IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 December 2023, accepted 6 January 2024, date of publication 9 January 2024, date of current version 24 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351871

RESEARCH ARTICLE

Hybrid Ant Lion Mutated Ant Colony Optimizer Technique With Particle Swarm Optimization for Leukemia Prediction Using Microarray Gene Data

T. R. MAHESH[1], (Senior Member, IEEE), D. SANTHAKUMAR[2], A. BALAJEE[1], H. S. SHREENIDHI[1], V. VINOTH KUMAR[3], (Member, IEEE), AND JONNAKUTI RAJKUMAR ANNAND[4]

[1]Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru 562112, India
[2]Saveetha School of Engineering, SIMATS, Chennai 602107, India
[3]School of Computer Science Engineering & Information Systems (SCORE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India
[4]Department of Electromechanical Engineering, Sawla Campus, Arba Minch University, Arba Minch 4400, Ethiopia

Corresponding author: Jonnakuti Rajkumar Annand (jonnakuti.rajkumar@amu.edu.et)

ABSTRACT Leukemia refers to a type of blood malignancy that develops due to certain hematological disorders. Identifying leukemia at its earlier stages through clinical operations are highly complicated task with invasive methods. Gene expression data could be collected and computational methods could be adopted which could lead to better prediction of leukemia that leads to prevention at its earlier stages. Today, feature selection has become an important step in pre-processing that helps bring improvement to the classification system and its performance that is done by choosing optimal feature subsets by means of reducing or eliminating redundant or irrelevant features. Particle Swarm Optimization (PSO) is a popular algorithm wherein certain solutions that are generated randomly move within the search space to obtain optimal solutions. Another relatively new and evolutionary method computation is the Ant Lion Optimization (ALO) algorithm that has lower computation cost compared to the other techniques. In this work, a new technique known as the Hybrid Ant Lion Mutated Ant Colony Optimize along with Particle Swarm Optimization (PSO) was proposed for the prediction of leukaemia with the microarray gene data. The proposed model that is used for identifying the optimal set of features from which the classification has been done using the Support Vector Machine (SVM) has produced a significant prediction accuracy of 87.88%.

INDEX TERMS Leukemia, gene expression data, feature selection, ant lion optimization (ALO) algorithm, evolutionary computation.

I. INTRODUCTION

Leukemia is the only type of blood cancer that is a result of haematological disorders and if leukaemia will ensue in the lymphocyte found in the bone marrow that is called the acute lymphoblastic leukaemia. At the same time, if a severe disorder follows within the bone marrow cells, platelets, or red blood cells and is known as acute myeloid leukemia. There are thousands of individuals are diagnosed of leukaemia that is evidenced as a mortal cancer among all different types of cancers. Both identification and classification have become explicitly important as the treatment will vary based on the subtype of leukaemia. This has been diagnosed on the basis of an examination of the test sample that recognizes all abnormalities of chromosome and various markers of cells to acknowledge other types of leukaemia. Imaging tests like Magnetic Resonance Imaging (MRI) are used.

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Kashif Bashir.

The Microarray Gene Expression Technology has today unlocked various means of examining the characteristics of many thousands of genes in a simultaneous manner. The gene expression profiles represent the profusion of measurement of the mRNA conforming to the genes [1]. Therefore, discriminant analysis of such microarray data will have a major advantage as a tool for medical diagnosis. The microarray data can be classified for building an effective and efficient model to recognize the genes that are expressed differently and can be used for prediction of class for the unknown samples. The primary difficulty faced in the classification are the small size of samples compared to the gene expression levels that are measured. There are a classifier design and a feature selection that is included in classifying gene expression data. Normally, there is only a small set of genes that have a robust association to a particular phenotype when compared to the total number of genes that are studied. So, for analysing the profiles of gene expression in the right manner, feature (gene) selection is important.

There are many challenges that are faced during classifying the microarray data. They are the small set of samples that are related to the high levels of dimensionality with disparities of experiments among the gene expression levels that are evaluated. Most commonly, there are only small numbers of genes that have a close association with a certain phenotype that corresponds to the genes examined. During analysing of the profiles of gene expressions in the right way, feature (gene) selection becomes important. The need for a trade-off between the performance of the algorithms for the NP based problems was identified at the early stages with the feature selection study that was focused on certain efficient and robust suboptimal methods. The primary goal of feature selection was the identification of the subset in differentially expressed genes relevant to distinguish sample classes. An ideal method to choose the genes that are relevant to the classification of samples is necessary for the accelerating rate of processing, decreasing its predictive error rate, and avoiding the incomprehensibility of spurious data correlations. This is based on the genes investigated.

Different methods are used for simulating the physical and biological systems for which powerful global optimization techniques have been proposed. The EC algorithms were motivated by the biological or social behaviour (of birds, animals, fish, antlions, bats, spiders, wolves, and firefly) found in a pack [2]. Different researchers have been suggesting several methods of computation that imitated the behaviour of such animals to investigate their optimal solutions. The EC algorithms have been applied in order to search for an optimal solution with a set of simple entities which can communicate socially which will adaptively search the space. Particle swarm optimization (PSO) [3] refers to a stochastic population-based strategy of optimization that was stimulated based on the social behaviour shown by birds in searching for optimal paths. This is a meta-heuristic model wherein the swarm (the set of particles) navigate within the search space using some velocity for obtaining the best solution set. Each particle will specify a certain solution to the problem of optimization. For this, the TS algorithm will be a new process of metaheuristic to accept the initial solution as its input by making use of the memory structures until such time the stopping criterion is met.

