International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 1, February 2025, pp. 604~613
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i1.pp604-613

Ant lion and ant colony optimization integrated ensemble machine learning model for effective cancer diagnosis

Pinakshi Panda[1], Sukant Kishoro Bisoy[1], Amrutanshu Panigrahi[2], Abhilash Pati[2]
[1]Department of Computer Science and Engineering, C. V. Raman Global University, Bhubaneswar, Odisha, India
[2]Department of Computer Science and Engineering, Institute of Technical Education and Research (ITER), Siksha 'O' Anusandhan (Deemed to be University), Odisha, India

Article Info

Article history:

Received Jun 18, 2024
Revised Aug 12, 2024
Accepted Aug 20, 2024

Keywords:

Ant colony optimization
Ant lion optimization
Averaging classifier
Cancer
Machine learning
Voting classifier

ABSTRACT

Statistics from reputable sources, including the World Health Organization (WHO), demonstrate that cancer is a leading cause of death globally, accounting for millions of deaths each year. When it comes to the early identification of cancer, machine learning (ML) is crucial. To analyze complex data and identify minute patterns that may indicate the presence of cancer, it employs robust computational approaches. Improving patient outcomes relies on early cancer detection since it paves the way for faster treatment and intervention, which might lead to better prognoses and higher survival rates. To choose features, this study intends to build an ML-based ensemble model utilizing ant colony optimization (ACO) and ant lion optimization (ALO). Next, ML classifiers are used as the initial predictions' basis learners. The last forecast is the result of combining two ensemble methods: voting and averaging classifiers. Four distinct cancer microarray datasets are used to assess the approach. With an accuracy of 99.08% on the Lung cancer dataset, the voting ensemble classifier outperforms the others, according to the empirical analysis.

Corresponding Author:

Amrutanshu Panigrahi
Department of Computer Science and Engineering, Institute of Technical Education and Research (ITER), Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneswar, Odisha, India
Email: amrutansup89@gmail.com

1. INTRODUCTION

Cancer significantly contributes to the increasing worldwide mortality rate. It is a powerful competitor, causing the death of almost 10 million people each year, as reported by the World Health Organization (WHO) [1]. The effect of cancer is magnified by its many manifestations, which include lung, breast, colorectal, prostate, and stomach cancers. Various factors contribute to the high occurrence of the disease, including genetic predispositions, lifestyle decisions such as tobacco intake and sedentary behaviors, environmental contaminants like radiation, and infectious agents like hepatitis and human papillomavirus [2]. The difficulty increases since early diagnosis is challenging, as many malignancies do not exhibit recognizable symptoms until they have reached the late stages, which increases the overall number of cancer-related fatalities worldwide [3].

Machine learning (ML) is essential in transforming cancer detection and therapy. Integrating this technology into healthcare systems has facilitated the precise and efficient identification of malignant tissues, assisting in early cancer diagnosis and the development of personalized treatment plans [4]. Machine learning models may use extensive datasets and advanced algorithms to analyze intricate patterns and biomarkers that may be overlooked by human observers. Analyzing extensive genetic and clinical data may aid in identifying

new biomarkers, forecasting patient prognosis, and expediting medication development. Microarray data is crucial in cancer detection since it allows for thoroughly examining gene expression patterns in malignant cells. This method enables researchers and physicians to detect distinct gene signatures linked to various forms of cancer, facilitating precise categorization, prognosis prediction, and therapy selection [5].

Microarray datasets often have several genes or characteristics, a significant portion of which may be extraneous or duplicative for predictive modeling. Feature selection techniques aid in identifying the most useful and discriminative subset of features, hence lowering the dimensionality of the data while preserving significant information [6]. This technique not only boosts the efficiency and performance of machine learning models but also improves their interpretability and ability to generalize to new data. Metaheuristic algorithms are efficient and effective techniques for selecting features while working with microarray data in cancer research. Due to the complex and multi-dimensional nature of microarray datasets, metaheuristic algorithms such as genetic algorithms, particle swarm optimization, and simulated annealing provide a systematic method to identify important genes or features that play a critical role in cancer diagnosis, prognosis, and prediction of treatment response [7].

