 23, 24, 36	1.167	2.073
36	400	5	4	10, 15, 16, 25	1.267	2.506
36	500	5	4	3, 4, 7, 29, 33	0.979	3.184
<i>Experiment 2: Fitness value taken on different iteration interval for population size 5(tenfold)</i>						
36	100	5	4	4, 11, 15, 29	1.026	0.309
36	200	5	5	6, 11, 16, 32, 33	1.111	1.414
36	300	5	4	4, 5, 18, 25	1.109	1.744
36	400	5	5	9, 16, 20, 26, 36	1.127	2.570
36	500	5	5	3, 10, 13, 23, 28	0.971	3.479
<i>Experiment 3: Fitness value taken on different iteration intervals for population size 10(fivefold)</i>						
36	100	10	10	4,10,14,15,18,19,24,28,33,36	0.009	0.542
36	200	10	9	1,5,8,11,13,19,21,22,25	0.009	2.446
36	300	10	9	7,10,11,13,18,21,25,26,31	0.010	3.026
36	400	10	10	1,2,12,17,18,19,22,31,32,33	0.007	4.539
36	500	10	9	4,7,10,11,13,18,20,25,34	0.009	5.718
<i>Experiment 4: Fitness value taken on different iteration interval for population size 10(tenfold)</i>						
36	100	10	8	4,7,8,14,17,20,25,28	0.009	1.247
36	200	10	10	6,7,8,13,14,15,17,21,25,35	0.009	1.095
36	300	10	8	2,8,15,21,22,27,30,35	0.008	3.593
36	400	10	9	6,7,9,12,15,17,20,24,29	0.010	4.540
36	500	10	9	2,6,7,13,27,33,34,35,36	0.009	5.055
<i>Experiment 5: Fitness value taken on different iteration interval for population size 15(fivefold)</i>						
36	100	15	14	1,5,7,11,12,13,14,16,17,21,26,31,33,34	0.008	1.487
36	200	15	13	1,3,4,7,8,9,12,17,20,24,27,31,32	0.007	2.948
36	300	15	15	5,9,10,12,14,15,16,22,24,27,29,30,32,34,36	0.009	4.280
36	400	15	15	1,5,7,9,11,13,17,19,21,22,24,26,29,32,34	0.008	8.006
36	500	15	12	3,9,10,11,14,20,22,23,27,28,30,31	0.008	4.914
<i>Experiment 6: Fitness value taken on different iteration intervals for population size 15(tenfold)</i>						
36	100	15	12	3,4,5,8,11,12,17,23,27,28,33,35	0.007	1.074
36	200	15	14	2,8,9,12,19,21,23,24,26,27,30,31,34,35	0.008	3.054
36	300	15	13	1,2,4,6,11,21,24,25,26,27,32,34,36	0.009	4.861
36	400	15	13	3,7,12,14,17,18,20,28,29,30,32,35,36	0.008	5.662
36	500	15	13	3,5,12,13,16,20,22,25,29,31,32,34,35	0.008	6.150
<i>Experiment 7: Fitness value taken on different iteration intervals for population size 20(fivefold)</i>						
36	100	20	15	3,4,6,10,11,12,15,16,18,20,22,24,28,32,36	0.008	1.026
36	200	20	12	2,4,6,8,10,15,16,20,21,27,28,30	0.008	4.465
36	300	20	15	2,4,8,9,10,11,14,15,18,19,25,27,30,35,36	0.007	5.183
36	400	20	14	1,3,4,7,9,10,13,19,21,26,28,30,31,34	0.009	7.652
36	500	20	15	3,4,5,9,10,17,18,19,21,22,26,27,28,33,34	0.007	7.487
<i>Experiment 8: Fitness value taken on different iteration intervals for population size 20(tenfold)</i>						
36	100	20	14	4, 12, 13, 16, 19, 21, 22, 24, 27, 29, 30, 31, 34, 35	0.893	2.235
36	200	20	17	1, 2, 6, 7, 8, 9, 12, 16, 19, 20, 21, 24, 26, 32, 33, 35, 36	1.002	2.462
36	300	20	14	1, 5, 11, 13, 16, 18, 19, 21, 22, 27, 28, 29, 33, 36	0.906	6.715
36	400	20	14	2, 5, 6, 9, 12, 14, 15, 17, 24, 25, 26, 28, 34, 36	1.015	9.227

Table 5 (continued)

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution Time
36	500	20	13	3, 7, 8, 11, 16, 19, 20, 22, 23, 27, 30, 34, 36	0.830	11.230

utilizing a single process. As a result, it shortens computation time. Feng and Li (2008) used chaos theory and evolutionary algorithms to provide highly accurate and speedy patient information retrieval. Wu et al. (2018) produced an outcome for the grading of gliomas by using a semiautomatic segmentation technique. Additionally, genetic programming (GP) has been used to do FS (Muni et al. 2006). The majority of GP techniques use a tree-based representation where the properties of the leaf nodes are selected. A wrapper-based general purpose strategy that uses the naive Bayes algorithm for classification has been published by Neshatian and Zhang (2009). Additionally, PSO has been used for FS (continuous and binary). In binary PSO, bit strings 1 and 0 signify the selection and non-selection of respective features, respectively. The continuous PSO uses a threshold to determine whether or not to choose a certain feature (attribute). Lane et al. (2013) coupled the advantages of PSO with statistical clustering for FS. Later, they (Lane et al. 2013) enhanced the performance of the preexisting approach by allowing the selection of several features from the same group. Ke et al. (2008) used constrained pheromone values to choose relevant parameters while using ACO. O'Boyle et al. (2008) developed an ACO-based FS method and enhanced the SVM's parametric values to identify the best feature subset. The chance that a certain feature would be picked or not has also been determined using a weighted method. Khushaba et al. (2008) created a hybrid FS method where DE selects the ideal feature subset from the output of ACO. A combined ACO approach with two colonies for attribute selection has been given by Vieira et al. (2010), in which the first colony estimates the requisite number of features and the second colony selects particular features. Ke et al. (2010) proposed many-objective ACO to hasten the convergence of filter FS. Additionally, an adaptive DE was used for FS, with parameters that self-adapted to the difficulties (Ghosh et al. 2013). Rodrigues et al. (2014) introduced a novel FS method that is based on the binary version of the bat algorithm and the optimum-path forest. Using a memetic wrapper and Relief-F algorithm, a two-stage feature selection technique has graded each feature (Yang and Deb 2009). Furthermore, a memetic wrapper is used to choose the ideal combination of traits. Gu et al. (2018) selected the important features and dimensionality reduction with competitive swarm optimizer (CSO), a relatively recent PSO variant. A binary black hole

algorithm (BBHA) was employed to solve the FS in biological data (Pashaei and Aydin 2017). Additionally, Emary et al. (2016) developed a binary variant of the grey wolf optimization (GWO) to increase the efficacy of the classifier. Mafarja and Mirjalili (2017) suggested a hybrid meta-heuristic FS strategy based on the whale optimization algorithm (WOA) and simulated annealing (SA), where SA method is implemented in WOA, in order to improve the exploitation process. Tang et al. (2016) devised two FS selection strategies, namely maximum discrimination (MD) and MD-2 techniques using Jerefs multi-hypothesis (JMH) divergence, to prevent the early convergence of BPSO-SVM FS. Kang et al. (2016) suggested an outlier-insensitive hybrid approach that both selects the suitable feature subset and recognizes outliers in data-driven diagnostics. Barani et al. (2017) used a binary-inspired gravitational search algorithm (bGSA) to solve the FS objective. Mafarja et al. (2018) used binary dragonfly optimization to choose the ideal subset of features using time-varying transfer functions. To address the problem of FS, a novel chaotic salp swarm approach has also been described (Sayed et al. 2018). Shunmugapriya and Kanmani (2017) created a hybrid FS approach based on ACO and bee colony optimization (BCO) to stop ant stagnation behavior and choose the most excellent features subset. Jayaraman and Sultana (2019) created a unique hybrid metaheuristic strategy based on the efficiency of artificial gravitational cuckoo search and particle Bee optimized associative memory neural network in order to solve the FS dilemma. Ibrahim et al. (2019) developed a novel hybrid metaheuristic method named SSAPSO based on the slap swarm algorithm (SSA) and particle swarm optimization to extract the best subset of features from a high-dimensional dataset (PSO). Prabukumar et al. (2019) presented an excellent diagnostic model based on cuckoo search characteristics for the early detection of lung cancer. Sayed et al. (2019) introduced the chaotic crow search algorithm (CCSA), a novel metaheuristic optimization method, to extract the best features from a dataset. Mafarja et al. (2019) presented on a binary version of the grasshopper optimization approach based on sigmoid and V-shaped transfer functions for FS. Nematzadeh et al.'s (2019) frequency-based selection technique also uses mutual congestion to identify signals. The binary cuckoo search approach was developed by Rodrigues et al. (2013) to choose the ideal set of attributes. The Lévy flight is used in

Table 6 Performance of different five classifiers including ensemble on population size 5,10,15,20 (k-fold cross validation approach)

Classifier	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution Time	Two tailed test (p-value)
<i>Experiment 1: Using fivefold cross validation measure the classifier performance on 5 selected features of minimum cost 0.89409 (population size 5)</i>										
Logistic Regression (LR)	0.970	0.652	0.927	0.766	0.665	0.687	0.811	0.8545	0.282	0.070141
Random Forest(RF)	0.860	0.782	0.763	0.773	0.639	0.639	0.822	0.8319	9.134	0.068455
Decision Tree(DT)	0.931	0.748	0.862	0.801	0.699	0.703	0.839	0.8644	0.272	0.065915
K-nearest neighbor (KNN)	0.912	0.722	0.826	0.770	0.652	0.655	0.817	0.8429	0.406	0.069217
Support vector machine (SVM)	0.950	0.614	0.877	0.723	0.603	0.624	0.802	0.8279	0.741	0.071545
Ensemble	0.948	0.727	0.891	0.801	0.703	0.711	0.838	0.8679	0.483	0.066062
<i>Experiment 2: Using tenfold cross-validation measure the classifier performance on 5 selected features of minimum cost 0.97123 (population size 5)</i>										
LR	0.978	0.698	0.950	0.805	0.718	0.736	0.836	0.8765	0.372	0.066358
RF	0.941	0.805	0.888	0.844	0.761	0.763	0.873	0.8916	20.086	0.061068
DT	0.922	0.792	0.854	0.822	0.725	0.727	0.857	0.8748	0.357	0.063311
KNN	0.962	0.782	0.923	0.847	0.770	0.776	0.872	0.8968	0.431	0.061207
SVM	0.942	0.793	0.888	0.838	0.838	0.756	0.868	0.8881	1.048	0.061762
Ensemble	0.924	0.846	0.865	0.855	0.774	0.774	0.885	0.8957	0.540	0.05943
<i>Experiment 3: Using fivefold cross-validation measure the classifier performance on 10 selected features of minimum cost 0.00798 (population size 10)</i>										
LR	0.968	0.637	0.922	0.753	0.648	0.671	0.803	0.8476	0.805	0.071388
RF	0.883	0.757	0.788	0.772	0.646	0.646	0.820	0.8371	10.737	0.068759
DT	0.869	0.637	0.737	0.703	0.521	0.524	0.813	0.8044	0.278	0.069832
KNN	0.903	0.591	0.769	0.698	0.512	0.545	0.802	0.8080	0.348	0.071545
SVM	0.924	0.602	0.820	0.694	0.557	0.572	0.823	0.8064	0.702	0.068303
<i>Experiment 4: Using tenfold cross-validation measure the classifier performance on 8 selected features of minimum cost 0.00855 (population size 10)</i>										
LR	0.959	0.711	0.910	0.798	0.704	0.715	0.835	0.8690	0.567	0.066506
RF	0.950	0.849	0.908	0.877	0.811	0.812	0.900	0.9136	23.358	0.057433
DT	0.946	0.820	0.897	0.857	0.781	0.783	0.883	0.9003	0.357	0.059710
KNN	0.957	0.778	0.914	0.840	0.760	0.765	0.868	0.8922	0.416	0.061762
SVM	0.947	0.778	0.895	0.832	0.746	0.751	0.863	0.8858	1.033	0.062462
Ensemble	0.905	0.730	0.815	0.770	0.690	0.652	0.817	0.8412	0.542	0.069217
<i>Experiment 5: Using fivefold cross-validation measure the classifier performance on 13 selected features of minimum cost 0.0079 (population size 15)</i>										
LR	0.949	0.719	0.891	0.796	0.697	0.706	0.834	0.8655	0.432	0.066654
RF	0.833	0.820	0.740	0.778	0.639	0.642	0.827	0.8290	10.006	0.067701
DT	0.848	0.754	0.741	0.747	0.600	0.600	0.801	0.8140	0.407	0.071703
KNN	0.933	0.765	0.868	0.813	0.716	0.720	0.849	0.8719	0.359	0.064458
SVM	0.926	0.763	0.857	0.808	0.707	0.709	0.845	0.8673	0.629	0.065038
Ensemble	0.871	0.671	0.750	0.709	0.555	0.557	0.802	0.8083	0.403	0.071545
<i>Experiment 6: Using tenfold cross validation measure, the classifier performance on 12 selected features of minimum cost 0.00792 (population size 15)</i>										
LR	0.945	0.738	0.885	0.805	0.708	0.715	0.841	0.8696	2.123	0.065622
RF	0.938	0.843	0.888	0.865	0.790	0.791	0.890	0.9038	22.762	0.058758
DT	0.920	0.740	0.842	0.788	0.678	0.681	0.830	0.8545	0.427	0.06725
KNN	0.953	0.778	0.905	0.837	0.754	0.759	0.865	0.8893	0.775	0.062181
SVM	0.929	0.767	0.862	0.812	0.713	0.716	0.848	0.8702	1.141	0.064608
Ensemble	0.899	0.814	0.823	0.819	0.715	0.715	0.857	0.8684	0.580	0.063311

Table 6 (continued)

Classifier	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution Time	Two tailed test (p-value)
<i>Experiment 7: Using fivefold cross-validation measure, the classifier performance on 15 selected features of minimum cost 0.00715 (population size 20)</i>										
LR	0.939	0.733	0.875	0.798	0.697	0.703	0.836	0.8644	0.523	0.066358
RF	0.858	0.828	0.771	0.799	0.676	0.677	0.843	0.8476	10.588	0.065329
DT	0.789	0.827	0.693	0.754	0.592	0.598	0.808	0.8030	0.308	0.070607
KNN	0.898	0.705	0.800	0.749	0.619	0.622	0.801	0.8279	0.395	0.071703
SVM	0.909	0.765	0.829	0.796	0.686	0.687	0.837	0.8568	0.697	0.06621
Ensemble	0.865	0.706	0.723	0.598	0.502	0.501	0.805	0.8045	0.416	0.071075
<i>Experiment 8: Using tenfold cross validation measure, the classifier performance on 13 selected features of minimum cost 0.8308 (population size 20)</i>										
LR	0.968	0.817	0.938	0.873	0.808	0.813	0.893	0.9136	0.938	0.058358
RF	0.948	0.833	0.903	0.867	0.795	0.797	0.891	0.9067	27.062	0.058624
DT	0.953	0.860	0.914	0.886	0.824	0.825	0.906	0.9194	0.5300	0.05665
KNN	0.960	0.809	0.922	0.862	0.791	0.794	0.885	0.9055	0.5715	0.05943
SVM	0.957	0.841	0.920	0.879	0.814	0.816	0.899	0.9154	1.321	0.057564
Ensemble	0.910	0.885	0.850	0.868	0.789	0.789	0.898	0.9015	0.801	0.057696

the binary cuckoo search method to explore the whole search space. Lévy flight produces a random walk and follows the heavy tailed probability distribution. As a consequence, more iterations result in larger step sizes, which affect the convergence accuracy. All metaheuristic-based FS strategies often suffer from stability issues since different sets of features are selected in different runs. It was shown that conventional and metaheuristic-based FS methods suffer from stability issues and are computationally expensive. The metaheuristic FS methods show premature convergence due to a lack of variance.