The primary motivation behind the hybridization of various algorithmic concepts was for obtaining better performing systems to exploit the advantages of pure strategies as hybrids benefit only from synergy. The selection of the right combination of several algorithmic concepts can be the key for achieving top performance and solving several problems of optimization. In this work, a hybrid technique that is based on the Ant Lion Optimization (ALO) with PSO has been proposed. The uniqueness of the proposed feature selection mechanism is that it adopts optimization strategies to identify the crisp set of features from the dataset which has higher dimensional space. It is essential for the selected features must not only be a crisp set but also an optimized one which can efficiently identify the target variable. The remainder of the study has been organised as follows: All review of available work was explained in Section II. The different techniques employed were discussed in Section III. Section IV presents the results and the paper is concluded in Section V.

II. RELATED WORK

For the performance of clustering, informative gene selection was used to discover certain useful phenotypes and this can be a major issue since there is not information on a class that is available. Deepthi and Thampi [4] had proposed a new wrapper-based feature selection technique for the performance of a sample-based clustering made on gene expression data. This work makes use of the PSO for proper subset selection with the k-means as the wrapper algorithm to evaluate subsets. The experiments proved that the chosen features could produce good quality clusters. The accuracy of clustering was about 70 to 80% for various datasets. An Improved Binary PSO (IBPSO) was used in Chuang et al. [5] for implementing feature selection with the K-nearest neighbor (K-NN) method that was used for evaluation of the IBPSO used for the problems of identification of gene expression data. Experimental results demonstrated that the method was able to simplify feature selection to lessen the total number of features that are required. The accuracy of classification that was obtained by the method proposed was higher in 9 out of 11 test problems of gene expression data and can be compared to the accuracy of classification of other problems.

Dutta et al. [6] had reported regarding another automated approach that measured the actual degree of relevance of genes to predict pathway activity. Since there is a larger search space that had to be explored, the properties of exploration of the PSO was used in this context. The PSO particles represent various scores of relevance for their member genes belonging to different pathways.

For dealing with this relevance-score, another popular t-score used widely to measure pathway activity was expanded and was known as the weighted t-score. The PSO-based weighted framework proposed was further assessed based on three different gene expression data sets. The top 50% of the pathway markers were chosen for every dataset with the quality of the measures duly checked was performed with respect to various measures of quality. The results were evaluated by making use of biological significance tests.

Zawbaa et al. [7] had proposed another model used for feature selection that was based on ALO. Normally, feature sets have a dependent and redundant set of correlated features that affect the performance of classification to the increase time taken for training. Thus, feature selection is critical for removing irrelevant features and to enhance generalization of classification. The Wrapper-based feature selection refers to a method that chooses feature sets that maximizes classifier performance criteria therefore needed an efficient method of search to identify optimal feature combinations. Recently, an Ant Lion Optimization has been proposed with good capability of search. The ALO was exploited with a good capability of search. The ALO was exploited by this study in the form of a method of search that finds optimal feature sets that maximize performance of classification. The ALO algorithm further imitates a hunting behaviour of the antlions in nature. This model was evaluated with various evaluation criteria based on 18 different datasets that was compared to two different methods of search which are the Genetic Algorithm (GA) and the PSO. Experimental results demonstrated the efficiency of the proposed ALO. Apart from feature selection, feature extraction mechanisms are also implemented directly towards the gene expression data by applying the deep learning models which demand the data with a large number of samples stated Vinoth Kumar et al. [18]. Hybridization can be implemented to the feature selection mechanism which in turn could produce the better feature vector stated by Balajee et al. [19]. Rupara et al. [20] proposed a hybrid feature selection mechanism that integrates resampling and statistical tests to calculate the set of features that could lead to better prediction. Wahid and Banday [21] implemented a hybridization of feature selection and classification mechanism for leukemia prediction using the microarray gene data where feature selection is carried through the Artificial Bee Colony optimization and classification is performed through CNN which has achieved better accuracy. Ilyas et al. [22] implemented linear programming to perform feature selection and classification from the gene expression data to predict leukemia disease. Linear programming utilizes the statistical method as a step-by-step process to perform the computational methods for the available data. Machine learning models can be built by computing the key features based on the characteristics of the data considered for analysis stated by Rafi et al. [23]. Alphonse et al. [24] stated that features can be extracted directly from the data by analyzing the different domains based on the nature of the samples considered. Once it is extracted it must be optimized to remove irrelevant attributes that could lead to misclassification. Fuzzy-based approaches could rank the features based on their effectiveness towards the target variable which could optimize the feature selection based on the ranking attained through each of those attributes stated by Nasir et al. [25]. From the various literature gone through it is evitable that to handle the clinical data two approaches are widely adopted. One is to extract the features from the raw data and the optimization techniques will be annexed along with those features to eliminate the outliers. An another approach is to create an optimization strategy that could perform feature selection from the feature vector based on its probability towards the target variable. We had adopted the second approach since in recent researches, the available feature vectors are used directly without analyzing the importance of features towards the target outcome which in turn could reduce the performance of the classification.

III. METHODOLOGY

Feature selection refers to a problem of a global combinatorial optimization observed in machine learning to reduce features and remove redundant, noisy, and irrelevant data. To search for an optimal feature subset was called the NP-complete problem. Normally, the FS algorithms include a random or heuristic search to avoid prohibitive complexity. However, the degree of optimality of its final feature subset is generally reduced. The objectives of such feature selection are varied. The most important among them were improving model performance, providing faster and cost-effective models, and finally, obtaining a new and profound insight into the process that generates this data. The Leukaemia Dataset, Tabu Feature Selection, PSO feature selection, and Hybrid PSO-ALO are described in this section. The final classification is done through Support Vector Machine (SVM) classifier since the data was now linearly separable due to the optimized feature set computed. The workflow of the proposed methodology is given in figure 1.

A. LEUKAEMIA DATASET

The dataset of leukaemia had been obtained with samples taken from the patients reported as in Golub et.al. (1999) [8]. The dataset served as a standard for the methods of microarray classification investigation. The Acute Lymphoblast Leukaemia (ALL) and the Acute Myeloid Leukaemia (AML) samples obtained from peripheral blood and bone marrow is included in the dataset. This dataset had 72 samples among which 49 samples were of ALL and 23 samples were of AML. Every sample has over a total of 7,129 genes.