Aziz [8] proposed a hybrid model with independent component analysis (ICA) with two metaheuristic approaches, Cuckoo search (CS), artificial bee colony (ABC), and genetic algorithm (GA), to propose two hybrid models with naïve Bayes (NB) classifier. Nekouie *et al.* [9] proposed an ensemble-based model for cancer diagnosis. For this, the author used a two-stage feature selection process. Initially, multimodal optimization and the Firefly algorithm are applied to the microarray data in the first stage. The next stage of feature selection is particle swarm optimization (PSO). Machine learning classifiers are applied to the selected features, and then to the initial prediction, the soft voting ensemble classifier is applied. Naji *et al.* [10] proposed a ML model for cancer diagnosis. The model uses five different ML classifiers to classify the cancer disease. To evaluate the model's accuracy, the F-1 score, specificity, and sensitivity have been calculated. Lu *et al.* [11] proposed an ensemble model for effective cancer diagnosis. Finally, different ML classifiers are applied to predict cancer disease. Then, the voting method is applied as the ensemble technique to enhance the prediction result. Tavasoli *et al.* [12] have proposed an ensemble machine-learning model for effective cancer prediction with water cycle algorithm and support vector machine (SVM) for feature selection and classification purposes. Sun *et al.* [13] have integrated the fuzzy roughest, entropy-based feature selection with the Fisher score. The developed model uses the fisher score to reduce the number of genes present in the dataset. Shukla *et al.* [14] have proposed a ML-based hybrid model with two different feature selection stages. In the initial phase of the feature selection, the minimum redundancy and maximum relevance are applied to select the relevant genes from the dataset. Then, the next feature selection phase integrates two metaheuristic approaches, including the teaching learning algorithm and the gravitational search algorithm. The naïve Bayes classifier calculates the fitness function and classifies the cancer. Meenachi *et al.* [15] proposed another hybrid model based on the ant colony optimization algorithm (ACO), genetic algorithm (GA), and tabu search algorithm (TSA) followed by a fuzzy rough set classifier to classify cancer. Table 1 shows the summary of the reported literatures.

Table 1. Literature survey summary

Reference	Feature selection algorithm	Classifier	Ensemble classifier	Cancer dataset
Aziz [8]	ICA, CS, ABC, GA	NB	--	Colon, Lung II, Prostate, Acute leukemia and Leukemia
Nekouie *et al.* [9]	Multimodal, Firefly algorithm (FA), PSO	SVM, K-nearest neighbors (KNN), extreme learning machine (ELM)	Soft voting	Brain, Colon, Leukemia, Lung, Prostate, Breast, SRBCT, Ovarian
Naji *et al.* [10]	--	SVM, random forest (RF), decision tree (DT), KNN	--	Breast
Lu *et al.* [11]	--	Linear regression (LR), KNN, SVM, multilayer perceptron (MLP), DT	Voting	Cervical
Tavasoli *et al.* [12]	Water cycle algorithm (WCA)	SVM	Soft voting	Leukemia, Colon, Prostate, DLBCL
Sun *et al.* [13]	Fisher score	Fuzzy rough set	--	SRBCT, Colon, Brain, Lymphoma, Leukemia
Shukla *et al.* [14]	Teaching learning-based algorithm (TLBO), gravitational search algorithm (GSA), minimum redundancy maximum relevance (mRMR)	NB	--	Leukemia 1 and 2, DLBCL, Prostate, Colon
Meenachi *et al.* [15]	ACO, GA, TS	Fuzzy rough set	--	DLBCL, SRBCT, Breast, Leukemia

The current work aims to propose an ensemble ML model for cancer diagnosis. The proposed work adopts two metaheuristic approaches, ant lion optimizer (ALO) and ant colony optimizer (ACO), to select relevant features from the microarray data. Then, three ML-based classifiers are applied along with the voting and averaging ensemble technique for classification purposes. The objectives of the current work can be summarized as follows: i) to implement ALO and ACO as the metaheuristic feature selection method; ii) to implement ML-based classifiers such as SVM, AdaBoost, XGBoost, and RF classifiers for initial prediction; iii) to apply the voting and averaging ensemble technique to form an ensemble ML-based diagnosis model; iv) to evaluate the proposed model over three kinds of cancer microarray dataset; and v) to measure the model's performance using different ML-based evaluative parameters, including accuracy, precision, recall, specificity, and F1-score.

2. MATERIALS AND METHODS

This section describes the dataset and the methods used to develop the proposed model. To develop the model, ant lion optimization and ant colony optimization are applied sequentially as the feature selection algorithm. Then, to the selected features, three machine learning-based classifiers, including support vector machine, AdaBoost, and XGBoost, are applied as the base classifiers to perform initial prediction. The two ensemble strategies (voting and averaging) are applied to the initial prediction to build the proposed model.

2.1. Dataset description

The proposed approach was evaluated on four cancer datasets [16]. Table 2 shows the dataset description. All these four datasets contain microarray data with binary class values. A brief description of these datasets is depicted in Table 2.

Table 2. Dataset description

Dataset	Number of Features	Number of Samples	Class
Ovarian	15,154	253	2
Lung	12,533	181	2
Breast	24,481	97	2
Leukemia	7,129	72	2

2.2. Ant lion optimization algorithm

The ant lion optimization (ALO) method is derived from the predatory strategy used by antlions, where they capture ants by constructing conical trenches. Within the context of ALO, the actions of antlions and ants are simplified and represented as an optimization algorithm [17], [18]. In this algorithm, antlions symbolize potential solutions, while ants symbolize the process of searching and investigating the space of possible solutions. A represents the ant, and AL represents the ant lion. The random walk of A with random step size t to search the food can be modeled as (1):

$$A(t) = [0,\ CS(2\mathrm{P}(t1) - 1, \ldots . (tn) - 1,\ )\ ] \quad (1)$$

where $P(t1)$ is the stochastic function, CS is the cumulative sum, and n is the maximum number of iterations. $P(t)$ can be defined using (2), with r as a random variable between [0, 1].