The extraction of wavelet features was implemented in this research (Singh et al. 2016), and then genetic feature optimization, many learning methods, and different parameter settings were included. This research uses feature extraction from the segmented, blood vessel-free optic disc to increase identification accuracy. They justified that wavelet characteristics of the segmented optic disc image are clinically more relevant in comparison with features of the full or sub-fundus image. In this article (Khan et al. 2022), authors suggested method takes advantage of image denoising of digital fundus pictures by minimize the statistics of wavelet coefficients of glaucoma images using a non-Gaussian bivariate probability distribution function. The least square support vector machine classifier, which uses a variety of kernel functions, is then fed the chosen features. This study makes use of nonparametric GIST descriptor and optic disc (Raghavendra et al. 2018). After revolutionary area-based optic disc segmentation, the Radon transformation was proposed in the approach (RT). Modified census transformation was used to account for

changes in the light levels of the Radon-converted picture (MCT). The spatial envelope energy spectrum was then extracted from the MCT pictures using the GIST descriptor. Utilizing locality-sensitive discriminant analysis (LSDA), the resultant GIST descriptor dimension was lowered, and then, different FS and ranking algorithms are used. In this study (Maheshwari et al. 2017), practitioners provide an innovative technique for an automated glaucoma diagnosis using digital fundus pictures. The iterative variational mode decomposition (VMD) approach is used for picture decomposition. From VMD components, a number of properties were retrieved, including fractal dimensions, Yager entropy, Renyi entropy, and Kapoor entropy. The discriminating features were chosen using the ReliefF method, and the least square-SVM uses these features to classify the data. The approach described in the study uses information from higher-order spectra (HOS), trace transform (TT), and discrete wavelet transform (DWT) (Krishnan and Faust 2013). In this study, the SVM classifier with a polynomial order 2 kernel function performed well in differentiating between glaucoma and healthy pictures.

The automated glaucoma diagnosis approach presented in this work uses quasi-bivariate variational mode decomposition (QB-VMD) (Agrawal et al. 2019). This approach is employed to deconstruct 505 fundus pictures altogether, yielding band-limited sub-band images (SBIs) centered on a certain frequency. These SBIs are fault-free and have no issues with mode mixing. The most helpful elements that effectively acquired the necessary data are what determine how accurately glaucoma is detected. QB-VMD SBIs are

used to extract 70 features. The ReliefF approach selects the extracted features. The dimensionality of selected characteristics is then decreased by feeding them through singular value decomposition. The least square SVM classifier is then used to classify the reduced features. Glaucoma detection is performed using time-invariant feature cup-to-disc ratio and anisotropic dual-tree complex wavelet transform features (Kausu et al. 2018). Fuzzy C-Means clustering is employed for optic disc segmentation, while Otsu's thresholding is used for optic cup segmentation. The extraction of GIST and PHOG characteristics from preprocessed fundus pictures was the main emphasis of this paper (Gour and Khanna 2020). To choose relevant characteristics, the extracted features are sorted and chosen using principal component analysis (PCA).

Super-pixel sorting for glaucoma screening has been done using center-surround data and histograms (Cheng et al. 2013). Fractal dimensions (FDs) and power spectral characteristics were combined with a SVM by Kolar and Jan (2008). Pixel intensities and the fast Fourier transform (FFT) were employed by Nyul (2009). Raja and Gan-gatharan (2013) employed complicated wavelet transform and higher-order spectra. They employed wavelet packet decomposition (WPD) together with entropy and energy geographies as a feature. The DWT and histogram functions have been proposed by Kirar and Agrawal (2018) for this infection identification. 2D-DWT has been suggested by Kirar and Agrawal (2019) to separate glaucoma from healthy pictures, and classification has been carried out using histogram-based features. Glaucoma fundus pictures were employed by Yadav et al. (2014) for texture-based feature extraction and categorization. Maheshwari et al. (2016) used empirical wavelet transform (EWT) to deconstruct fundus pictures. The least square-SVM has been shown to be useful for applying on fundus pictures in two categories in the newly published research on glaucoma diagnosis using fundus images (Martins et al. 2020; Parashar and Agrawal 2020). The dimensions of the optic disc and optic cup were calculated and used to estimate the CDR (cup-to-disc ratio) using fundus images in this study (Shanmugam et al. 2021). The initial step was to segment the optic disc (OD) and optic cup (OC). The second goal of the study was to reduce the number of features and errors, which are two conflicting objectives. Phases such as contrast amplification, picture collection, feature extraction, and glaucoma assessment are included in the suggested identification system. During the feature extraction phase, the OD and OC borders were segmented. In order to evaluate glaucoma in the images, the CDR ratio of an overused image is then determined. The glaucomatous photos have then been categorized using a random forest classifier according to the CDR values. Different versions

Table 7 The best values of the performance metrics when the minimum number of features was selected (in case of EPO)

Case of minimum features extracted	List of features	Max Specificity (Classifier)	Max Sensitivity (Classifier)	Max Precision (Classifier)	Max F1-score (Classifier)	Max Kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max Accuracy (Classifier)	Minimum Execution Time
5	12, 18, 27, 31, 33	0.970(LR)	0.782(RF)	0.927(LR)	0.801(DT)	0.703 (Ensemble)	0.711 (Ensemble)	0.839(DT)	0.86790 (Ensemble)	0.272 (DT)

Table 8 The maximum values of the performance metrics (in case of EPO algorithm)

Max specificity (Classifier)	Max sensitivity (Classifier)	Max precision (Classifier)	Max F1-score (Classifier)	Max Kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max accuracy (Classifier)	Minimum execution time
0.978(LR)	0.885(Ensemble)	0.950(LR)	0.886(DT)	0.824(DT)	0.816(SVM)	0.906(DT)	0.9194(DT)	0.272(DT)

Overall summary of feature selection using EPO

This table depicts the best results generated on population size 5, 10, 15 and 20 for K-fold approach (5- and 10-)

of U-Net that have been tweaked have been used to assess the performance of the suggested technique. Deformable U-Net, Full-Deformable U-Net, and Original U-Net have all been used to assess the performance of the suggested approach. In this study (Mrad et al. 2022), the authors reported an automated technique for Smartphone Captured Fundus Images (SCFIs) glaucoma screening. The installation of an optical lens for retinal image capture results in a mobile-assisted glaucoma screening system. The central idea was based on vessel displacement inside the optic disk (OD), where the vessel tree is adequately represented using SCFIs. Within this aim, the most significant contribution consists of proposing: (1) a robust method for finding vessel centroids in order to accurately model vessel distribution and (2) a feature vector that reflects two prominent glaucoma biomarkers in terms of vessel displacement. In this work (Deperlioglu et al. 2022), explainable artificial intelligence (XAI) was used to provide a hybrid solution with image processing and deep learning for Glaucoma screening. Enhancing colored fundus imaging data were performed using histogram equalization (HE) and contrast-limited adaptive HE (CLAHE). An explainable CNN utilized the augmented image data for the diagnosis. The XAI was accomplished by Class Activation Mapping (CAM), which enabled heat map-based explanations for the CNN's picture interpretation. Using three public retinal image datasets, the performance of the hybrid approach was examined. The optic cup (OC) borders in retinal fundus pictures are not very clear (Haider et al. 2022); as a result, correct segmentation of the OC is extremely difficult, and the OD segmentation performance must also be enhanced. For precise pixel-wise segmentation of the OC and OD, researchers suggested new networks: separable linked segmentation network (SLS-Net) and separable linked segmentation residual network (SLSR-Net). In SLS-Net and SLSR-Net, it is possible to preserve a big final feature map, which improves the performance of OC and OD segmentation by limiting spatial information loss. SLSR-Net uses external residual connections to enable features. Both suggested networks have a detachable convolutional connection to increase computational performance and decrease network cost. On four publicly accessible retinal

fundus image datasets, the segmentation abilities of the proposed networks were examined.

In this glaucoma detection approach, Elmoufidi et al. 2022, the Regions of Interest (ROI) were divided into components (BIMFs + residue) using the Bi-dimensional Empirical Mode Decomposition (BEMD) technique. To extract features from deconstructed BEMD components, VGG19 was applied. The same ROI features are combined into a bag of features, and principal component analysis (PCA) are used to reduce the feature dimension. To SVM, this condensed subset of characteristics is sent. ACRIMA and REFUGE were used for training, while RIM-ONE, ORIGA-light, Drishti-GS1, and sjchoi86-HRF were used for testing together with a portion of ACRIMA and REFUGE. An strategy based on the wrapper method employing soft-computing methods and a Kernel-Extreme Learning Machine (KELM) classifier were recommended for glaucoma diagnosis in this study by Balasubramanian and Ananthamoorthy (2022). The nature-inspired algorithms use a correlation-based feature selection (CFS) method to extract three feature sub-sets from the preprocessed fundus pictures. The salp-swarm optimization-based KELM, which determines the ideal parameters of the KELM classifier network, was trained using the chosen features. The objective of this study was to develop machine learning (ML) models with high predictive power and interpretability for glaucoma diagnosis based on multiple candidate features extracted from the examination retinal nerve fiber layer (RNFL) thickness and visual field (VF) (Kim et al. 2017). Using a process of feature evaluation, they then chose the optimal characteristics for classification (diagnosis). Authors considered C5.0, random forest (RF), support vector machine (SVM), and k-nearest neighbor as ML algorithms (KNN). Using multiple metrics, we analyzed the models' quality. They utilized 100 data cases as the test dataset and 399 data cases as the training and validation datasets. The random forest model performs the best. Clinicians may use the prediction findings to make more informed recommendations. This article examines the many problems involved with multi-objective optimization and the current efforts made to overcome these obstacles (Chand and Wagner 2015). It also describes how the current methodologies and body of knowledge have been used

Table 9 Feature selection on population size 5, 10, 15 and 20 for k-fold cross-validation ($k = 5$ and 10)

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution time
<i>Experiment 1: Fitness value taken on different iteration intervals for population size 5(fivefold)</i>						
36	100	5	7	1,2,8,10,14,27,32	0.01109	3.012
36	200	5	11	1,6,15,20,22,23,24,25,30,31,34	0.00988	2.914
36	300	5	10	6,9,12,18,23,31,32,34,35,36	0.00841	2.847
36	400	5	6	13,14,15,28,31,35	0.00842	2.723
36	500	5	10	4,5,6,14,16,18,23,24,29,35	0.00975	3.028
<i>Experiment 2: Fitness value taken on different iteration interval for population size 5(fivefold)</i>						
36	100	5	8	2,5,6,13,21,27,30,35	0.00935	2.950
36	200	5	10	2, 4, 5, 8, 10, 11, 13, 25, 34, 35	0.01033	2.995
36	300	5	6	3,14,17,18,24,27	0.00991	2.824
36	400	5	9	1, 8, 16, 17, 19, 29, 32, 33, 35	0.00999	2.999
36	500	5	6	10, 15, 20, 26, 28, 35	0.00994	3.062
<i>Experiment 3: For population size 10 and on the selected features, different machine learning classifiers using the fivefold cross-validation for training and testing</i>						
36	100	10	14	5,7,8,10,11,14,15,18,21,22,25,26,28,30	0.01000	16.063
36	200	10	14	6,7,10,12,13,19,26,28,29,30,31,33,34,35	0.00988	15.947
36	300	10	17	1,2,4,5,6,8,9,10,11,12,16,18,19,25,26,27,28	0.01071	15.946
36	400	10	16	2,4,5,6,10,12,17,18,19,21,27,31,33,34,35,36	0.00910	15.746
36	500	10	18	1,4,5,6,7,13,14,15,18,20,25,26,27,28,30,32,34,36	0.01080	16.143
<i>Experiment 4: For population size 10 and on the selected features different machine learning classifiers using the tenfold cross-validation for training and testing</i>						
36	100	10	17	2,5,9,10,12,14,17,19,23,27,30,31,32,33,34,35,36	0.00907	15.978
36	200	10	17	1,3,4,11,12,13,14,16,18,19,24,25,26,27,28,34,35	0.00991	15.875
36	300	10	17	2,3,4,6,7,9,12,13,18,19,20,21,24,25,31,32,33	0.00863	16.092
36	400	10	16	3,4,5,10,13,14,16,19,22,23,25,26,27,29,30,34	0.00929	15.955
36	500	10	13	2,4,5,7,8,11,14,23,25,27,28,29,32	0.00894	16.254
<i>Experiment 5: For population size 15 and on the selected features different machine learning classifier using the fivefold cross validation for training and testing</i>						
36	100	15	20	1,2,6,9,10,12,13,16,18,22,24,26,28,29,30,33,34,35,36	0.00956	31.692
36	200	15	20	3,7,8,10,11,12,13,14,17,18,21,22,25,26,28,29,30,31,34,35	0.01055	30.872
36	300	15	20	4,5,6,8,9,11,13,15,16,17,20,21,22,23,27,28,31,33,35,36	0.00954	31.443
36	400	15	23	1,2,3,4,5,6,7,8,12,15,18,19,20,21,22,26,28,29,30,33,34,35,36	0.00912	31.583
36	500	15	23	1,4,9,10,13,15,17,18,19,21,23,25,26,27,28,29,30,31,32,33,34,35,36	0.00857	31.008
<i>Experiment 6: For population size 15 and on the selected features different machine learning classifier using the tenfold cross-validation for training and testing</i>						
36	100	15	20	2,3,4,8,11,13,14,19,20,21,22,23,24,26,27,29,30,31,34,35	0.00853	31.265
36	200	15	19	3,4,5,7,8,13,16,17,18,19,20,25,30,31,32,33,34,35,36	0.01040	32.113
36	300	15	21	1,2,4,5,6,8,14,15,16,18,21,23,24,25,27,28,30,31,32,34,35	0.00880	31.047
36	400	15	22	1,2,3,8,10,12,13,15,17,18,20,21,22,24,25,26,28,30,32,33,34,35	0.00913	30.825
36	500	15	19	1,4,6,7,8,10,11,13,17,18,20,24,25,26,28,31,32,35,36	0.00994	31.595
<i>Experiment 7: For population size 20 and on the selected features different machine learning classifiers using the fivefold cross-validation for training and testing</i>						
36	100	20	26	1,2,3,4,5,6,7,8,9,10,11,16,17,18,19,20,21,25,26,27,,30,31,,33,34,35,36	0.00851	61.385
36	200	20	25	3,4,7,8,9,10,12,13,15,16,17,21,22,23,24,25,26,27,29,30,31,32,33,34,36	0.00883	61.671
36	300	20	26	2,3,6,7,8,9,10,11,12,13,14,15,16,19,20,21,22,23,24,25,27,29,31,32,33,34	0.00922	60.486
36	400	20	26	1,2,3,4,5,6,7,9,10,11,13,14,15,18,19,21,22,24,25,26,27,28,29,31,33,34	0.00997	60.435
36	500	20	25	2,3,5,6,7,8,9,10,11,12,14,17,21,22,25,26,27,28,29,31,32,33,34,35,36	0.00853	60.384

Table 9 (continued)

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution time
<i>Experiment 8: For population size 20 and on the selected features different machine learning classifiers using the tenfold cross-validation for training and testing</i>						
36	100	20	23	2,3,5,6,7,9,14,15,16,18,19,22,24,25,26,27,28,29,30,31,33,34,35	0.00859	61.503
36	200	20	24	1,3,4,5,6,8,10,11,14,16,19,20,23,24,26,27,28,29,30,32,33,34,35,36	0.00786	60.646
36	300	20	21	2,4,6,10,11,13,14,16,17,19,21,22,24,28,29,30,31,32,33,34,36	0.00981	60.771
36	400	20	25	1,3,4,6,7,9,11,12,13,14,16,18,19,20,21,22,23,24,25,26,27,28,29,30,32	0.00910	60.632
36	500	20	25	1,5,6,8,9,10,12,13,14,15,16,18,19,20,21,25,26,27,29,30,31,32,34,35,36	0.00876	60.941

to meet the many multi-objective situations in the actual world. The paper concludes by highlighting future research prospects associated with multi-objective optimization. In this article, the authors describe what is often used, whether it algorithms or test issues, so that the reader is aware of the benchmarks and other possibilities.