B. PARTICLE SWARM OPTIMIZATION (PSO) FEATURE SELECTION

PSO [9] refers to a stochastic population-based optimization approach based on the behaviour of the birds while searching

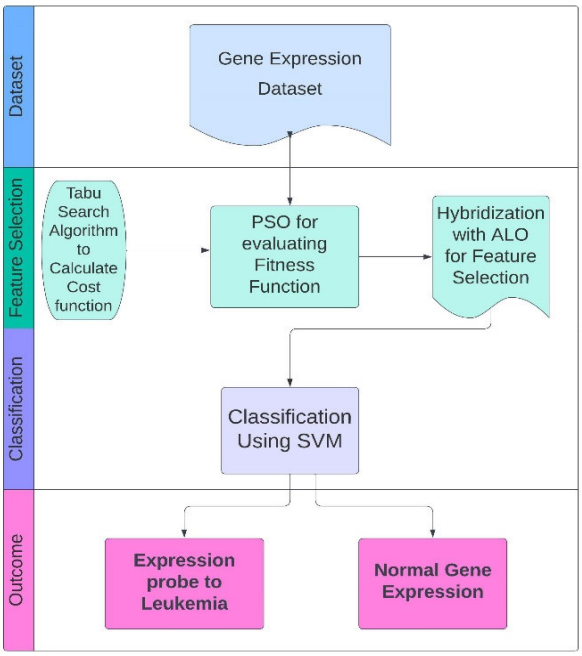

FIGURE 1. Workflow of the proposed methodology.

for optimal paths. This is a metaheuristic method where the swarm of a particles navigate within the search space in a certain velocity to meet a set of solutions. Each particle here will specify a certain solution to the problem of optimization. For the binary PSO, the position vector of each particle has been represented using a binary value which is either 0 or 1. By means of certain successive generations, these particles will update their positions and will move to the best solution found within the search space. The PSO algorithm is depicted in [4]. The objective function of each particle has been evaluated and duly stored. Fitness value for any optimum particle is known as the pBest. At the time all the populations are generated, the best value is taken into consideration among the population and that best value is known as the gBest.

Algorithm 1 PSO Algorithm

1: Generate particles randomly; each particle represents a subset of features
2: Randomly assign initial position and velocity to particles
3: Calculate fitness value
4: pbest, pbestloc of each particle identified using initial fitness value
5: update pbest, pbestloc of the subset
6: Find gbest, gbestloc
7: Calculate new velocity
8: Update location of subset
9: Calculate fitness value using current location
10: Termination - Maximum number of iterations
11: Output gbestloc as best subset of genes

C. TABU FEATURE SELECTION

Tabu search was for the first time proposed by Glover in the year 1986 as a method of meta-heuristic optimization to solve problems of hybrid optimization that were constructed using the local search algorithm for overcoming its various failings. The structure of such a TABU search has been described as given below: For achieving an optimal solution to a problem of optimization, the TABU search will move from its initial solution [10].

It will choose the best neighbour solution from among the neighbours of its current solution. In case, if the Tabu list does not include the solution, the algorithm proceeds to the neighbour solution. If not, it checks the criterion of aspiration. The flowchart for the TABU search algorithm is as in Figure 1 [11]. On the basis of the aspiration criterion, in case the neighbour solution is found to be better compared to the best one, the algorithm moves to the solution even if it is not available on the TABU list. Once this is done, there is an update done to the TABU list. This list proves to be the core to the Tabu Search through which it is prohibited from moving into the local optimum. A certain number of moves are then placed to the TABU list and this is decided by the TABU tenure parameter. A move from its current solution to that of the neighbor solution will continue until such time a stop criterion has been met. Various stop criteria are used with the algorithm. For instance, the actual number of moves to a neighbor solution can also be a stop criterion.

D. PROPOSED PARTICLE SWARM OPTIMIZATION-ANT LION FEATURE SELECTION

The inspiration for the ALO algorithm was the intelligent behaviour of hunting of the antlions of their prey, the ants. The primary steps of hunting of the antlions have been modelled mathematically in this algorithm. Here, the ants are the search agents and they move throughout the solution space. The antlions are permitted to hunt them to get fitter. For every iteration, the position of the ant has been updated in connection to the chosen ant based on the elite and the roulette wheel (the best one obtained until now). The solutions having better fitness functions are chosen for better hunting.

For the ALO algorithm, all initial positions of the antlions and the ants have been randomly initialized and its fitness values have been computed and thus the elite antlion is chosen. For each iteration, an ant lion is chosen by its roulette wheel operator and its position is duly updated using a random walk around its roulette selected antlion and the elite. All new positions are assessed using fitness values. furthermore, the elite is updated and the best antlion in the current iteration gets fitter than the elite. The same steps are repeated.

The ALO and its mathematical model [12] is described below. The ants move in a stochastic manner looking for food and thus a random walk for every ant for every step in the process of optimization can be defined

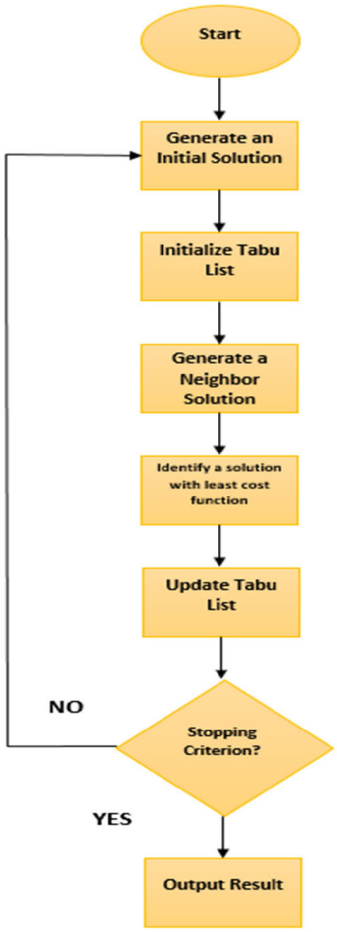

FIGURE 2. Flowchart of general tabu search algorithm.