Researchers, Maheshwari et al. (2019), present a novel approach for glaucoma diagnosis based on bit-plane slicing (BPS) and local binary patterns (LBP). Initially, the method divides the red (R), green (G), and blue (B) channels of the input color fundus picture into bit planes. Then, they extracted LBP-based statistical features from each of the individual channels' bit planes. Then, the characteristics of the individual channels are supplied independently to three distinct SVMs for categorization. Finally, the decisions from the individual SVMs are combined at the decision level in order to classify the input fundus image as either normal or glaucoma. Using tenfold cross validation, the experimental findings indicate that the suggested method can distinguish between normal and infected cases with an accuracy of 99.30%. This paper presents a model for glaucoma diagnosis that employs multiple features (such as inferior, superior, nasal, and temporal region areas and cup-to-disc ratio) to detect glaucoma in retinal fundus images (Singh et al. 2021). Using support vector machine (SVM), the proposed model provides a maximum classification accuracy of 98.60%. In contrast to other existing models such as SVM, K-nearest neighbors (KNN), and Naive Bayes, the proposed model combines four machine learning techniques to achieve a classification accuracy of 98.60%. CAD system utilizing machine learning (ML) algorithms (for classification) and nature-inspired computing (for feature selection/reduction) resolved glaucoma detection problems (Singh et al. 2022c). Two novel two-layered techniques (BA-BCS, BCS-PSO) based on Particle Swarm Optimization (PSO), Binary Cuckoo Search (BCS), and the Bat Algorithm (BA) were presented in this recommended empirical study. In addition, they have independently examined the performances of BA, BCS, and PSO. These five (single- and two-layered) approaches are

used to generate subsets of reduced features that, when given to three machine learning (ML) classifiers, provide the maximum accuracy. Using benchmark publicly accessible datasets, ORIGA and REFUGE, as well as their combinations, the proposed method is validated. In addition, the researcher's community is provided with a multitude of different possibilities, including trade-offs.

Following this extensive survey of prior premium studies, we conclude that there is still a large scope for working on effective feature selection through soft-computing approaches for glaucoma screening, as very little work has been implemented in this direction and newly proposed algorithms are rarely used for this work. Furthermore, our work is novel in that we extracted features from different classes while also satisfying the subject image count (selected for training and testing). The subject images are a combination of different public datasets and private images taken from the internet and hospitals located near the authors' home. This study implements innovative BFO and EPO techniques and provides a hybrid method to solve the drawbacks of metaheuristics and other transformation-based FS techniques. The performance of the proposed approach is assessed using widely established benchmark fundus pictures. The proposed method is used to discover the optimal collection of features (characteristics), with the main objective of enhancing classification accuracy while limiting the number of features chosen and the error rate. The best subset of features from different benchmark datasets is created using specified and hybrid data transformation procedures. We provide the best qualities to researchers through this study: an efficient and quick support system for ophthalmologists (that they can rely on) and a software-based tool for the human race to slow down human eye sight loss by ensuring early, efficient, and effective diagnosis of this illness. The tool may be modified to function with mobile and wearable medical equipment, and it can be utilized in areas where there aren't enough skilled medical practitioners.

Table 10 Performance of different five classifiers including ensemble on population size 5,10,15,20 (K-fold cross-validation approach)

Classifiers	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution time	Two tailed test (p-value)
<i>Experiment 1: Using fivefold cross-validation measure, the classifier performance on 10 selected features of minimum cost 0.00841 (population size 5)</i>										
LR	0.938	0.705	0.869	0.869	0.670	0.679	0.964	0.8534	0.477	0.04953
RF	0.815	0.825	0.720	0.769	0.621	0.625	0.820	0.8192	11.35	0.068759
DT	0.898	0.762	0.812	0.786	0.669	0.670	0.830	0.8487	0.366	0.067251
KNN	0.905	0.729	0.817	0.770	0.649	0.652	0.817	0.8412	0.502	0.069217
SVM	0.932	0.730	0.861	0.790	0.685	0.690	0.831	0.8586	0.739	0.067101
Ensemble	0.847	0.738	0.736	0.737	0.585	0.585	0.803	0.8076	0.476	0.071388
<i>Experiment 2: Using tenfold cross validation measure, the classifier performance on 6 selected features of minimum cost 0.00991 (population size 5)</i>										
LR	0.978	0.711	0.949	0.813	0.728	0.744	0.952	0.8806	0.945	0.050938
RF	0.945	0.877	0.902	0.889	0.827	0.828	0.911	0.9206	24.659	0.056005
DT	0.937	0.839	0.886	0.862	0.786	0.787	0.888	0.9020	0.453	0.059026
KNN	0.981	0.741	0.959	0.836	0.759	0.773	0.861	0.8939	0.502	0.062744
SVM	0.984	0.721	0.963	0.825	0.745	0.762	0.852	0.8881	1.269	0.064026
Ensemble	0.957	0.798	0.914	0.852	0.776	0.780	0.877	0.8991	0.691	0.060518
<i>Experiment 3: Using fivefold cross validation measure, the classifier performance on 16 selected features of minimum cost 0.00910 (population size 10)</i>										
LR	0.932	0.708	0.857	0.776	0.665	0.672	0.948	0.8505	0.544	0.051415
RF	0.831	0.801	0.733	0.766	0.621	0.623	0.816	0.8209	11.280	0.06937
DT	0.884	0.741	0.786	0.763	0.633	0.634	0.812	0.8319	0.375	0.06937
KNN	0.916	0.694	0.804	0.703	0.541	0.554	0.805	0.8089	0.477	0.071075
SVM	0.928	0.690	0.838	0.725	0.598	0.610	0.814	0.8232	0.873	0.069678
Ensemble	0.881	0.692	0.760	0.702	0.550	0.554	0.807	0.8077	0.501	0.070762
<i>Experiment 4: Using tenfold cross validation measure, the classifier performance on 17 selected features of minimum cost 0.00863 (population size 10)</i>										
LR	0.951	0.781	0.902	0.837	0.754	0.758	0.964	0.8893	0.986	0.04953
RF	0.948	0.884	0.908	0.906	0.837	0.838	0.916	0.9252	30.556	0.055365
DT	0.959	0.832	0.922	0.875	0.808	0.811	0.895	0.9130	0.4914	0.058092
KNN	0.957	0.795	0.914	0.850	0.773	0.778	0.876	0.8980	0.674	0.060655
SVM	0.960	0.773	0.919	0.839	0.759	0.766	0.867	0.8922	1.285	0.061902
Ensemble	0.977	0.749	0.949	0.837	0.760	0.772	0.863	0.8939	0.795	0.062462
<i>Experiment 5: Using fivefold cross validation measure, the classifier performance on 23 selected features of minimum cost 0.00857 (population size 15)</i>										
LR	0.876	0.729	0.773	0.750	0.613	0.613	0.959	0.8227	0.598	0.050133
RF	0.845	0.852	0.760	0.804	0.680	0.683	0.849	0.8482	13.188	0.064458
DT	0.858	0.708	0.742	0.725	0.572	0.572	0.803	0.8035	0.427	0.071388
KNN	0.904	0.736	0.815	0.774	0.654	0.656	0.820	0.8429	0.620	0.068759
SVM	0.917	0.732	0.836	0.781	0.667	0.671	0.824	0.8499	0.847	0.068152
Ensemble	0.827	0.698	0.690	0.699	0.529	0.529	0.808	0.7694	0.590	0.070607
<i>Experiment 6: Using tenfold cross validation measure, the classifier performance on 20 selected features of minimum cost 0.00853 (population size 15)</i>										
LR	0.952	0.806	0.907	0.854	0.777	0.780	0.953	0.8991	0.944	0.05082
RF	0.947	0.882	0.907	0.894	0.835	0.835	0.915	0.9241	30.104	0.055493
DT	0.968	0.854	0.939	0.894	0.838	0.840	0.911	0.9264	0.540	0.056005
KNN	0.961	0.795	0.922	0.854	0.779	0.784	0.878	0.9009	0.677	0.060381
SVM	0.957	0.798	0.916	0.853	0.777	0.782	0.879	0.8997	1.315	0.060244
Ensemble	0.911	0.871	0.850	0.860	0.778	0.779	0.891	0.8968	0.815	0.058624

Table 10 (continued)

Classifiers	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution time	Two tailed test (p-value)
<i>Experiment 7: Using fivefold cross validation measure, the classifier performance on 28 selected features of minimum cost 0.00815 (population size 20)</i>										
LR	0.904	0.782	0.824	0.803	0.694	0.695	0.975	0.8597	0.649	0.048268
RF	0.829	0.874	0.746	0.805	0.679	0.685	0.852	0.8458	14.995	0.064026
DT	0.954	0.770	0.906	0.832	0.748	0.754	0.862	0.8870	0.407	0.062603
KNN	0.907	0.735	0.821	0.775	0.657	0.660	0.821	0.8447	0.617	0.068607
SVM	0.941	0.809	0.888	0.847	0.765	0.767	0.875	0.8933	0.837	0.60793
Ensemble	0.859	0.690	0.738	0.714	0.558	0.558	0.805	0.7977	0.616	0.071075
<i>Experiment 8: Using tenfold cross validation measure, the classifier performance on 24 selected features of minimum cost 0.00786 (population size 10)</i>										
LR	0.957	0.812	0.916	0.861	0.788	0.792	0.962	0.9044	0.977	0.049762
RF	0.943	0.881	0.899	0.890	0.828	0.828	0.912	0.9206	30.298	0.055876
DT	0.959	0.860	0.925	0.891	0.832	0.833	0.910	0.9235	0.549	0.056133
KNN	0.961	0.809	0.924	0.863	0.792	0.796	0.885	0.9061	0.713	0.05943
SVM	0.962	0.808	0.925	0.862	0.792	0.796	0.886	0.9062	1.366	0.059295
Ensemble	0.910	0.899	0.851	0.869	0.792	0.792	0.899	0.9026	0.879	0.057564

3 Materials and methods

In this section, we have discussed about the datasets selected for our study and the required details about three implemented soft-computing-based FS algorithms.

3.1 Dataset

Our dataset consists of mixture of images from 4 standard benchmark publically available dataset and one private dataset collected from hospitals located in nearby cities. The total images considered are 3112 (1226 Glaucomatic images and 1886 healthy images). These 3112 images are collected from ACRIMA (396 glaucomatous and 309 healthy images), DRISHTI (16 glaucomatous images), HRF (15 glaucomatous and 15 healthy images), ORIGA (168 glaucomatous and 482 healthy images) and PRIVATE 631 glaucomatous images and 1080 healthy images. From these images, 36 features are retrieved whose list is shown using Table 1. Figure 2 displays the framework of this glaucoma identification study.

The brief description about these features is given below:

1. **CDR (Cup-to-Disc Ratio):** The cup-to-disc ratio features are defined as ratio of cup diameter and disc diameter of fundus images.
2. **GLCM (Grey Level Co-occurrence Matrix)—**It is used to determine image texture by calculating the presence of paired pixels with certain values in spatial relationship inside an image.

3. **GLRM (grey-level run length matrix)—**It can be described as a row of identically intense pixels arranged in a particular orientation. The quantity of these pixels is known as the grey level run length, and the occurrences of these pixels are regarded as the run length value.
4. **SRE (Short-Run Emphasis)—**It is a measure of the distribution of short-run lengths. In this case, a higher SRE score suggests fine texture.
5. **LRE (Long-Run Emphasis)—**It is a measure of the distribution of long-run lengths, with greater values suggesting a coarse texture.
6. **GLU (Grey-Level Uniformity)—**When determining the true value of full length uniformity, we must greatly expand the scale.
7. **RLU (Run-Level Uniformity)—**In this, we will first convert the image to grayscale and then determine its uniformity and the amount to which the entire grey section is stretched. It will show us where we have the lightest intensity and where we need less light density.
8. **RPC (Rational polynomial coefficient)—**It is utilized to calculate the ratio of items preserved in space with changes in the regions that appear in other objects, allowing us to discriminate between them.
9. **Mat 0—**It is a correlation feature matrix that is used to extract the greatest degree, which is zero degree, of the GLCM image's features.
10. **Mat 45—**The goal of this feature is the same as the last one, but there are some differences. For example,

we are now looking at a GLCM image of glaucoma at a 45-degree angle from the origin and on an axis that is directly opposite the origin.

11. **Mat 0_avg**—In the feature, we have a correlation matrix including all the features shared by all the images.
12. **Mat 45_avg**—This feature is utilized to determine the correlation matrix coefficients of the glaucoma image characteristics. Each image following 45-degree rotation of both matrices.
13. **Mat90_avg**—This is the average used to determine the GLCM and correlation matrix characteristics of each image following 90-degree rotation of both matrices.
14. **Mat135_avg**—All of the matrices in our possession will be rotated by 135 degrees at this place. When rotate the matrices, the changes that happen will have an effect on each image.
15. **HOS (Higher Order Spectral)** —HOS characteristics are also employed to analyze picture attributes that are non-stationary, non-linear, or non-Gaussian.
16. **NRR (Neuro Retinal Rim)**—The rim that surrounds the eye is the neuroretinal rim. We discovered that the rim of a glaucoma-affected eye is somewhat larger than the rim of a glaucoma-free eye.
17. **DDL (Disc Damage Likelihood Scale)**—The ratio of the disc’s diameter to its Rim width, or the DDL, is used to predict the development of glaucoma.
18. **HOC (Higher-Order Cumulant)**—Higher-order statistics based on a picture’s moments and correlation moments are specific weights that are applied to the pixels or intensities of a picture and assessed by a function.
19. **DWT (Discrete Wavelet Transform)**—The separation of data, operators, or functions into discrete frequency components and analysis of each with a resolution proportional to scale are both carried out using wavelet transformations.
20. **FOS (First-Order Statistical)**—These traits are used to compute an image’s histogram, which shows the likelihood of an individual pixel occurring in an image, and use it to analyze the texture of the picture.
21. **Cum2est**—Cum2est is an eye feature in which we minimize the area and concentrate on the optic disc region.
22. **Bio coherence**—Bio coherence is the bispectrum that has been normalized. It is the moment spectrum of the third order.
23. **Energy**—Energy is a way to measure the size of voxels in an image. In this case, a higher value of energy means that the sum of the squares is higher.
24. **Homogeneity**—It is used for analysis the homogeneity of image.

Table 11 The best values of the performance metrics when the minimum number of features was selected (in case of BFO)

Case of minimum features extracted	List of features	Max Specificity (Classifier)	Max Sensitivity (Classifier)	Max Precision (Classifier)	Max F1-score (Classifier)	Max Kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max Accuracy (Classifier)	Minimum Execution Time
6	3,14,17,18,24,27	0.984 (SVM)	0.877 (RF)	0.963 (SVM)	0.889 (RF)	0.827 (RF)	0.828 (RF)	0.952 (LR)	0.9206 (RF)	0.453 (DT)

Table 12 The maximum values of the performance metrics (in case of BFO algorithm)

Max specificity (Classifier)	Max sensitivity (Classifier)	Max precision (Classifier)	Max F1-score (Classifier)	Max kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max accuracy (Classifier)	Minimum execution time
0.984 (SVM)	0.899 (ENS)	0.963 (SVM)	0.906 (RF)	0.838 (DT)	0.840 (DT)	0.975 (LR)	0.9264 (DT)	0.366 (DT)

Overall summary of feature selection using BFO

This table depicts the best results generated on population size 5, 10, 15 and 20 fork-fold approach (5- and 10-)

25. **Correlation**—Correlation is a way to figure out how much the grey levels of the pixels at certain positions depend on each other.
26. **Contrast**—Contrast is a measure of intensity and variation. The greater the contrast value, the greater the discrepancy in image intensity values.
27. **Dissimilarity**—It is a measure of the relationship between pairs of similar and dissimilar intensities.
28. **Entropy**—Low first-order entropy is found in homogeneous scenes, while high entropy is found in heterogeneous scenes.

3.2 Preprocessing

The optic disc and optic cup regions in retinal fundus pictures are a key area for the diagnosis of glaucoma. In the retinal fundus image, the optic disc—where the location of the optic cup is in the centre of the optic disc—is the main area of concern. The picture of the retinal fundus has been reduced by $230 \times 230 \times 3$ to allow for additional processing at the outset. In a picture of a section of the retina, the optic disc is typically in the centre of the disc. The colors of red, green, and blue (RGB)—commonly known as red, blue, and green display formats for three networks—are shown in fundus images in Fig. 4(a). It could be challenging to perform a variety of operations on the blue pattern. The grey image in Fig. 4(c) is used to symbolize the vital part of our eyes, which houses a vast quantity of information. Initially, fundus imaging converts a person's eye scan into a "gray-green" color box. Equalization is the transformation or conversion feature that is only used to produce an output picture with a consistent histogram. Changing the brightness of the input picture's histogram's equalization procedure is a highly effective method. The input image's amplitude is evenly spread over the updated image. The main focus of this approach is applying strength in a way that is comparable to the overall image. This provides the linear trend for the cumulative probability function (cdf).