in equation (1):

$$X_i = [0; r(1); r(1) + r(2); \ldots .; \sum_{j=1}^{T-1} r(j); \sum_{j=1}^{T} r(j)] \quad (1)$$

Wherein, i = 1, …, dim, dim refers to the ant or the antlion dimension, T the maximum number of such iterations, $X = [X_1; \ldots ; X_{dim}]$, X_i is a (T + 1) × 1 matrix, and r(j) the stochastic function as expressed in equation (2):

$$r = \begin{cases} 1 & \text{if rand} > 0.5 \\ -1 & \text{if rand} \leq 0.5 \end{cases} \quad (2)$$

Wherein, rand refers to the generated random number with a uniform distribution of [0, 1]. The random walks of the ants should be changed to the actual search space position based on the lower and the upper boundary as in equation (3):

$$Y_i = \left(\frac{X_i - a_i}{b_i - a_i} \right) \times (d_i - c_i) + c_i \quad (3)$$

Wherein, a_i and b_i are the minimum and the maximum of X_i, c_i, and d_i is the minimum and the maximum of antlion of the ith dimension. $Y = [Y_1; \ldots ; Y_{dim}]$, Y_i are the (T + 1) ×1 matrix. X_i will be normalized in domain [0, 1] with $\frac{X_i - a_i}{b_i - a_i}$. After this, it is changed into the domain $[c_i, d_i]$ and the position of the chosen antlion. The movement of the ant is affected by traps and is shown in equations (4) and (5).

$$c' + Antlion \quad (4)$$

$$d = d' + Antlion \quad (5)$$

Wherein, c' and d' refer to the minimum and the maximum of the changing limit at its current iteration. The position of the antlion is chosen using the Roulette wheel based on its fitness. The chances of the antlion building traps can be proportional to their level of fitness. As soon as the antlions are aware of the fact that ants were trapped and have attempted escape, the ants conduct their sliding process. c' and d' are updated as per equations (6) and (7).

$$c' = \frac{lb}{10^w * (\frac{t}{T})} \quad (6)$$

$$d' = \frac{ub}{10^w * (\frac{t}{T})} \quad (7)$$

Wherein, t will be the current iteration, *lb* and *ub* will be the upper limit and the lower limit. *w*refers to the constant that is defined on the basis of its current iteration. It may be easy to identify if *Y* is a *(T + 1)×dim matrix* that is computed. Aside from this, the elitism that is implemented in the ALO algorithm means the best of the antlions will be chosen to be elite for the entire process of optimization. The update of the position for each of the ants will be dependent on their random walks around antlions selected by the Roulette wheel and also their elites. This is observed as per equation (8):

$$Ant = \frac{R_A + R_E}{2} \quad (8)$$

wherein, the Ant refers to the new position, R_A the random walk around an antlion chosen by the Roulette wheel, R_E refers to the random walk made around the elite. The ant's new position is modified in case it is beyond its boundary. In case the ant is able to reach the bottom point of the pit, it will be fitter than the antlion and takes its position. This is called catching prey as shown in equation (9):

$$Antlion - Ant, \text{ if f(Ant)<f(Antlion)} \quad (9)$$

Wherein, f(·) refers to the fitness function.

The flowchart for the Hybrid ALOPSO algorithm has been depicted in Figure 3.

The proposed hybrid ALO with PSO is explained in figure 3. These parameters of the PSO were initialized and fitness for the particles was evaluated. The velocity of the particle has performed a vital role in the management of particles to traverse this within the search space for it to become narrow to the target problems and they are made by modernizing the position. Optimal feature sets are chosen on the basis of their global best positions.

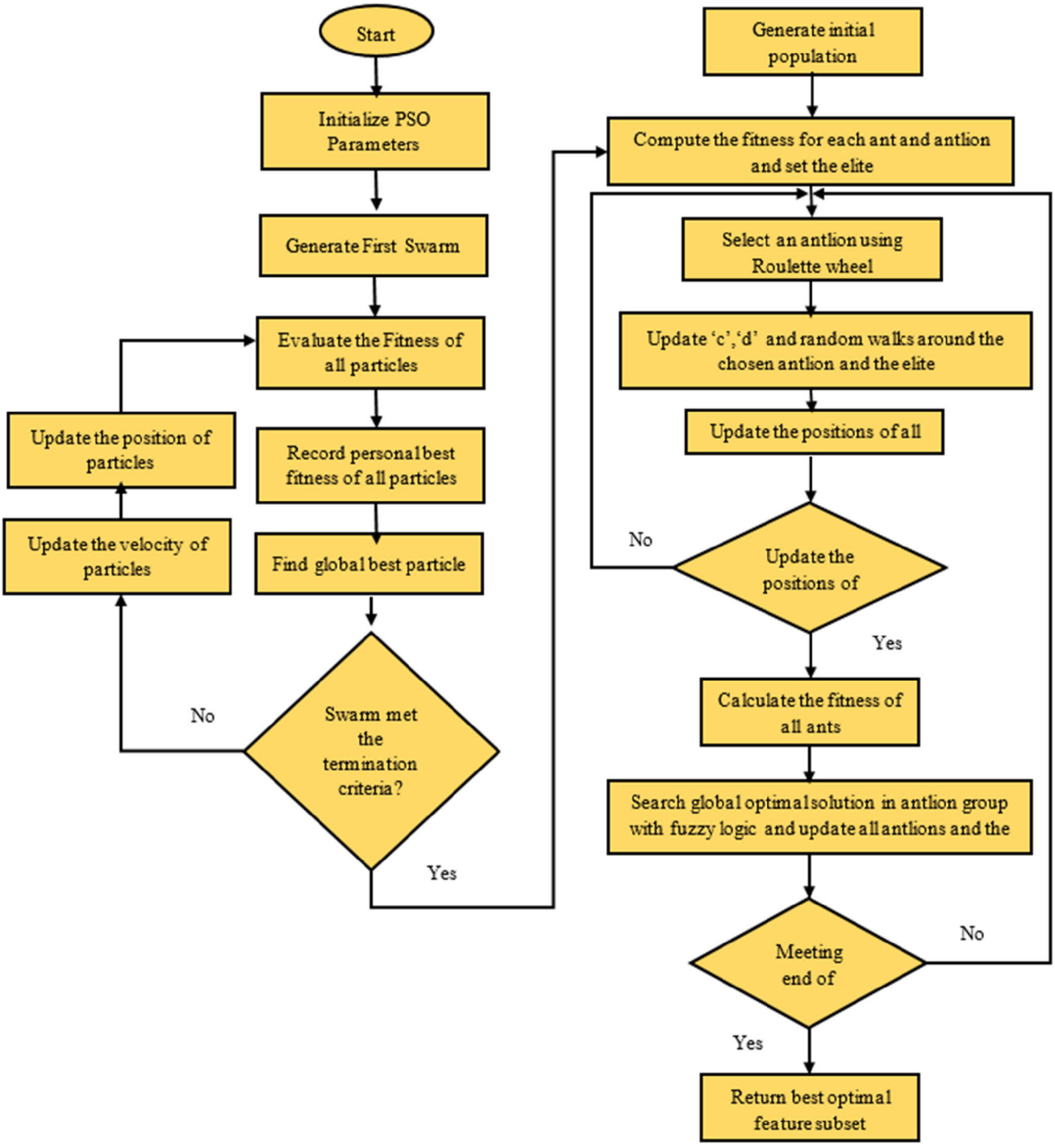

FIGURE 3. Proposed hybrid ant lion optimizer with particle swarm optimization.

For the ALO, the position of both ants and antlions is randomly initialized and fitness functions were measured. After this, the elite antlion will be defined. The ants and their new positions were accessed by measuring fitness functions and further connecting them to the antlions. In case the ant gets fitter, the position is recognized to be a new one for the subsequent iteration. Further, the elite will be updated in case the best antlion for the current iteration is found to be fitter compared to the elite.

IV. RESULTS AND DISCUSSIONS

The leukemia dataset is used for evaluating the proposed techniques. In this section, the accuracy, sensitivity, specificity, and f measure achieved by PSO feature Selection,

Tabu Feature Selection, and PSO-Ant Lion Feature Selection are presented. When anything is classified using binary classification, the output falls into one of four groups: True positive (TP), True negative (TN), False positive (FP), and False Negative (FN).

A classifier's sensitivity is the ratio of how many were accurately recognized as positive to how many were truly positive [13]. Sensitivity is also called recall. They are used where the classification of positives is a high priority. Sensitivity is computed using equation (10).

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

A classifier's specificity is the ratio of how much was accurately labeled as negative to how much was truly negative [14]. They are used where the classification of negatives is a high priority.

$$Specificity = \frac{TP}{TP + FN} \tag{11}$$

Precision is the value that is correctly classified as positive out of all positives [15] and is calculated using equation (12).

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

The proximity of a measured value to a standard or known value is referred to as accuracy [16] and it is computed using the equation (13).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{13}$$

TABLE 1. Summary of results.

Performance metrics	PSO feature Selection	Tabu Feature Selection	PSO-Ant Lion Feature Selection
Accuracy	83.33	80.3	87.88
Sensitivity -ALL	0.8636	0.8182	0.8636
Sensitivity -MLL	0.7222	0.7778	0.8333
Sensitivity -AML	0.8846	0.8077	0.9231
Specificity-ALL	0.9	0.8974	0.9286
Specificity-MLL	0.9333	0.8864	0.9556
Specificity-AML	0.8889	0.8889	0.9189
F- Measure - ALL	0.8444	0.8182	0.8636
F- Measure -MLL	0.7647	0.7567	0.8571
F- Measure-AML	0.8679	0.8235	0.9057

The harmonic mean of precision and recall gives a score called f1 score which is a measure of the performance of the model's classification ability [17]. The F score is thought to be a stronger predictor of classifier performance than the usual accuracy metric and it is calculated using equation (14).

$$Fscore = \frac{2 * (precision * sensitivity)}{(precision + sensitivity)} \tag{14}$$

A summary of the results is shown in Table 1. The accuracy, sensitivity, specificity, and f measure as shown in Figures 4 to 7. Figure 8 and 9 shows the fitness function for PSO and Proposed PSO-ALO.

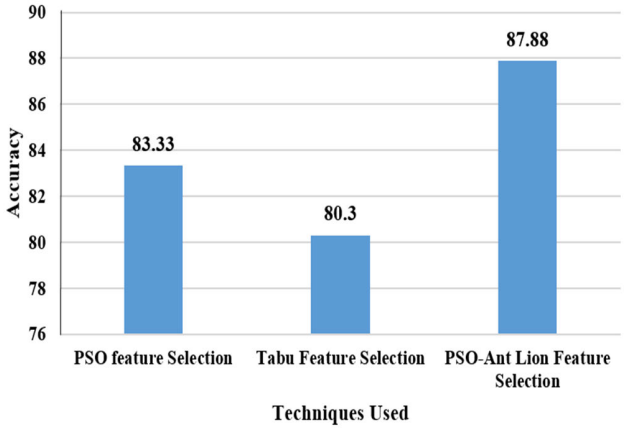

FIGURE 4. Accuracy for proposed PSO-Ant lion feature selection.