To further balance the signal, we used an adaptive median filter. Pulse noise has the potential to alter the image of the fundus. Impulsive noise is produced as a result of electromagnetic interference. The adaptive median filter

is a useful filtering technique to minimize the noise of the impulse in the spatial domain. It compares each pixel with its immediate neighbors in order to identify the noisy pixels. The adaptive median filter performs spatial analysis to determine which pixels in a picture have been impacted by impulsive noise. The adaptive median filter categorizes pixels as noise by comparing each pixel in the image to its immediate neighbors (Fig. 3). Both the reference criterion and the neighborhood's size may be altered. A pixel is referred to as "impulse noise" because it differs from any of its neighbors and is architecturally inconsistent with the pixels. The median pixel value of the nearby pixels that passed the noise marking test will subsequently be used to replace these noise pixels. Figure 4a–f displays the preprocessed glaucoma picture, whereas (g)–(l) displays the preprocessed healthy image. The whole process diagram of this preprocessing stage is shown below with the help of Fig. 3.

3.3 Feature selection algorithms

Three algorithms BFO, EPO and their hybrid are short-listed for this study, whose details are given below. Table 2 depicts the various parameters and their values assigned during the implementation of these three algorithms.

The Bacteria Foraging Optimization Algorithm (BFO) is a novel addition to the family of nature-inspired optimization algorithms (Das et al. 2009). A set of tensile flagella propels the genuine bacterium while foraging. Flagella assist an E.coli bacterium in its foraging activities by allowing it to tumble or swim. Each flagellum tugs on the cell as they spin the flagella clockwise. As a consequence, the flagella move independently, and the bacteria tumbles with fewer tumblings, while in a hazardous environment, it tumbles often to locate a nutritional gradient. Moving the flagella counterclockwise allows the bacteria to move extremely quickly. In the algorithm described above, bacteria engage in chemotaxis, in which they choose to travel toward a food gradient while avoiding unpleasant environments. In general, germs migrate further in a welcoming environment. When they obtain enough nourishment, they grow in length and, in the presence of an appropriate temperature; they break in the centre to form

an identical clone of themselves. This phenomenon prompted Passino to include a reproduction event in BFO algorithm. Because of unexpected environmental changes or an assault, chemotactic development may be disrupted, and a group of bacteria may travel to other locations or be incorporated into the swarm of concern. This is an example of an elimination-dispersal event in a genuine bacterial population, in which all of the bacteria in an area are killed or a group is dispersed into a different portion of the environment (Das et al. 2009; Chen et al. 2017). Table 2 displays the list of parameters used in this algorithm.

The steps of BFO algorithms are as follows.

Step 01: Parameter Initialization

Step 02: Elimination-Dispersal Loop : $i=i+1$

Step 03: Reproduction Loop: $k=k+1$

Step 04: Chemotaxis Loop: $j=j+1$

(a) Take a chemotactic Step ($i=1, 2, 3, \dots, S$)

(b) Compute Fitness function and save this value until we find a better one via run

(c) Tumble movement: in random direction $[-1, 1]$

(d) Swimming Process

(e) Goto next bacterium ($i+1$): if $i! = s$ process the next bacteria

Step 05: Check If $j <$ chemotactic Steps, continue chemotaxis since the life of the bacteria is not over.

Step 06: Reproduction Process

Step 07: Elimination and Dispersal

Step 08: Check if Elimination-Dispersal Loop $<$ Elimination and Dispersal Steps, then go to step 2, otherwise stop and display result

The emperor penguin is one of the biggest penguins, with male and female having about the same size. An emperor penguin has a black back, a white belly, golden ear patches, and grayish-yellow breasts. The emperor penguin’s wings serve as a fin while swimming. Emperor penguins move similarly like humans. Their whole existence is spent in Antarctica, and they are famous for their capacity to procreate through the brutal Antarctic winter, which may reach minus 60 °C. Although their unique feathers and body fat protect them from chilly winds, they must cluster together to stay warm in severe cold. During mating season, each female penguin lays a single egg,

The mathematical equations used in the algorithms are as follows

$$\mu(j) = \frac{\Delta(j)}{\sqrt{\Delta(j)^M \Delta(j)}} \tag{1}$$

$$\aleph^j((k + 1, l, m) = \aleph^j(k, l, m) + A(j)\mu(j) \tag{2}$$

$$\begin{aligned} & \text{Bcc}(\aleph, \aleph^i(k, l, m)) \\ &= \sum_{j=1}^o \left[-r_{\text{attract}} \exp 1 \left(x_{\text{attract}} \sum_s^h (\aleph_d - \aleph_d^j)^2 \right) \right] \\ &+ \sum_{j=1}^o \left[-h1_{\text{repellant}} \exp 1 \left(x_{\text{repellant}} \sum_s^h (\aleph_d - \aleph_d^j)^2 \right) \right] \end{aligned} \tag{3}$$

$$F(j, k, m, l) = F(j, k, m, l) + \text{Bcc}(\aleph, \aleph^j(k, l, m)) \tag{4}$$

$$Q_r = \frac{Q}{2} \tag{5}$$

which is then passed on to one of the males. Females will go up to 80 km in the open sea to hunt after the egg transfer. Male penguins utilize their brood pouches to keep the eggs warm until they hatch, allowing the eggs to survive until they hatch. A female emperor penguin normally returns to the nest after spending approximately two months in the ocean with food in her stomach, which she vomits up for the chicks to ingest and take over care of. Emperor penguins are excellent divers in addition to being excellent swimmers. They often go foraging and hunting together. They are the only species that huddle to live through the Antarctic winter. After dissecting the huddling behavior of emperor penguins into four stages, a mathematical model was built. To begin, emperor penguins form their huddle boundaries at random. Second, they calculate the temperature profile around them. Third, they use this approach to compute penguin distances in order to simplify exploring and exploiting emperor penguins. Finally, they choose the effective mover as the best option and update

Table 13 Feature selection on population size 5, 10, 15 and 20 using fivefold and tenfold cross-validation

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution time
<i>Experiment 1: Performance on Population size 5 (fivefold)</i>						
36	100	5	8	1, 3, 12, 16, 17, 19, 26, 31	0.999	0.505
36	200	5	9	3, 8, 10, 11, 18, 26, 29, 30, 34	0.829	1.966
36	300	5	7	13, 16, 19, 20, 24, 25, 35	0.867	2.024
36	400	5	8	11, 14, 15, 17, 22, 24, 28, 33	0.888	2.481
36	500	5	9	1, 4, 6, 9, 17, 19, 24, 27, 30	0.996	2.696
<i>Experiment 2: Performance on Population size 5(tenfold)</i>						
36	100	5	8	3, 10, 13, 15, 16, 21, 29, 35	0.831	0.528
36	200	5	11	9, 14, 17, 22, 23, 24, 25, 29, 32, 34, 35	1.094	1.693
36	300	5	6	3, 10, 19, 22, 26, 32	1.002	1.870
36	400	5	8	3, 9, 11, 13, 18, 24, 30, 35	0.865	2.378
36	500	5	9	4, 6, 9, 13, 15, 18, 21, 23, 35	0.990	2.740
<i>Experiment 3: Performance on Population Size 10(fivefold)</i>						
36	100	10	17	1, 3, 7, 9, 10, 12, 14, 15, 16, 21, 25, 26, 29, 30, 32, 33, 34	0.985	2.058
36	200	10	16	3, 4, 7, 8, 12, 16, 19, 20, 21, 22, 24, 27, 28, 29, 32, 36	1.025	6.447
36	300	10	16	1, 4, 5, 6, 8, 10, 13, 15, 16, 17, 20, 21, 23, 29, 32, 33	0.883	6.573
36	400	10	15	1, 3, 6, 7, 8, 12, 16, 18, 22, 23, 24, 29, 31, 32, 33	0.960	7.346
36	500	10	16	1, 2, 4, 9, 11, 15, 18, 20, 22, 24, 26, 27, 30, 32, 33, 35	0.946	7.775
<i>Experiment 4: Performance on Population Size 10(tenfold)</i>						
36	100	10	15	3, 4, 6, 7, 8, 11, 12, 20, 23, 24, 29, 31, 32, 34, 35	1.023	1.891
36	200	10	13	6, 8, 11, 12, 13, 15, 19, 20, 21, 22, 23, 32, 35	0.698	5.730
36	300	10	14	1, 2, 6, 11, 15, 21, 22, 23, 24, 28, 32, 33, 34, 36	1.014	6.547
36	400	10	13	5, 6, 7, 10, 11, 12, 18, 19, 21, 23, 26, 30, 35	0.757	6.686
36	500	10	13	2, 3, 4, 8, 11, 14, 16, 18, 19, 22, 23, 27, 28	1.036	8.460
<i>Experiment 5: Performance on Population Size 15(fivefold)</i>						
36	100	15	18	3, 4, 6, 13, 14, 15, 16, 18, 21, 24, 25, 26, 28, 31, 32, 33, 34, 36	0.991	6.032
36	200	15	21	1, 2, 3, 4, 9, 11, 13, 14, 15, 16, 17, 20, 21, 26, 28, 29, 30, 33, 34, 35, 36	0.992	7.915
36	300	15	17	1, 4, 7, 8, 9, 10, 12, 17, 18, 19, 20, 22, 25, 29, 30, 33, 36	0.916	10.05
36	400	15	20	1, 3, 4, 6, 12, 14, 16, 17, 18, 19, 23, 24, 26, 27, 30, 31, 32, 34, 35, 36	0.767	9.314
36	500	15	21	1, 4, 10, 11, 14, 16, 18, 19, 20, 21, 23, 24, 26, 29, 30, 31, 32, 33, 34, 35, 36	0.882	10.95
<i>Experiment 6: Performance on Population Size 15(tenfold)</i>						
36	100	15	21	2, 3, 4, 5, 8, 10, 11, 14, 15, 16, 18, 20, 23, 26, 27, 28, 29, 30, 31, 32, 36	0.935	6.959
36	200	15	15	1, 3, 4, 5, 6, 7, 10, 11, 13, 14, 15, 16, 19, 25, 28	0.968	6.752
36	300	15	20	2, 3, 6, 8, 9, 13, 14, 15, 18, 19, 21, 23, 26, 27, 29, 30, 31, 32, 34, 35	0.864	11.135
36	400	15	20	2, 4, 8, 9, 10, 11, 13, 15, 16, 17, 20, 25, 27, 28, 29, 30, 32, 33, 34, 35	0.999	12.619
36	500	15	20	1, 3, 4, 6, 9, 10, 12, 14, 16, 20, 22, 23, 24, 25, 29, 30, 31, 32, 34, 35	0.741	12.264
<i>Experiment 7: Performance on Population Size 20(fivefold)</i>						
36	100	20	25	1, 2, 3, 7, 9, 10, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 24, 25, 26, 27, 28, 31, 32, 35, 36	0.864	7.909
36	200	20	26	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 15, 16, 17, 18, 20, 22, 24, 26, 27, 28, 30, 31, 33, 34, 35	1.027	8.377
36	300	20	23	1, 2, 3, 4, 5, 8, 9, 10, 14, 16, 18, 19, 21, 22, 24, 25, 27, 28, 29, 31, 32, 34, 36	0.952	16.953

Table 13 (continued)

No. of features	No. of iteration	Population size	Features selected	Features number	Fitness value	Execution time
36	400	20	24	1, 2, 3, 4, 7, 9, 10, 11, 12, 13, 15, 17, 18, 19, 20, 26, 27, 28, 29, 30, 32, 33, 34, 36	0.920	17.677
36	500	20	24	1, 2, 4, 5, 6, 7, 10, 12, 16, 17, 18, 20, 21, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 36	0.970	17.373
<i>Experiment 8: Performance on Population Size 20(tenfold)</i>						
36	100	20	26	1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 20, 22, 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 36	0.814	7.479
36	200	20	24	2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16, 17, 20, 21, 22, 24, 25, 26, 27, 29, 30, 32	0.763	14.575
36	300	20	26	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 16, 17, 18, 19, 20, 22, 23, 25, 26, 27, 28, 31, 34, 36	1.019	14.141
36	400	20	26	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 20, 21, 22, 26, 27, 29, 31, 33, 35, 36	0.945	15.592
36	500	20	21	1, 5, 6, 8, 9, 10, 11, 13, 14, 17, 18, 19, 20, 21, 22, 24, 26, 27, 29, 30, 32	0.836	17.072

the emperor penguin placements to recalculate the huddle’s boundaries. This mathematical process’s main purpose is to discover the most effective mover. The huddle is considered to be on a 2D polygonal surface with a L shape. In reaction to this huddling behavior, the EPO algorithm was created (Dhiman and Kumar 2018; Khalid et al. 2022). Figure 5 depicts the flowchart of huddling process of EPO algorithm and complete EPO algorithm.

The steps of EPO algorithms are as follows.

Step 01: Generate the Emperor Penguins Population.

Step 02: Set Initial Parameters such as Maximum Iterations, Temperature, A,C

Step 03: Calculate the fitness values for all search agent.

Step 04: Determine the Huddle Boundary for Emperor Penguins Using:

$$\Re = \nabla\beta \tag{6}$$

$$\omega = \beta + i\zeta \tag{7}$$

Step 05: Calculate temperature profile (Temp’) around the Huddle using:

$$\text{Temp}' \leftarrow \left(\text{Temp} - \frac{Z_{\text{iter}}}{\psi - Z_{\text{iter}}} \right) \tag{8}$$

$$\text{Temp} = \begin{cases} 0, & \text{if } R^* > 1 \\ 1, & \text{if } R^* < 1 \end{cases} \tag{9}$$

Step 06: Compute the distance between the emperor penguins using:

$$\overrightarrow{\text{Dis}}_{\text{ep}} = \text{Abs} \left(S1(\overrightarrow{A}).P1(\overrightarrow{x1}) - c\overrightarrow{1}.P1_{\text{ep}}(\overrightarrow{x1}) \right) \tag{10}$$

Step 07: Update the position of Emperor Penguins

$$\overrightarrow{P1}_{\text{ep}}(x1 + 1) = \overrightarrow{P1}(x) - \overrightarrow{c} \cdot \overrightarrow{\text{Dis}}_{\text{ep}} \tag{11}$$

Step 08: If any emperor penguin goes beyond the Huddle Boundary improve its position.

Step 09: Calculate the fitness values for each search agent and update new optimal solution position.

Step 10: If Stopping Criteria met STOP else Goto Step 05.

Step 11: Return Best Emperor Penguins/Optimal Solutions.

The mathematical equations applied in the algorithm are as follows: Let γ define the wind velocity and \Im be the gradient of γ .

$$\Im = \Delta\gamma \tag{12}$$

Vector \Re is combined with γ to generate the complex potential

$$\text{AF} = \gamma + j\Re \tag{13}$$

where j denotes the imaginary constant and AF is an analytical function on the polygon plane.