From Figure 4, it is observed that the Accuracy of the Proposed PSO-Ant Lion Feature Selection performs better by 5.32% and by 9.01% than the PSO Feature selection and Tabu Search feature selection respectively.

From Figure 5, it is observed that the sensitivity of Proposed PSO-Ant Lion Feature Selection performs better by no change and by 5.4% than PSO Feature selection and Tabu Search feature selection respectively for Sensitivity-ALL. The sensitivity of the Proposed PSO-Ant Lion Feature Selection performs better by 14.3% and by 6.9% than the PSO Feature selection and Tabu Search feature selection respectively for Sensitivity-MLL. The sensitivity of the Proposed PSO-Ant Lion Feature Selection performs better by 4.3% and by 13.3% than the PSO Feature selection and Tabu Search feature selection respectively for Sensitivity-AML.

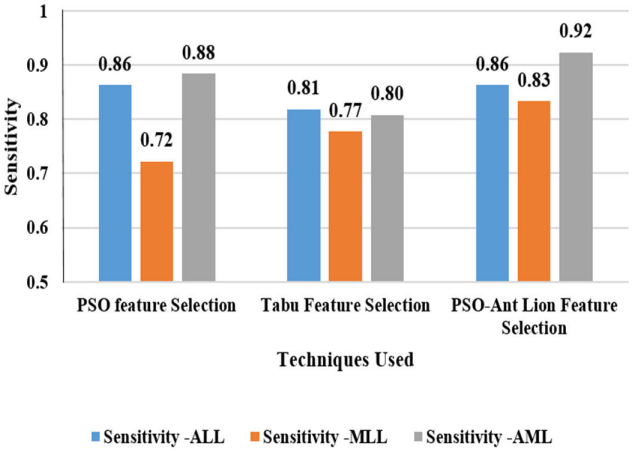

FIGURE 5. Sensitivity for proposed PSO-Ant lion feature selection.

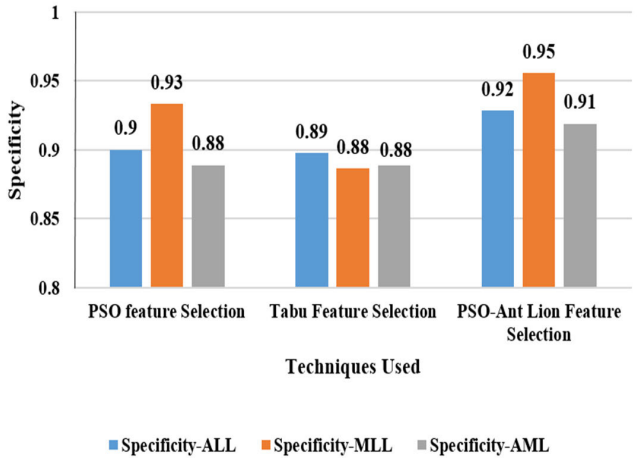

FIGURE 6. Specificity for proposed PSO-Ant lion feature selection.

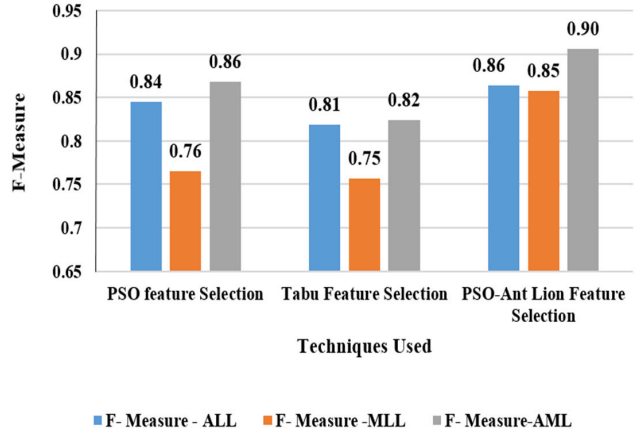

FIGURE 7. F Measure for proposed PSO-Ant lion feature selection.

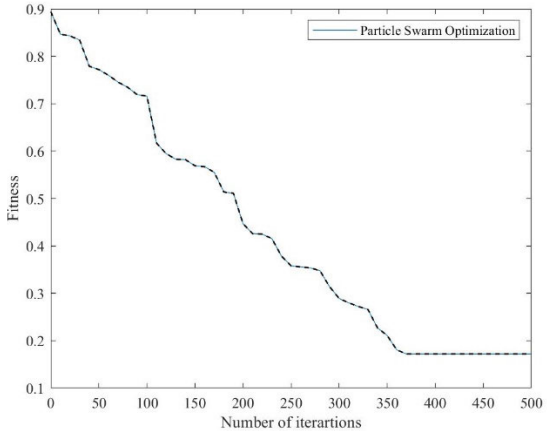

FIGURE 8. Fitness function for PSO.

From figure 6 it is observed that the specificity of Proposed PSO-Ant Lion Feature Selection performs better by 3.13% and by 3.42% than PSO Feature selection and Tabu Search feature selection respectively for specificity -ALL. The specificity of Proposed PSO-Ant Lion Feature Selection performs better by 2.4% and by 7.5% than PSO Feature selection and Tabu Search feature selection respectively for specificity -MLL. The specificity of the Proposed PSO-Ant Lion Feature Selection performs better by 3.32% and by 3.32% than the PSO Feature selection and Tabu Search feature selection respectively for specificity -AML.

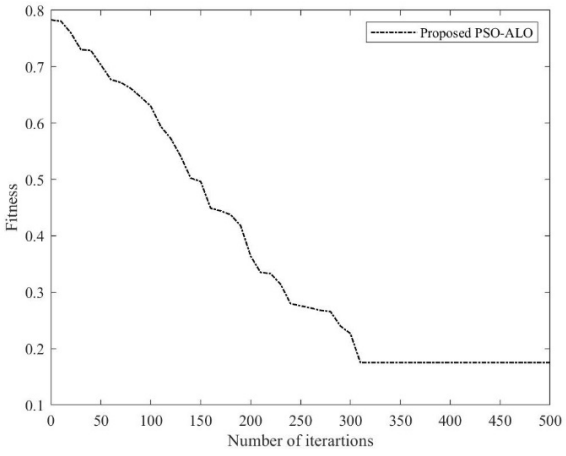

FIGURE 9. Fitness function for proposed PSO-ALO.