The temperature profile around the huddle A' is computed as follows:

$$A' = \left(A - \frac{\text{Maximum}_{\text{iteration}}}{y - \text{Maximum}_{\text{iteration}}} \right) \tag{14}$$

$$A = \begin{cases} 0, & \text{if } \text{Radius} > 1 \\ 1, & \text{if } \text{Radius} < 1 \end{cases} \tag{15}$$

Here y define the current iteration, $\text{Maximum}_{\text{iteration}}$ represents the maximum number of iteration. A is the time for finding best optimal solution in a search space.

$$\overrightarrow{\text{Dis}}_{\text{epn}} = \text{Abs} \left(k(\overrightarrow{A1}).\overrightarrow{l}(y) - c\overrightarrow{1}.l_{\text{epn}}(y) \right) \tag{16}$$

where $\overrightarrow{\text{Dis}}_{\text{epn}}$ shows the distance between the emperor

Table 14 Performance of different five classifiers including ensemble on population size 5,10,15,20 (K-fold Cross-Validation Approach)

Classifiers	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution Time	Two tailed test (p-value)
<i>Experiment 1: Using fivefold cross validation measure, the classifier performance on 9 selected features of minimum cost 0.82956 (population size 5)</i>										
LR	0.953	0.801	0.908	0.851	0.774	0.778	0.877	0.9000	0.504	0.060518
RF	0.837	0.827	0.745	0.784	0.649	0.651	0.832	0.8337	12.234	0.066951
DT	0.853	0.805	0.760	0.782	0.650	0.651	0.829	0.8560	0.376	0.0674
KNN	0.902	0.732	0.811	0.769	0.647	0.649	0.817	0.8500	0.445	0.69217
SVM	0.944	0.806	0.892	0.847	0.766	0.768	0.875	0.9039	0.812	0.060793
Ensemble	0.851	0.819	0.737	0.776	0.636	0.638	0.825	0.8373	0.622	0.068001
<i>Experiment 2: Using tenfold cross-validation measure, the classifier performance on 8 selected features of minimum cost 0.83156 (population size 5)</i>										
LR	0.952	0.727	0.898	0.803	0.708	0.717	0.839	0.8702	1.591	0.065915
RF	0.937	0.824	0.884	0.853	0.773	0.774	0.880	0.8962	23.565	0.060108
DT	0.897	0.793	0.817	0.805	0.695	0.696	0.845	0.8697	0.711	0.065038
KNN	0.942	0.814	0.890	0.850	0.771	0.772	0.878	0.9057	0.843	0.060381
SVM	0.951	0.830	0.908	0.867	0.796	0.798	0.891	0.9073	1.613	0.058624
Ensemble	0.863	0.830	0.777	0.803	0.683	0.684	0.846	0.8511	0.717	0.064892
<i>Experiment 3: Using fivefold cross validation measure the classifier performance on 16 selected features of minimum cost 0.88324 (population size 10)</i>										
LR	0.931	0.741	0.861	0.797	0.693	0.698	0.836	0.8621	0.564	0.066358
RF	0.863	0.874	0.786	0.828	0.720	0.723	0.868	0.8673	12.159	0.061762
DT	0.929	0.771	0.863	0.815	0.717	0.720	0.850	0.8719	0.359	0.064314
KNN	0.897	0.746	0.807	0.775	0.654	0.655	0.822	0.8424	0.496	0.068455
SVM	0.941	0.843	0.892	0.867	0.793	0.794	0.892	0.9055	0.768	0.058491
Ensemble	0.870	0.763	0.772	0.768	0.635	0.635	0.817	0.8314	0.614	0.069217
<i>Experiment 4: Using tenfold cross validation measure, the classifier performance on 13 selected features of minimum cost 0.69869 (population size 10)</i>										
LR	0.957	0.803	0.916	0.856	0.781	0.785	0.880	0.9015	0.876	0.060108
RF	0.947	0.934	0.915	0.901	0.875	0.845	0.915	0.9541	28.447	0.055493
DT	0.959	0.858	0.924	0.890	0.831	0.832	0.909	0.9229	0.438	0.056262
KNN	0.961	0.805	0.923	0.860	0.788	0.792	0.883	0.9044	0.591	0.0597
SVM	0.957	0.797	0.916	0.852	0.776	0.780	0.877	0.9091	1.169	0.060518
Ensemble	0.911	0.877	0.850	0.864	0.784	0.784	0.894	0.9091	0.785	0.058225
<i>Experiment 5: Using fivefold cross validation measure, the classifier performance on 20 selected features of minimum cost 0.76739 (population size 15)</i>										
LR	0.871	0.754	0.771	0.762	0.628	0.628	0.812	0.8385	0.634	0.069986
RF	0.828	0.835	0.737	0.783	0.645	0.648	0.831	0.8308	13.94	0.067101
DT	0.863	0.754	0.760	0.757	0.618	0.618	0.808	0.8232	0.406	0.070607
KNN	0.909	0.730	0.823	0.774	0.655	0.658	0.820	0.8441	0.609	0.068759
SVM	0.909	0.740	0.825	0.780	0.664	0.666	0.824	0.8576	0.825	0.068152
Ensemble	0.827	0.668	0.690	0.679	0.519	0.519	0.808	0.8094	0.647	0.070607
<i>Experiment 6: Using tenfold cross validation measure, the classifier performance on 20 selected features of minimum cost 0.74174 (population size 15)</i>										
LR	0.953	0.811	0.909	0.857	0.782	0.785	0.882	0.9015	1.094	0.059836
RF	0.952	0.881	0.914	0.907	0.840	0.840	0.916	0.9364	31.096	0.055365
DT	0.973	0.858	0.931	0.893	0.836	0.837	0.911	0.9252	0.540	0.056005
KNN	0.960	0.816	0.922	0.866	0.796	0.799	0.888	0.9178	0.663	0.059026
SVM	0.957	0.803	0.916	0.856	0.781	0.785	0.880	0.9015	1.235	0.060108
Ensemble	0.930	0.879	0.879	0.879	0.810	0.810	0.905	0.9119	0.854	0.069999

Table 14 (continued)

Classifiers	Specificity	Sensitivity	Precision	F1-score	Kappa score	MCC	AUC value	Accuracy	Execution Time	Two tailed test (p-value)
<i>Experiment 7: Using fivefold cross validation measure, the classifier performance on 25 selected features of minimum cost 0.86435 (population size 20)</i>										
LR	0.884	0.770	0.792	0.781	0.658	0.658	0.827	0.8424	0.748	0.05678
RF	0.840	0.879	0.760	0.815	0.696	0.701	0.859	0.8545	16.951	0.063027
DT	0.954	0.773	0.907	0.834	0.751	0.756	0.863	0.8981	0.455	0.062462
KNN	0.913	0.719	0.826	0.769	0.650	0.654	0.816	0.8424	0.689	0.06937
SVM	0.926	0.744	0.852	0.795	0.689	0.693	0.835	0.8597	1.103	0.066506
Ensemble	0.857	0.716	0.713	0.761	0.588	0.591	0.807	0.8294	0.665	0.070762
<i>Experiment 8: Using tenfold cross validation measure, the classifier performance on 24 selected features of minimum cost 0.76353 (population size 20)</i>										
LR	0.957	0.808	0.917	0.859	0.786	0.789	0.883	0.9032	1.054	0.0597
RF	0.943	0.881	0.899	0.890	0.828	0.828	0.912	0.9206	35.035	0.055876
DT	0.959	0.858	0.944	0.890	0.831	0.832	0.909	0.9229	0.569	0.056262
KNN	0.959	0.797	0.919	0.853	0.778	0.783	0.878	0.9003	0.639	0.060381
SVM	0.957	0.797	0.916	0.852	0.776	0.780	0.877	0.9091	1.433	0.060518
Ensemble	0.920	0.876	0.864	0.870	0.794	0.794	0.898	0.9044	0.892	0.057696

Table 15 The best values of the performance metrics when the minimum number of features was selected (in case of hybrid algorithm)

Case of minimum features extracted	List of features	Max specificity (Classifier)	Max sensitivity (Classifier)	Max precision (Classifier)	Max F1-score (Classifier)	Max Kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max accuracy (Classifier)	Minimum execution time
8	3, 10, 13, 15, 16, 21, 29, 35	0.952 (LR)	0.830 (SVM)	0.908 (SVM)	0.867 (SVM)	0.796 (SVM)	0.798 (SVM)	0.891 (SVM)	0.9073 (SVM)	0.711 (DT)

Table 16 The maximum values of the performance metrics (in case of hybrid algorithm)

Max specificity (Classifier)	Max sensitivity (Classifier)	Max precision (Classifier)	Max F1-score (Classifier)	Max Kappa score (Classifier)	Max MCC (Classifier)	Max AUC value (Classifier)	Max accuracy (Classifier)	Minimum execution time
0.973 (DT)	0.934 (RF)	0.944 (DT)	0.907 (RF)	0.840 (RF)	0.840 (RF)	0.916 (RF)	0.9541 (RF)	0.406 (DT)

Overall summary of feature selection using Hybrid

This table depicts the best results generated on population size 5, 10, 15 and 20 for k-fold approach (5- and 10-)

penguin and best fittest search agent. y shows the current iteration. \vec{AI} and \vec{cI} are used to avoid the collision between neighbors. $k()$ defines the social forces of emperor penguins.

$$\vec{AI} = (\text{Mov} \times (A' + \text{Poly_grid}(\text{Acc})) \times \text{Random}()) - A' \tag{17}$$

$$\text{Poly_grid}(\text{Acc}) = \text{Abs}\left(\vec{l} - \vec{l}_{\text{epn}}\right) \tag{18}$$

$\text{Poly_grid}(\text{Acc})$ defines the polygon grid accuracy by comparing difference between emperor penguins and random function $\text{Random}()$.

$$\vec{cI} = \text{Random}() \tag{19}$$

$$k(\vec{AI}) = (\sqrt{f \cdot e^{-x/l}} - e^{-x})^2 \tag{20}$$

where e defines the expression function. f and l are control parameters for better exploration.

Table 3 depicts the parameters and their corresponding values used in the algorithms. The objective function for all the optimization algorithms is according to Eq. (21).

$$\text{fitness}(x_i(t)) = \sum (x_i(t))^2 / 10^5 \quad (21)$$

Here, Eq. (21) is basically a sphere function, one of many benchmark optimization functions used for solving optimization in single variable (Tang et al. 2007). So in the proposed work we have used single variable sphere function for updating the cost on different set of population. The function is suitable for single objective optimization. This means that it presents one mode and has a single global optimum. In our case, our objective is to finding the minimum fitness value and after that pass, we have forwarded set of features retrieved on minimum fitness value.

The operating system used is Windows 7 Professional. System type is 64-bit operating system with Intel(R) Core ((TM) i3-4150 CPU @ 3.50 GHz GHz processor and Python 3.7.15 is the programming language used.

The pseudocode of the hybrid BFOAEPO algorithm is given below:

```

Step 01: A dataset for experimentation has been loaded.
Step 02: Set the initial BFOA parameters
Step 03: Produce a Sample Size for Further Study  $X(i = 1, 2, \dots, \text{PopSize})$  (By BFOA)
Step 04: for t=1 to Max_iters
Step 05: The loops of Elimination and Dispersion
Step 06: Loop of reproduction
Step 07: Perform the computation on the fitness function
Step 08: Eliminate unpromising bacteria
Step 09: Update Population (By EPO)
Step 10: Set the initial EPO parameters
Step 11: Perform the computation on the fitness function
Step 12: if ( getting the better solution) then Update the current best
        endif
        Update y=y+1
    endfor
Step 13: Return the Optimal Solution

```

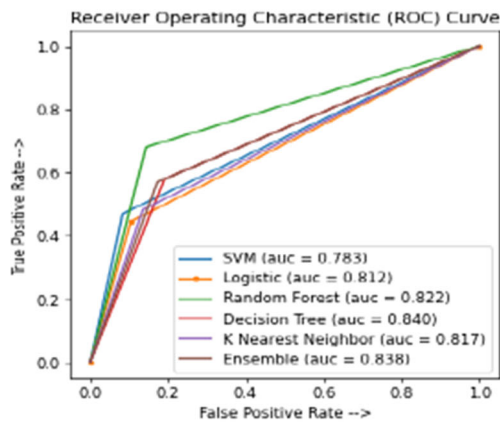
4 Results

Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15 assemble the calculated results in tabular style. Tables 4 and 5 are devoted to presenting the findings of the EPO algorithm. Similarly, Tables 8 and 9 are devoted to presenting the results created by the BFO algorithm, while Tables 12 and 13 are dedicated to presenting the results obtained by the hybrid method. Apart from this, we have also created two

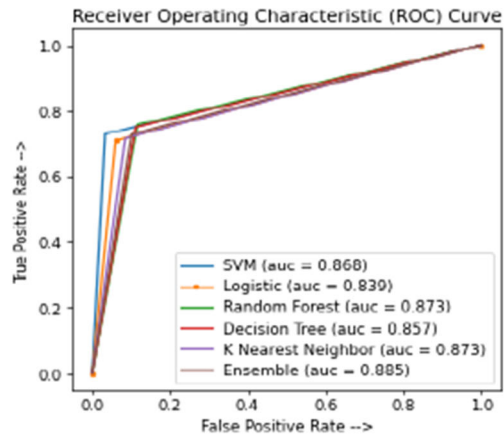
Fig. 6 ROC curve of different classifiers on k-fold cross-validation, **a** For minimum cost **0.89409** (population size 5), **b** For minimum cost **0.97123** (population size 5), **c** For minimum cost **0.00798** (population size 10), **d** For minimum cost **0.00855** (population size 10), **e** For minimum cost **0.00790** (population size 15), **f** For minimum cost **0.00792** (population size 15), **g** For minimum cost **0.00715** (population size 20), **h** For minimum cost **0.83086** (population size 20)

small tables (per algorithm) depicting vital information such as what were the values of different performance measuring metrics (like sensitivity, specificity, accuracy, F1-score etc.) when the number of features was at a minimal during the whole experimentation phase. When the features were minimal, researchers rarely showed this type of data (values of various performance measuring measures) in previous studies. These traits are shown in the table along with the outcomes they produced (Tables 7, 11, and 15). If one does not have the time to calculate and extract all the features, they may still utilize these features without making too many sacrifices in terms of getting the

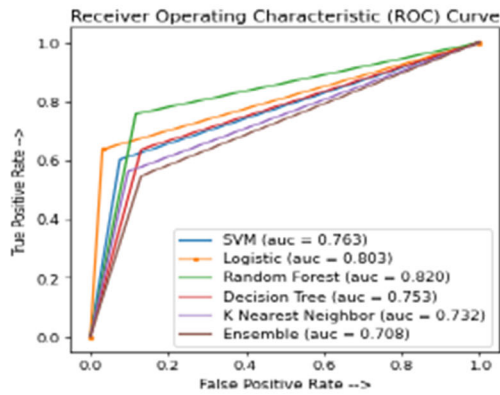
best results. The results are quite satisfying, even with a relatively small number of features. The best value for each algorithm created throughout the course of all eight tests with each algorithm is shown in the second table (Tables 8, 12, and 16). Figures 6, 7, and 8 display the collection of ROC graphs for various experiments performed. Figure 9 depicts the convergence graphs (iteration vs. fitness) of 12 experiments (out of 24 experiments performed) due to space constraints. Along with this, hypothesis testing was



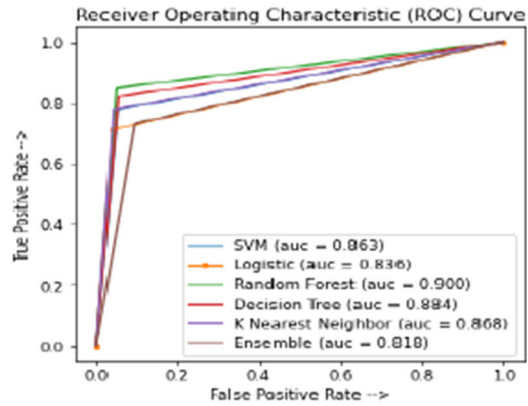
(a) For minimum cost **0.89409** (population size 5)



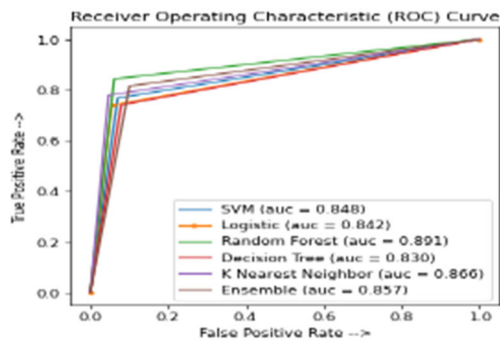
(b) For minimum cost **0.97123** (population size 5)



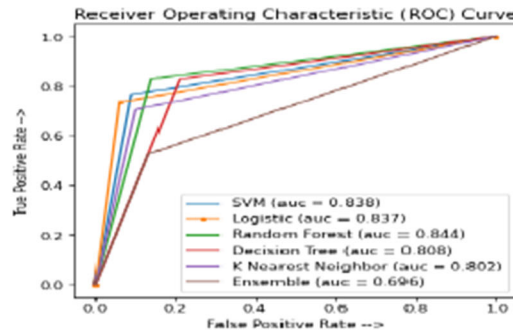
(c) For minimum cost **0.00798** (population size 10)



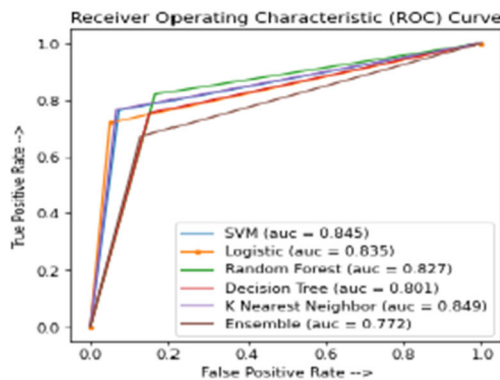
(d) For minimum cost **0.00855** (population size 10)



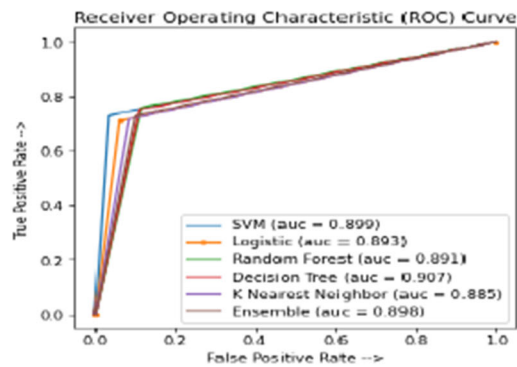
(e) For minimum cost **0.00790** (population size 15)



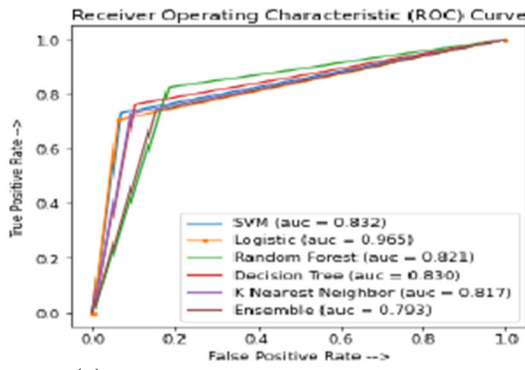
(f) For minimum cost **0.00792** (population size 15)



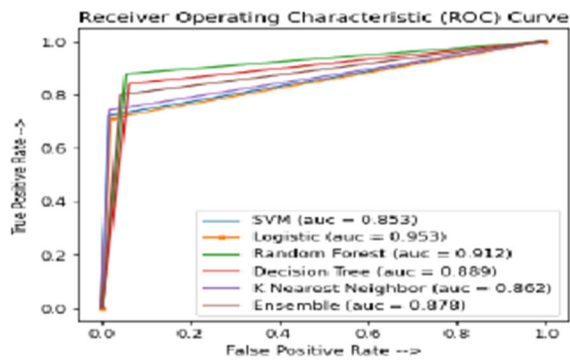
(g) For minimum cost **0.00715** (population size 20)



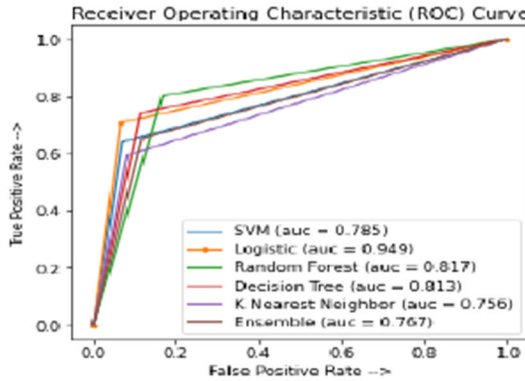
(h) For minimum cost **0.83086** (population size 20)



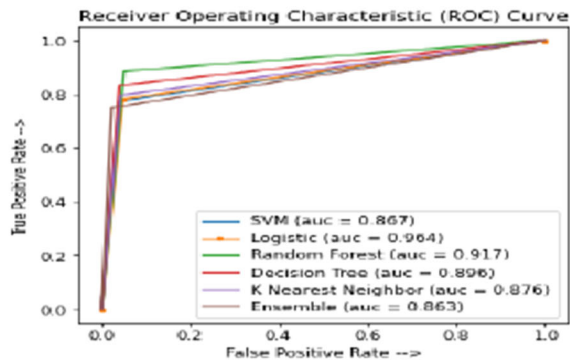
(a) For minimum cost **0.00841** (population size 5)



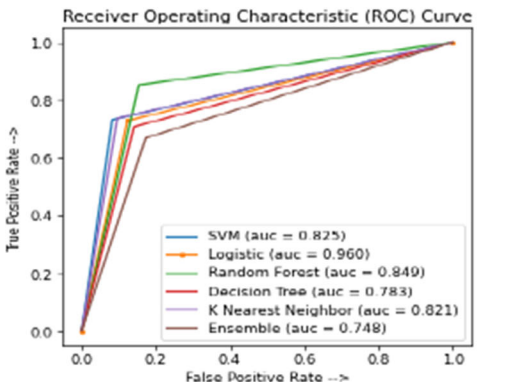
(b) For minimum cost **0.00991** (population size 5)



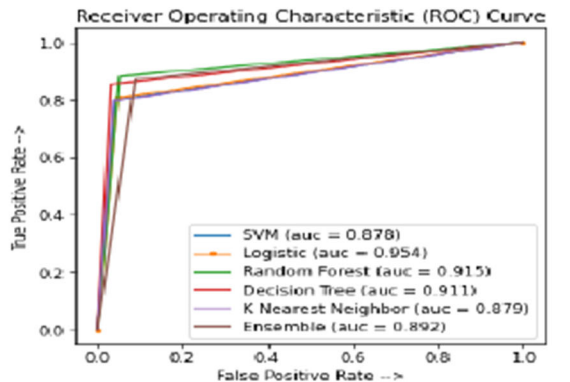
(c) For minimum cost **0.00910** (population size 10)



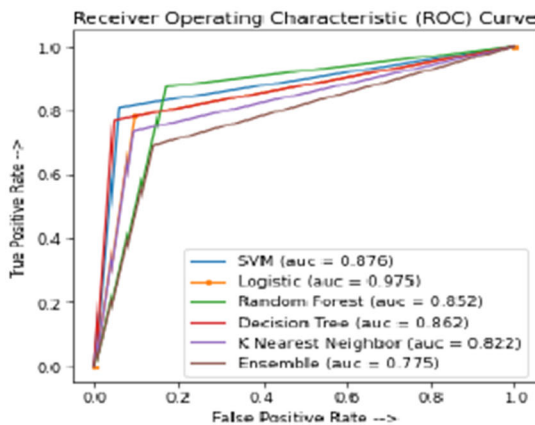
(d) For minimum cost **0.00863** (population size 10)



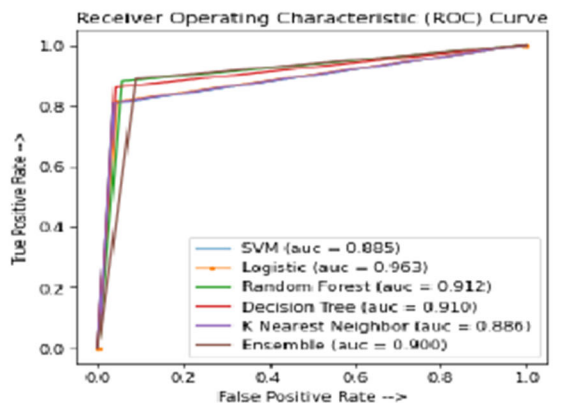
(e) For minimum cost **0.00857** (population size 15)



(f) For minimum cost **0.00853** (population size 15)



(g) For minimum cost **0.00815** (population size 20)



(h) For minimum cost **0.00786** (population size 20)

◀**Fig. 7** Combined ROC curve of different classifiers including ensemble based on k-fold approach, **a** For minimum cost **0.00841** (population size 5), **b** For minimum cost **0.00991** (population size 5), **c** For minimum cost **0.00910** (population size 10), **d** For minimum cost **0.00863** (population size 10), **e** For minimum cost **0.00857** (population size 15), **f** For minimum cost **0.00853** (population size 15), **g** For minimum cost **0.00815** (population size 20), **h** For minimum cost **0.00786** (population size 20)

also performed in the experiments. To determine statistically whether one model has a substantially different AUC from another, we applied DeLong's test to provide a p -value (DeLong et al. 1988; Sun and Xu 2014). In our study, the hypothesis was set up so that, assuming the two models performed as expected, $H_0: AUC1 = AUC2$. The alternative hypothesis, $H_1: AUC1 \neq AUC2$, states that the two models' AUCs differ. We computed the z -value and associated p -value. The overall ML model's p -values were less than 0.05 for all comparisons. After getting the Z -score, we use the "two-tailed test" to see whether the AUC of model A is different from the AUC of model B. A brief description of the terms used in the remainder of the paper is shown in Table 4.

4.1 Experiment based on the Emperor penguin optimization (FS using EPO)

4.2 Experiment based on the bacterial foraging optimization algorithm using K-fold approach (FS using BFO)

4.3 Experiment based on the hybrid of BFO and EPO using K-fold approach (FS using Hybrid approach)

The following convergence graphs (figures) of 12 experiments have been shown here (out of 24 experiments performed), due to space constraints.

5 Discussion and comparison with prior studies

This section is divided into two subsections. The first is dedicated to a discussion and findings of our work. In the later half, we have focused on comparison of our work with prior state-of-the-art studies.

5.1 Discussion

Three algorithms have been implemented in this proposed approach: BFO, EPO, and an amalgamation of these two. The goal of these methods is to divide the set of original

images into two classes. The extracted features were forwarded to various machine learning classifiers for categorization of the subject fundus images. Twenty-four trials on these three algorithms were conducted (eight individually). All of the experiments have a common objective function. During a specific experiment, the population size remains constant, and the performance of this population is evaluated by varying the number of iterations from 100 to 500 (with a gap of 100). The case in which we find the minimum value of the objective function is chosen, and the features returned by this case are forwarded to the classifier; the remaining four cases are discarded and will not be considered further. The population size is then varied in fivefold increments from 5 to 20. According to the above tables, the minimum number of features returned is 5, and the maximum number of features returned is 26. Thus, the feature reduction scale ranges from 86.11% (the highest) to 27.77% (the lowest). The execution time is also computed in two ways: first, we show the iteration time of soft-computing algorithms; second, we show the training and testing time of machine learning classifiers. Six machine learning classifiers were chosen and put into action. The first five are classical, while the last is an ensemble of these five. Accuracy, sensitivity, specificity, precision, the F1-Score, the Kappa-Score, the MCC, and the AUC are all used to evaluate the performance of these classifiers. All of these criteria are important in predicting human diseases using medical pictures. All of these indicators are rarely computed in a single study.

We know that accuracy is a vital indicator of performance, and our findings on this parameter are remarkable; we were able to get the best accuracy of 0.9541 in the combination of a hybrid algorithm and a random forest classifier. In the hybrid example, the least accuracy generated is 0.8094; similarly, in the BFO scenario, the least accuracy generated is 0.8035 (and a maximum of 0.9264(with DT)); and finally, in the EPO case, the least accuracy generated is 0.8030 (and a maximum of 0.9194 (with DT)). As a result, we can conclude that we were never able to obtain an accuracy of less than 0.80 in any of the cases, despite using a few selections of features; this demonstrates that the suggested method is satisfactory. Another important measurement aspect is sensitivity; a 100% sensitive test will identify all people with the illness by testing positive. When generated with an ensemble, the maximum sensitivity for EPO is 0.88589. For BFO, the best match was BFO-ensemble, which had a maximum value of 0.89906. Finally, the maximum value for hybrid is 0.93431 when generated with random forest.

Specificity refers to a test's capacity to correctly exclude healthy people who are free of a problem. A positive test result would unequivocally confirm the existence of the sickness, because a test with 100% specificity would

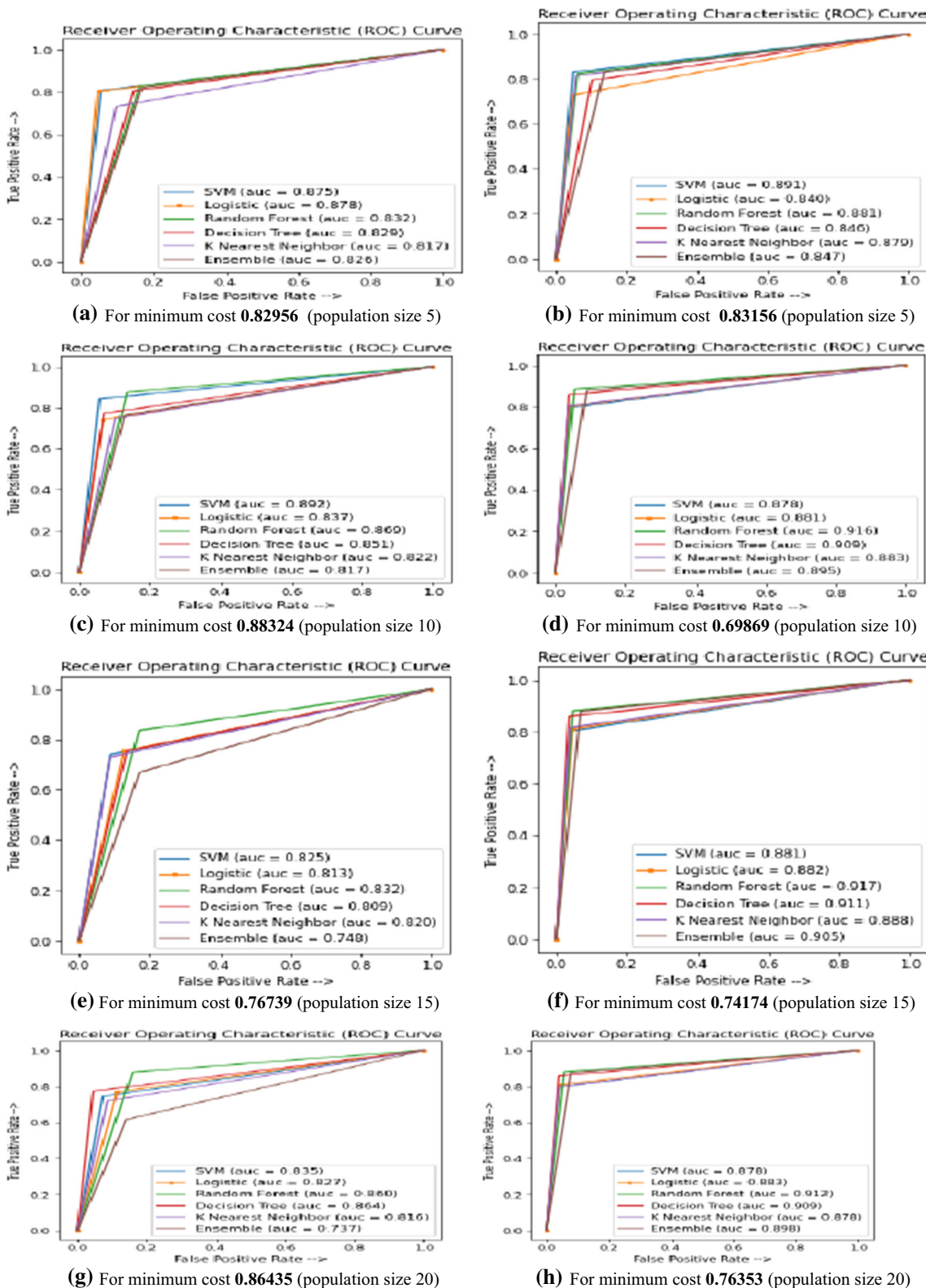


Fig. 8 Combined ROC curve of different classifier including Ensemble based on 70:30 splitting approach, **a** For minimum cost **0.82956** (population size 5), **b** For minimum cost **0.83156** (population size 5), **c** For minimum cost **0.88324** (population size 10), **d** For

minimum cost **0.69869** (population size 10), **e** For minimum cost **0.76739** (population size 15), **f** For minimum cost **0.74174** (population size 15), **g** For minimum cost **0.86435** (population size 20), **h** For minimum cost **0.76353** (population size 20)