From figure 6 it is observed that the F Measure of Proposed PSO-Ant Lion Feature Selection performs better by 2.25% and by 5.4% than PSO Feature selection and Tabu Search feature selection respectively for F Measure -ALL. The F Measure of Proposed PSO-Ant Lion Feature Selection performs better by 11.4% and by 12.4% than PSO Feature selection and Tabu Search feature selection respectively for F Measure -MLL. The F Measure of Proposed PSO-Ant Lion Feature Selection performs better by 4.3% and by 9.51% than PSO Feature selection and Tabu Search feature selection respectively for F Measure -AML.

Figures 8 and 9 show that the fitness function of proposed PSO-ALO attains convergence in a better way on iteration number 300 than PSO.

V. CONCLUSION

Leukemia prediction at its earlier stage from the gene expression data can be performed computationally which demands high-performance results. Analyzing the gene expression data is to calculate the significant characteristics of each expression which as a whole could predict the occurrence. In this article, we have proposed a bio-inspired and hybrid feature selection algorithm to calculate the optimal set of features that provides better performance during the classification phase. The ALO is a method that encourages swarm intelligence-based algorithms to use random walking in performing operations of both exploration and exploitation. There is a hybrid method to merge into the efficiency of computation for both strategies. This can degrade the dimension of the feature space to choose an optimal feature subset. The results have proved that the accuracy of the

Proposed PSO-ALO Feature Selection has outperformed the generic PSO feature selection by 5.32% and the TABU Search feature selection by 9.01% respectively.

DATA AVAILABILTY

The digital mammogram data that support the findings of this study will be available from the corresponding author upon request.

CONFLICTS OF INTEREST

There is no conflict of interest among the authors.

REFERENCES

[1] D. Santhakumar and S. Logeswari, "Efficient attribute selection technique for leukaemia prediction using microarray gene data," *Soft Comput.*, vol. 24, no. 18, pp. 14265–14274, Sep. 2020, doi: 10.1007/s00500-020-04793-z.

[2] D. Santhakumar and S. Logeswari, "Hybrid ant lion mutated ant colony optimizer technique for leukemia prediction using microarray gene data," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 2, pp. 2965–2973, Feb. 2021, doi: 10.1007/s12652-020-02454-5.

[3] A. K. Dwivedi, "Artificial neural network model for effective cancer classification using microarray gene expression data," *Neural Comput. Appl.*, vol. 29, no. 12, pp. 1545–1554, Jun. 2018.

[4] E. Emary and H. M. Zawbaa, "Feature selection via Lèvy antlion optimization," *Pattern Anal. Appl.*, vol. 22, no. 3, pp. 857–876, 2019.

[5] S. Yadav, A. Ekbal, and S. Saha, "Information theoretic-PSO-based feature selection: An application in biomedical entity extraction," *Knowl. Inf. Syst.*, vol. 60, no. 3, pp. 1453–1478, Sep. 2019.

[6] P. S. Deepthi and S. M. Thampi, "PSO based feature selection for clustering gene expression data," in *Proc. IEEE Int. Conf. Signal Process., Informat., Commun. Energy Syst. (SPICES)*, Feb. 2015, pp. 1–5.

[7] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Comput. Biol. Chem.*, vol. 32, no. 1, pp. 29–38, Feb. 2008.

[8] P. Dutta, S. Saha, and A. B. Chauhan, "Predicting degree of relevance of pathway markers from gene expression data: A PSO based approach," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2018, pp. 3–14.

[9] H. M. Zawbaa, E. Emary, and B. Parv, "Feature selection based on antlion optimization algorithm," in *Proc. 3rd World Conf. Complex Syst. (WCCS)*, Nov. 2015, pp. 1–7.

[10] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, Oct. 1999.

[11] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Human Sci.*, 1995, pp. 39–43.

[12] K. Lekshmi, K. Rubasoundar, E. Abinaya, P. Gayathri, and T. Keerthana, "Automated selection of parameters using Tabu search in image segmentation," in *Proc. 10th Int. Conf. Intell. Syst. Control (ISCO)*, Jan. 2016, pp. 1–4.

[13] A. M. Reddy, K. S. Reddy, M. Jayaram, N. V. M. Lakshmi, R. Aluvalu, T. R. Mahesh, V. V. Kumar, and D. S. Alex, "An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network," *J. Sensors*, vol. 2022, pp. 1–11, Nov. 2022, doi: 10.1155/2022/4093658.

[14] B. Mokashi, V. S. Bhat, J. D. Pujari, S. Roopashree, T. R. Mahesh, and D. S. Alex, "Efficient hybrid blind watermarking in DWT-DCT-SVD with dual biometric features for images," *Contrast Media Mol. Imag.*, vol. 2022, pp. 1–14, Sep. 2022.

[15] X. Li, Z. Peng, B. Du, J. Guo, W. Xu, and K. Zhuang, "Hybrid artificial bee colony algorithm with a rescheduling strategy for solving flexible job shop scheduling problems," *Comput. Ind. Eng.*, vol. 113, pp. 10–26, Nov. 2017.

[16] M. Petrovic, J. Petronijevic, M. Mitic, N. Vukovic, A. Plemic, Z. Miljkovic, and B. Babic, "The ant lion optimization algorithm for flexible process planning," *J. Production Eng.*, vol. 18, no. 2, pp. 65–68, 2015.