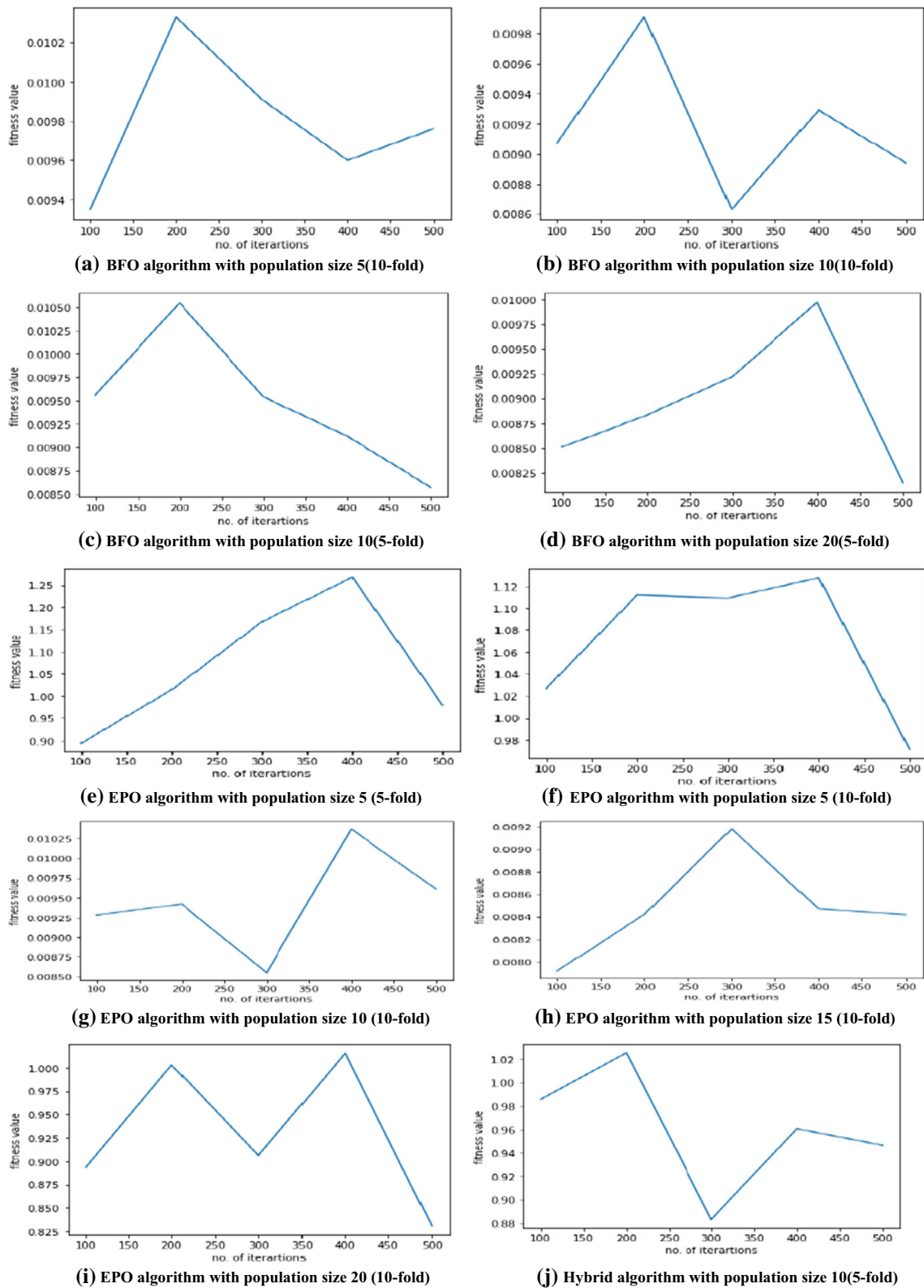


Fig. 9 Fitness Vs Iterations graphs of 12 experiments

identify all people who do not have the ailment by testing negative. The range of results demonstrates the success of

this strategy. The next essential factor that is computed is specificity. The range for EPO is 0.88378 to 0.97899 (with

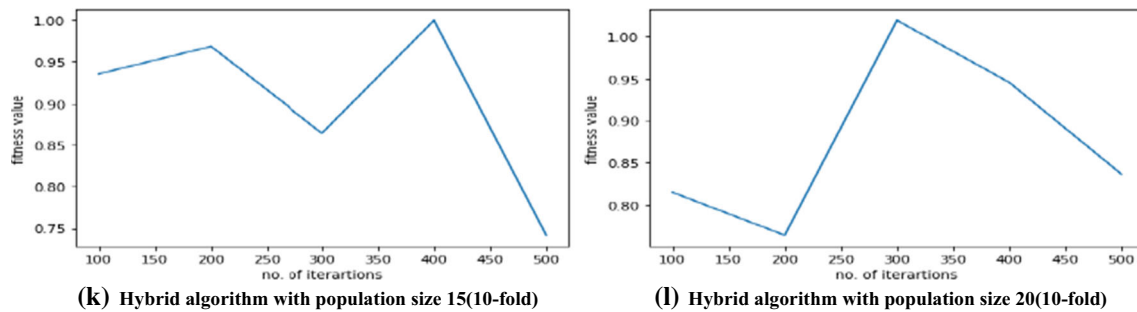


Fig. 9 continued

LR), the range for BFO is 0.81552 to 0.98447 (with SVM), and the range for the hybrid method is 0.82739 to 0.97347 (with DT). As a result, we were able to achieve at least 81% specificity in all situations. Because the F1-score is the harmonic mean of accuracy and recall, it keeps the classifier in balance. The maximum value of 0.88635 is determined when using the EPO-DT combination, 0.90638 when using the BFO-RF combination, and 0.90749 when using the hybrid-RF combination. Precision is defined as the ratio of true positives to all positives. It is critical that we do not start treating a patient who does not have a true glaucoma issue but does according to our model. When we computed precision, we got the highest value of 0.95043 in EPO, the highest value of 0.96398 in BFO, and the highest value of 0.94491 in the hybrid algorithm.

There is a set of diagnostic sensitivity and specificity values for each cutoff. To make a ROC graph, we plotted these pairs of values on a graph with specificity on the x-axis and sensitivity on the y-axis. The area under the curve (AUC) and the shape of the ROC curve both contribute to establishing a test's discriminative capacity. As the curve approaches the upper-left hand corner and the area under the curve increases, the test performs better at distinguishing between sick and non-sick situations. The area under the curve, a trustworthy indicator of how well the test was performed, can have any value between 0 and 1. The area of a non-discriminatory test is 0.5, whereas the AUC of a perfect diagnostic test is 1.0. Finally, we generated AUC scores, and the approach performed admirably in this situation as well, with a range of 0.80138 to 0.90698 in the EPO algorithm case, 0.80300 to 0.97542 (with LR) in the BFO algorithm case, and 0.80701 to 0.91567 in the hybrid scenario (with RF classifier). The closer the value is to 1.000, the better the outcome. We can see that the EPO has the best range (up to 0.97542). The results are also shown to include ROC curves for all cases, selected confusion metrics, and the computation of MCC and Kappa scores for all experiments.

The created glaucoma assessment pipeline should be tested, performed offline on mobile devices as well, and the findings presented in a short amount of time. This is one of

the limitations of this effort. The proposed method must still be tested on large-scale clinical imaging before it can be used in real-world circumstances worldwide. Future studies could include training the proposed model on a large number of varieties of images in order to improve its performance on other datasets and generalize it to all fundus images. Furthermore, as researchers are constantly proposing new optimization algorithms, in the future we will try to implement and analyze the performance of these algorithms on our problem. The same ideology can be applied to the diagnosis of various disorders such as diabetes, retinopathy, fatty liver disease, thyroid cancer, and ovarian cancer, among others.

5.2 Comparison with prior published literature

A comparison of the suggested approach to the existing state-of-the-art glaucoma prediction methods is shown in Table 17. The comparative table suggests the effectiveness of the implemented approach in glaucoma identification as compared to previous studies. This table provides irrefutable evidence that the proposed novel approach is robust and efficient in categorizing fundus images with an accuracy of up to 95.410%. The sensitivity and specificity of the performance are both high, as with other metrics. Our method (which relies on soft computing and machine learning) has been shown to outperform deep learning methods in a few scenarios. However, there may be a few instances when the datasets on which the other methods were evaluated are different from ours. We have worked on several datasets to further investigate the performance of the technique we created, as previously described, which indicates the generalizability of our suggested approach.

6 Limitations and future work directions

After 300–400 iterations (or even 200 in some experiments), fitness starts to go down. This is one of the limitations of the proposed study. Also, the results we report show that the hybrid approach is less effective than at least

Table 17 Comparison of proposed approach with the existing work

Study	Classification method	Dataset and/or No. of images	Accuracy	Sensitivity	Specificity
Abad et al. (2021)	DL-based approach	10,658 images from REFUGE, and other datasets	0.953	0.841	0.958
Acharya et al. (2017)	Local Configuration Pattern features and texton and KNN	702 images	0.957 (Number of features required—6)	0.962 (Number of features required—6)	0.937 (Number of features required—6)
Das et al. (2016)	Region growing method and watershed transformation	Images from HRF, MESSIDOR, DRIONS-DB and DIARETDB1 and the images from a local eye hospital (Sri Sankaradeva Netralaya)	0.9385	0.9259	0.9447
Elangovan et al. (2021)	DL	ORIGA	0.7832	0.5806	0.9244
Elmoufidi et al. (2021)	CNN and SVM	ORIGA	0.9667	0.9665	0.9671
Fu et al. (2019)	DL	ORIGA and Singapore Chinese Eye Study	0.8429	0.8478	0.8380
Gour and Khanna (2019)	GIST and PHOG descriptors	60 images of HRF and remaining from Drishti GS images	0.834 (Number of features required- 4 top GIST and 4 top PHOG)	Not Reported	Not Reported
Guo et al. (2018)	Convolutional neural network(CNN)	ORIGA	0.7690	0.799	0.738
Guo et al. (2020)	Gradient boosting decision tree (GBDT)	650 images of ORIGA	0.843 (Number of features required—13 to 17)	0.894 (Number of features required—13 to 17)	0.793 (Number of features required—13 to 17)
Jerith and Kumar (2020)	FS through gray wolf algorithm	Not Reported	0.9310 (Number of features required—features of 4 classes)	0.916777 (Number of features required—features of 4 classes)	0.94117 (Number of features required—features of 4 classes)
Juneja et al. (2020b)	DL-based approach; proposed a new DL model	DRISHTI –DS	0.9751	0.9878	0.9620
Kim et al. (2017)	Features from the examination of retinal nerve fiber layer (RNFL) thickness and visual field (VF). Classification by RF	500 images from local hospital	0.9800 (Number of features required—7)	0.983 (Number of features required—7)	0.979 (Number of features required—7)
Kirar et al. (2020)	Discrete wavelet transform and LS-SVM	505 images of RIM-1 dataset	0.8495 (Number of features required—26)	0.86 (Number of features required—26)	0.8385 (Number of features required—26)
Liu et al. (2019)	Deep Learning(DL)	650 images of ORIGA dataset and 400 images from REFUGE dataset	ORIGA (0.7657) REFUGE (0.8278)	ORIGA (0.7273) REFUGE (0.7)	ORIGA (0.8041) REFUGE (0.956)
Maheshwari et al. (2017)	Feature normalization, selection, and ranking; LS-SVM	488 private images	0.9519 (Number of features required—13)	0.9362 (Number of features required—13)	0.9671 (Number of features required—13)

Table 17 (continued)

Study	Classification method	Dataset and/or No. of images	Accuracy	Sensitivity	Specificity
Maheshwari et al. (2019)	Bit-plane slicing (BPS) and local binary patterns (LBP). Classification by RF	1426 images from local hospital	0.9930 (Number of features required—20)	0.9884 (Number of features required—20)	0.9964 (Number of features required—20)
Martins et al. (2020)	DL	2482 images from ORIGA and other datasets	0.88	0.83	0.90
Orlando et al. (2019)	DL	REFUGE	Not Reported	0.9752	Not Reported
Parashar and Agrawal (2020)	FS and SVM classifier	941 images of RIM-ONE + ORIGA + DRISHTI + HRF	0.9076 (Number of features required—24)	0.945 (Number of features required—24)	0.8784 (Number of features required—24)
Raghavendra et al. (2018)	SVM and 19 features (Radon + MCT + GIST + LSDA)	1000 Private and Public images	0.9700 (Number of features required—19)	0.9780 (Number of features required—19)	0.9580 (Number of features required—19)
Sreng et al. (2020)	11 DL models, then SVM and finally their ensemble	2787 images from REFUGE, ACRIMA, ORIGA and other datasets	From 0.8000 (ORIGA) to 0.9559 (REFUGE)	Not Reported	Not Reported
Tulsani et al. (2021)	DL	650 images of ORIGA	0.90	0.87	0.81
Zilly et al. (2017)	Entropy sampling and ensembles learning method based on CNN architecture	50 images of Drishti GS Dataset	0.941	0.923	0.941
Proposed method	Soft computing and machine learning approaches	3112 images from the combination of public Datasets and private images	0.9541 (Number of features required—13 Execution time—28.447)	0.934 (Number of features required—13 Execution time—28.447)	0.984 (Number of features required—6 Execution time—1.269)

one of the individual approaches for most of the performance metrics we looked at. Another limitation is that the proposed system still needs to be tested on huge datasets. That can also be a future direction for the researchers. Future work could integrate AI-provided data with all other patient data (presence of a positive family history, medical history, ethnicity, age, gender). The utilized public datasets did not contain this information. It would be extremely intriguing to conduct additional analysis using these variables to detect relationships. Future work might include generalizing the suggested model to all fundus pictures, training it on a wider range of images to increase performance on other datasets, and expanding the proposed model's use beyond glaucoma to include heart disorders, lung cancer, and Alzheimer's disease. To boost the system's performance, several features will be used, such as extraction and reduction techniques. We will also attempt

to develop another imaging technique for glaucoma detection called optical coherence tomography, or OCT. Additionally, it may be expanded for automation in the near future to increase accuracy using deep learning techniques with more images. In the future, the proposed system's performance can be measured on datasets of other chronic human diseases like COVID-19 detection, breast cancer screening, and skin cancer screening. Researchers are frequently proposing new soft computing algorithms, so the performance of these new algorithms can be evaluated in the future. Moreover, upgrading the proposed system to an IoT-integrated web-based system can be a future extension that will be extremely useful for the human race.

7 Conclusion

The key to preventing this visual degradation is early identification, which, if left untreated, might go further and result in irreversible vision loss. Early identification and diagnosis of this illness are substantially hampered by standard diagnostic techniques' laborious and lengthy approach. Moreover, the most crucial phase in the development of a glaucoma detection system is the selection of the most appropriate features. The primary purpose of the suggested approach is to reduce the size of the feature space in order to enhance the classification system's performance. This article implements three algorithms: EPO, BFO, and a hybrid of these two algorithms. To the best of our knowledge, these algorithms have only rarely been employed in this way for the detection of glaucoma. The characteristics (features) chosen by these three methods are assessed using six machine learning classifiers that have been shortlisted. The performance of the proposed technique is evaluated using benchmark fundus image datasets. In total, twelve experiments have been conducted. Typically, more than 50% of the characteristics are decreased most of the time. This decrease gets up to 86.11% (five features are restored from the original 36 features without a significant compromise in accuracy). In addition to being computationally efficient, the suggested method requires relatively little time to process and analyze high-resolution retinal pictures. Expert ophthalmologists may use the proposed method as a second opinion when diagnosing glaucoma. Consequently, the suggested method might be highly valuable for glaucoma patients' first screening. The proposed solution yields trustworthy results and is simple to implement in medical departments for successful prediction.

Author's contribution All authors contributed equally to the study conception, design and coding. Material preparation, data collection and analysis were performed by [LKS], [MK], [HG] and [RS]. All authors read and approved the final submitted manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets analyzed during the current study are easily and publically available in the internet's repository. We also confirm that data will be available on demand.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval No ethical approval required.