[17] K. K. Raghunath, V. V. Kumar, M. Venkatesan, K. K. Singh, T. R. Mahesh, and A. Singh, "XGBoost regression classifier (XRC) model for cyber attack detection and classification using inception V4," *J. Web Eng.*, vol. 21, no. 4, pp. 1295–1322, 2022, doi: 10.13052/jwe1540-9589.21413.

[18] V. K. Venkatesan, K. R. K. Murugesan, K. A. Chandrasekaran, M. T. Ramakrishna, S. B. Khan, A. Almusharraf, and A. Albuali, "Cancer diagnosis through contour visualization of gene expression leveraging deep learning techniques," *Diagnostics*, vol. 13, no. 22, p. 3452, Nov. 2023.

[19] A. Balajee, R. Murugan, and K. Venkatesh, "Security-enhanced machine learning model for diagnosis of knee joint disorders using vibroarthrographic signals," *Soft Comput.*, vol. 27, no. 11, pp. 7543–7553, Jun. 2023.

[20] V. Rupapara, F. Rustam, W. Aljedaani, H. F. Shahzad, E. Lee, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model," *Sci. Rep.*, vol. 12, no. 1, p. 1000, Jan. 2022.

[21] A. Wahid and M. T. Banday, "Classification of DNA microarray gene expression Leukaemia data through ABC and CNN method," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, pp. 119–131, Jul. 2023.

[22] M. Ilyas, K. M. Aamir, S. Manzoor, and M. Deriche, "Linear programming based computational technique for leukemia classification using gene expression profile," *PLoS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0292172.

[23] T. H. Rafi, R. M. Shubair, F. Farhan, Md. Z. Hoque, and F. M. Quayyum, "Recent advances in computer-aided medical diagnosis using machine learning algorithms with optimization techniques," *IEEE Access*, vol. 9, pp. 137847–137868, 2021.

[24] B. Alphonse, V. Rajagopal, S. Sengan, K. Kittusamy, A. Kandasamy, and R. Periyasamy, "Modeling and multi-class classification of vibroarthographic signals via time domain curvilinear divergence random forest," *J. Ambient Intell. Humanized Comput.*, vol. 23, pp. 1–3, Feb. 2021.

[25] M. U. Nasir, M. F. Khan, M. A. Khan, M. Zubair, S. Abbas, M. Alharbi, and M. Akhtaruzzaman, "Hematologic cancer detection using white blood cancerous cells empowered with transfer learning and image processing," *J. Healthcare Eng.*, vol. 2023, pp. 1–20, May 2023.

T. R. MAHESH (Senior Member, IEEE) is currently the Program Head of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India. He has to his credit more than 90 research articles in Scopus/WoS and SCIE indexed journals of high repute. His research interests include image processing, machine learning, deep learning, artificial intelligence, the IoT, and data science. He has been an editor of books on emerging and new age technologies with publishers like Springer, IGI Global, and Wiley. He has served as a reviewer and a technical committee member for multiple conferences and journals of high reputation.

D. SANTHAKUMAR received the Ph.D. degree in information and communication engineering from Anna University, Chennai. He is currently a Professor with the Saveetha School of Engineering, SIMATS, Chennai, India. He has 13 years of teaching experience with good academic background. He is also a Certified RPA Trainer and completed two levels of certification with UIPath Academia and also certified in various courses from Edx, Coursera, Udemy, and NPTEL. He has published articles in national and international journals, conferences, and symposiums. His research interests include clinical data analytics, machine learning, optimization techniques, and cloud technology. His contributions toward professional bodies include a Life Member Indian Society of Technical Education, India, and a member of the Computer Society of India. He has been an active member of ICTACT, since 2010.

A. BALAJEE received the Ph.D. degree from SASTRA Deemed University, Thanjavur. He is currently an Assistant Professor with JAIN (Deemed-to-be University), specialized in machine learning, data analytics, artificial intelligence, and data science. With almost eight years of teaching and research experience, he has published four SCIE articles that belong to Q1 and have a cumulative impact factor of 18.57. He has also published in Scopus and UGC care-indexed journals. He holds a patent and his writings were published in book chapters of Springer and Taylor & Francis. He has presented at various national and international conferences. He also acted as a reviewer of SCIE, Scopus journals, and also for national and international conferences. He had worked for NBA and NAAC accreditation. He is a lifetime member of ISTE. In collaboration with ICTACT, he organized conferences and workshops for undergraduate students.

H. S. SHREENIDHI received the B.Tech. degree from MIT, Manipal University, Karnataka, India, in 2015, and the M.Tech. degree from the SJB Institute of Technology, Bengaluru, India, in 2017. Currently, he is an Assistant Professor with JAIN (Deemed-to-be University), Bengaluru. He is also a Distinguished Expert in the field of Internet of Things (IoT), with a prolific record of impactful publications. His research interests include cutting-edge applications, machine learning, and the advancements in IoT technology for solving various real-world problems.

V. VINOTH KUMAR (Member, IEEE) is currently an Associate Professor with the School of Computer Science Engineering & Information Systems (SCORE), Vellore Institute of Technology (VIT), Vellore, India. His current research interests include wireless networks, the Internet of Things, machine learning, and big data applications. He is the author/coauthor of papers in international journals and conferences, including SCI indexed papers. He has published as over than 60 papers in IEEE Access, Springer, Elsevier, IGI Global, and Emerald. He is an Associate Editor of *International Journal of e-Collaboration* (IJeC) and *International Journal of Pervasive Computing and Communications* (IJPCC). He is an editorial member of various journals.

JONNAKUTI RAJKUMAR ANNAND received the B.Tech. degree in mechanical engineering from JNT University, Hyderabad, and the M.Tech. degree in mechanical machine design from JNTUK University, India, in 2012. He is currently an Assistant Professor with the Department of Electromechanical Engineering, Sawla Campus, Arba Minch University, Ethiopia. His research interests include mechanical elements design, composite materials, and design analysis of various mechanical parts.