References

- Abad PF, Coronado-Gutierrez D, Lopez C, Burgos-Artizzu XP (2021) Glaucoma patient screening from online retinal fundus images via Artificial Intelligence. medRxiv
- Acharya UR, Bhat S, Koh JE, Bhandary SV, Adeli H (2017) A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images. *Comput Biol Med* 88:72–83
- Agrawal DK, Kirar BS, Pachori RB (2019) Automated glaucoma detection using quasi-bivariate variational mode decomposition from fundus images. *IET Image Proc* 13(13):2401–2408
- Balasubramanian K, Ananthamoorthy NP (2022) Correlation-based feature selection using bio-inspired algorithms and optimized KELM classifier for glaucoma diagnosis. *Appl Soft Comput* 128:109432
- Barani F, Mirhosseini M, Nezamabadi-Pour H (2017) Application of binary quantum-inspired gravitational search algorithm in feature subset selection. *Appl Intell* 47(2):304–318
- Cheng J, Liu J, Xu Y, Yin F, Wong DWK, Tan NM, Tao D, Cheng CY, Aung T, Wong TY (2013) Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans Med Imaging* 32(6):1019–1032
- Chen YP, Li Y, Wang G, Zheng YF, Xu Q, Fan JH, Cui XT (2017) A novel bacterial foraging optimization algorithm for feature selection. *Expert Syst Appl* 83:1–17
- Chen LH, Hsiao HD (2008) Feature selection to diagnose a business crisis by using a real GA-based support vector machine: an empirical study. *Expert Syst Appl* 35(3):1145–1155
- Da Silva SF, Ribeiro MX, Neto JDEB, Traina-Jr C, Traina AJ (2011) Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decis Support Syst* 51(4):810–820
- Dash M, Liu H (1997) Feature selection for classification. *Intell Data Anal* 1(1–4):131–156
- Das P, Nirmala SR, Medhi JP (2016) Diagnosis of glaucoma using CDR and NRR area in retina images. *Netw Model Anal Health Inform Bioinform* 5(1):1–14
- Das S, Biswas A, Dasgupta S, Abraham A (2009) Bacterial foraging optimization algorithm: theoretical foundations, analysis, and applications. *Foundations of computational intelligence*, vol 3. Springer, Berlin, Heidelberg, pp 23–55
- DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837–845
- Deperlioglu O, Kose U, Gupta D, Khanna A, Giampaolo F, Fortino G (2022) Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: analysis with doctor evaluation. *Futur Gener Comput Syst* 129:152–169
- Derrac J, García S, Herrera F (2009) A first study on the use of coevolutionary algorithms for instance and feature selection. In: *International conference on hybrid artificial intelligence systems*. Springer, Berlin, Heidelberg, 557–564
- Elangovan P, Nath MK (2021) Glaucoma assessment from color fundus images using convolutional neural network. *Int J Imaging Syst Technol* 31(2):955–971
- Elmoufidi A, Skouta A, Jai-Andaloussi S, Ouchetto O (2022) CNN with multiple inputs for automatic glaucoma assessment using fundus images. *Int J Image Graph*, 2350012
- Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. *Neurocomputing* 172:371–381
- Fu H, Cheng J, Xu Y, Liu J (2019) Glaucoma detection based on deep learning network in fundus image. In: *Deep learning and*

- convolutional neural networks for medical imaging and clinical informatics. Springer, Cham. 119–137
- Ghosh A, Datta A, Ghosh S (2013) Self-adaptive differential evolution for feature selection in hyperspectral image data. *Appl Soft Comput* 13(4):1969–1977
- Gour N, Khanna P (2020) Automated glaucoma detection using GIST and pyramid histogram of oriented gradients (PHOG) descriptors. *Pattern Recogn Lett* 137:3–11
- Gu S, Cheng R, Jin Y (2018) Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Comput* 22(3):811–822
- Guo F, Mai Y, Zhao X, Duan X, Fan Z, Zou B, Xie B (2018) Yanbao: a mobile app using the measurement of clinical parameters for glaucoma screening. *IEEE Access* 6:77414–77428
- Guo F, Li W, Tang J, Zou B, Fan Z (2020) Automated glaucoma screening method based on image segmentation and feature extraction. *Med Biol Eng Comput* 58(10):2567–2586
- Haider A, Arsalan M, Lee MB, Owais M, Mahmood T, Sultan H, Park KR (2022) Artificial Intelligence-based computer-aided diagnosis of glaucoma using retinal fundus images. *Expert Syst Appl* 207:117968
- Ibrahim RA, Ewees AA, Oliva D (2019) Improved salp swarm algorithm based on particle swarm optimization for feature selection. *J Ambient Intell Humaniz Comput* 10:3155–3169
- Jayaraman V, Sultana HP (2019) Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. *J Ambient Intell Human Comput*, 1–10
- Jerith GG, Kumar PN (2020) Recognition of Glaucoma by means of gray wolf optimized neural network. *Multimed Tools Appl* 79(15):10341–10361
- Juneja M, Thakur S, Wani A, Uniyal A, Thakur N, Jindal P (2020) DC-Gnet for detection of glaucoma in retinal fundus imaging. *Mach Vis Appl* 31(5):1–14
- Juneja M, Thakur N, Thakur S, Uniyal A, Wani A, Jindal P (2020) GC-NET for classification of glaucoma in the retinal fundus image. *Mach Vis Appl* 31(5):1–18
- Kang M, Islam MR, Kim J, Kim JM, Pecht M (2016) A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Trans Ind Electron* 63(5):3299–3310
- Kausu TR, Gopi VP, Wahid KA, Doma W, Niwas SI (2018) Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images. *Biocybern Biomed Eng* 38(2):329–341
- Ke L, Feng Z, Ren Z (2008) An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recogn Lett* 29(9):1351–1357
- Ke L, Feng Z, Xu Z, Shang K, Wang Y (2010) A multiobjective ACO algorithm for rough feature selection. In: 2010 second pacific-Asia conference on circuits, communications and system, IEEE Vol. 1, pp 207–210
- Khan SI, Choubey SB, Choubey A, Bhatt A, Naishadhkumar PV, Basha MM (2022) Automated glaucoma detection from fundus images using wavelet-based denoising and machine learning. *Concurr Eng* 30(1):103–115
- Khushaba RN, Al-Ani A, AlSukker A, Al-Jumaily A (2008) A combined ant colony and differential evolution feature selection algorithm. In: International Conference on ant colony optimization and swarm intelligence. Springer, Berlin, Heidelberg. pp 1–12
- Kim SJ, Cho KJ, Oh S (2017) Development of machine learning models for diagnosis of glaucoma. *PLoS One* 12(5):e0177726
- Kirar BS, Agrawal DK, Kirar S (2020) Glaucoma detection using image channels and discrete wavelet transform. *IETE J Res*, 1–8
- Kirar BS, Agrawal DK (2018) Glaucoma diagnosis using discrete wavelet transform and histogram features from fundus image. *Int J Eng Technol* 7(4):2546–2551
- Kirar BS, Agrawal DK (2019) Computer aided diagnosis of glaucoma using discrete and empirical wavelet transform from fundus images. *IET Image Proc* 13(1):73–82
- Kolář R, Jan J (2008) Detection of glaucomatous eye via color fundus images using fractal dimensions. *Radioengineering* 17(3):109–114
- Krishnan MMR, Faust O (2013) Automated glaucoma detection using hybrid feature extraction in retinal fundus images. *J Mech Med Biol* 13(01):1350011
- Lane MC, Xue B, Liu I, Zhang M (2013) Particle swarm optimisation and statistical clustering for feature selection. In: Australasian Joint Conference on Artificial Intelligence. Springer, Cham. pp 214–220
- Liu Y, Tang F, Zeng Z (2014) Feature selection based on dependency margin. *IEEE Trans Cybern* 45(6):1209–1221
- Liu S, Hong J, Lu X, Jia X, Lin Z, Zhou Y, Liu Y, Zhang H (2019) Joint optic disc and cup segmentation using semi-supervised conditional GANs. *Comput Biol Med* 115:103485
- Mafarja MM, Mirjalili S (2017) Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* 260:302–312
- Mafarja M, Aljarah I, Heidari AA, Faris H, Fournier-Viger P, Li X, Mirjalili S (2018) Binary dragonfly optimization for feature selection using time-varying transfer functions. *Knowl-Based Syst* 161:185–204
- Mafarja M, Aljarah I, Faris H, Hammouri AI, AlaM AZ, Mirjalili S (2019) Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Syst Appl* 117:267–286
- Maheshwari S, Pachori RB, Acharya UR (2016) Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. *IEEE J Biomed Health Inform* 21(3):803–813
- Maheshwari S, Pachori RB, Kanhangad V, Bhandary SV, Acharya UR (2017) Iterative variational mode decomposition based automated detection of glaucoma using fundus images. *Comput Biol Med* 88:142–149
- Maheshwari S, Kanhangad V, Pachori RB, Bhandary SV, Acharya UR (2019) Automated glaucoma diagnosis using bit-plane slicing and local binary pattern techniques. *Comput Biol Med* 105:72–80
- Mrad Y, Elloumi Y, Akil M, Bedoui MH (2022) A fast and accurate method for glaucoma screening from smartphone-captured fundus images. *IRBM* 43(4):279–289
- Martins J, Cardoso JS, Soares F (2020) Offline computer-aided diagnosis for Glaucoma detection using fundus images targeted at mobile devices. *Comput Methods Programs Biomed* 192:105341
- Muni DP, Pal NR, Das J (2006) Genetic programming for simultaneous feature selection and classifier design. *IEEE Trans Syst Man Cybern Part B (Cybern)* 36(1):106–117
- Nematzadeh H, Enayatifar R, Mahmud M, Akbari E (2019) Frequency based feature selection method using whale algorithm. *Genomics* 111(6):1946–1955. <https://doi.org/10.1016/j.ygeno.2019.01.006>
- Neshatian K, Zhang M (2009) Dimensionality reduction in face detection: a genetic programming approach. In: 2009 24th International Conference Image and Vision Computing New Zealand. IEEE. pp 391–396
- Nyúl LG (2009) Retinal image analysis for automated glaucoma risk evaluation. In: MIPPR 2009: medical imaging, parallel processing of images, and optimization techniques. SPIE. Vol. 7497, pp 332–340

- O'Boyle NM, Palmer DS, Nigsch F, Mitchell JB (2008) Simultaneous feature selection and parameter optimisation using an artificial ant colony: case study of melting point prediction. *Chem Cent J* 2(1):1–15
- Orlando JI, Fu H, Breda JB, van Keer K, Bathula DR, Diaz-Pinto A et al (2019) REFUGE Challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal* 2020(59):101570
- Pandey AC, Kulhari A (2018) Semi-supervised spatiotemporal classification and trend analysis of satellite images. In: *Advances in Computer and Computational Sciences*. Springer, Singapore. pp 353–363
- Parashar D, Agrawal DK (2020) Automated classification of glaucoma stages using flexible analytic wavelet transform from retinal fundus images. *IEEE Sens J* 20(21):12885–12894
- Pashaei E, Aydin N (2017) Binary black hole algorithm for feature selection and classification on biological data. *Appl Soft Comput* 56:94–106
- Prabukumar M, Agilandeeswari L, Ganesan K (2019) An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. *J Ambient Intell Humaniz Comput* 10(1):267–293
- Raghavendra U, Bhandary SV, Gudigar A, Acharya UR (2018) Novel expert system for glaucoma identification using non-parametric spatial envelope energy spectrum with fundus images. *Biocybern Biomed Eng* 38(1):170–180
- Raja C, Gangatharan N (2013) Glaucoma detection in fundal retinal images using trispectrum and complex wavelet-based features. *Eur J Sci Res* 97(1):159–171
- Rodrigues D, Pereira LA, Nakamura RY, Costa KA, Yang XS, Souza AN, Papa JP (2014) A wrapper approach for feature selection based on bat algorithm and optimum-path forest. *Expert Syst Appl* 41(5):2250–2258
- Rodrigues D, Pereira LA, Almeida TNS, Papa JP, Souza AN, Ramos CC, Yang XS (2013) BCS: a binary cuckoo search algorithm for feature selection. In: *2013 IEEE International symposium on circuits and systems (ISCAS)*. IEEE. pp 465–468
- Saraswat M, Arya KV (2014) Feature selection and classification of leukocytes using random forest. *Med Biol Eng Comput* 52(12):1041–1052
- Sayed GI, Khoriba G, Haggag MH (2018) A novel chaotic salp swarm algorithm for global optimization and feature selection. *Appl Intell* 48(10):3462–3481
- Sayed GI, Hassanien AE, Azar AT (2019) Feature selection via a novel chaotic crow search algorithm. *Neural Comput Appl* 31(1):171–188
- Shanmugam P, Raja J, Pitchai R (2021) An automatic recognition of glaucoma in fundus images using deep learning and random forest classifier. *Appl Soft Comput* 109:107512
- Shunmugapriya P, Kanmani S (2017) A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm Evol Comput* 36:27–36
- Singh A, Dutta MK, ParthaSarathi M, Uher V, Burget R (2016) Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. *Comput Methods Programs Biomed* 124:108–120
- Singh LK, Garg H, Khanna M, Bhadoria RS (2021) An enhanced deep image model for glaucoma diagnosis using feature-based detection in retinal fundus. *Med Biol Eng Comput* 59(2):333–353
- Singh LK, Khanna M, Thawkar S, Singh R (2022) Collaboration of features optimization techniques for the effective diagnosis of glaucoma in retinal fundus images. *Adv Eng Softw* 173:103283
- Singh LK, Khanna M (2022) A novel multimodality based dual fusion integrated approach for efficient and early prediction of glaucoma. *Biomed Signal Process Control* 73:103468
- Singh LK, Khanna M, Thawkar S (2022) A novel hybrid robust architecture for automatic screening of glaucoma using fundus photos, built on feature selection and machine learning-nature driven computing. *Expert Syst* 39:e13069
- Sreng S, Maneerat N, Hamamoto K, Win KY (2020) Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Appl Sci* 10(14):4916
- Sun X, Xu W (2014) Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett* 21(11):1389–1393
- Tang B, Kay S, He H (2016) Toward optimal feature selection in naive Bayes for text categorization. *IEEE Trans Knowl Data Eng* 28(9):2508–2521
- Tang K, Yáo X, Suganthan PN, MacNish C, Chen YP, Chen CM, Yang Z (2007) Benchmark functions for the CEC'2008 special session and competition on large scale global optimization. *Nat Inspired Comput Appl Lab USTC China* 24:1–18
- Tulsani A, Kumar P, Pathan S (2021) Automated segmentation of optic disc and optic cup for glaucoma assessment using improved UNET++ architecture. *Biocybern Biomed Eng* 41:819–832
- Vieira SM, Sousa JM, Runkler TA (2010) Two cooperative ant colonies for feature selection using fuzzy models. *Expert Syst Appl* 37(4):2714–2723
- Wei J, Zhang R, Yu Z, Hu R, Tang J, Gui C, Yuan Y (2017) A BPSO-SVM algorithm based on memory renewal and enhanced mutation mechanisms for feature selection. *Appl Soft Comput* 58:176–192
- Winkler SM, Affenzeller M, Jacak W, Stekel H (2011) Identification of cancer diagnosis estimation models using evolutionary algorithms: a case study for breast cancer, melanoma, and cancer in the respiratory system. In: *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation* pp 503–510
- Wu Y, Liu B, Wu W, Lin Y, Yang C, Wang M (2018) Grading glioma by radiomics with feature selection based on mutual information. *J Ambient Intell Humaniz Comput* 9(5):1671–1682
- Xue B, Zhang M, Browne WN, Yao X (2015) A survey on evolutionary computation approaches to feature selection. *IEEE Trans Evol Comput* 20(4):606–626
- Yadav D, Sarathi MP, Dutta MK (2014) Classification of glaucoma based on texture features using neural networks. In: *2014 seventh international conference on contemporary computing (IC3)*. IEEE. pp 109–112
- Yang J, Honavar V (1998) Feature subset selection using a genetic algorithm. In: *Feature extraction, construction and selection*. Springer, Boston, MA. pp 117–136
- Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: *2009 World congress on nature & biologically inspired computing (NaBIC)*. IEEE. pp 210–214
- Zilly J, Buhmann JM, Mahapatra D (2017) Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Comput Med Imaging Graph* 55:28–41

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.