$$\mathrm{P}(t) = \{\ 1\ \ if\ r > 0.5, 0\ if\ r \leq 0.5 \quad (2)$$

Using the random walk, the next position of the ant can be updated using (3). However, in (1) cannot used directly as each search space of an ant has a boundary. So, to update, the position needs to be normalized by using (3), which is also used to define the next position of A.

$$A(t+1) = \frac{A(t) + Elite(t)}{2} \quad (3)$$

where $A(t)$ is the current position of the Ant and $Elite(t)$ is the best candidate solution found so far. In the final stage of hunting, when A reaches the pit of the trap of AL, AL drags A to the sand and then consumes the prey. Then, the AL position is updated using (3) [19].

2.3. Ant colony optimization algorithm

Ant colony optimization (ACO) is a heuristic optimization technique that emulates the foraging behavior of ants to solve optimization issues. The method works on a graph representing the issue, with nodes representing potential solutions and edges indicating connections between solutions. The ACO algorithm employs artificial ants to create solutions iteratively [20]. The probability of movement of an ant (a) from location i to location j can be mathematically modeled as (4):

$$P_{ij}(a) = \begin{cases} \frac{(\gamma_{ij}{}^{\alpha})\cdot(\vee_{ij}{}^{\beta})}{\sum_{l\in N}(\gamma_{il}{}^{\alpha})\cdot(\vee_{il}{}^{\beta})} \; if \; j < N \\ 0 \qquad\qquad Otherwise \end{cases} \tag{4}$$

where the γ_{ij} is the pheromone level on the path $i-j$ at time t. $\vee_{ij}$ is the inverse of the distance between two locations i and j, α, and β are the constant, N is the total number of allowed locations an ant can move, and l is an intermediate location between j and N. The pheromone level can be updated by using the (5):

$$\gamma ij(t+1) = (1-\rho)\gamma ij + \upsilon_{ij} \tag{5}$$

where ρ is the pheromone evaporation rate, A is the total number of ants; upsilon, υ_{ij} is the total pheromone level deposited at the edge $i-j$ at time t. These two processes will be repeated by the ant until the iteration does not exceed the maximum one [21], [22].

2.4. Voting and averaging ensemble technique

Voting and averaging are often used ensemble methods in machine learning to enhance forecast accuracy by aggregating the results of many base models. These strategies are especially efficient when separate models possess distinct strengths and limitations since they may mutually enhance each other and result in more resilient and precise forecasts [23]. A voting ensemble involves using numerous base models, such as decision trees, support vector machines, or neural networks, independently providing predictions on the same dataset. The ultimate forecast of the ensemble is established by consolidating the individual estimates using a voting method. An averaging ensemble, sometimes called an averaging or mean ensemble, combines the predictions of base models by calculating their average [24], [25].

2.5. Workflow of the proposed work

The proposed methods adopt two phases for cancer classification, including the two-phase feature selection and classification processes. In the first phase, the feature selection methods ALO and ACO are applied to select the subset of initial features. Then, the second phase starts by applying the base learners SVM, RF, AdaBoost, and XGBoost to have an initial prediction. Then, to the initial prediction, voting and averaging are applied to make the final prediction. Figure 1 shows the working principle of the proposed model. The workings of the proposed method can be explained below.

Step 1: Consider the microarray data for the normalization process using the standard scaler method.

Step 2: Split the dataset into two sets, the training set and the testing set, with a ratio of 80:20.

Step 3: To the train data, apply a two-stage feature selection process.

Step 4: Apply the ALO feature selection algorithm

- Initiate the ant and ant lion populations with maximum iteration T_{max}.
- Calculate the fitness function of both populations using the k-fold cross-validation method with accuracy as a key for fitness function calculation.
- $Fit(Ant, Antlion) = \frac{\sum_{i=1}^{n} Accuracy_i}{n}$, with n as the number of folds.
- Identify the next position of the ant and ant lion.
- Re-calculate the fitness function
- If $Fit_{old} > Fit_{new}$, then replace $Fit()$ with Fit_{new}
- Update the Fit () until $t = T_{max}$
- Update the feature set

Step 5: Apply ACO feature selection algorithm

- Initiate the ant population, maximum iteration T_{max}
- Calculate the fitness function of both populations using the k-fold cross-validation method with accuracy as a key for fitness function calculation
- $Fit(Ant) = \frac{\sum_{i=1}^{n} Accuracy_i}{n}$, with n as the number of folds
- Identify the next position of the ant

- Re-calculate the fitness function
- If $Fit_{old} > Fit_{new}$, then replace $Fit()$ with Fit_{new}
- Update the $Fit()$ until $t = T_{max}$
- Update final features for Phase 2 of the proposed model.

Step 6: Apply base classifiers (SVM, RF, XGBoost, and AdaBoost).
Step 7: Apply ensemble classifier voting and averaging.
Step 8: Apply test data to two trained models obtained from Step 7.
Step 9: Evaluate the performance of the trained models.

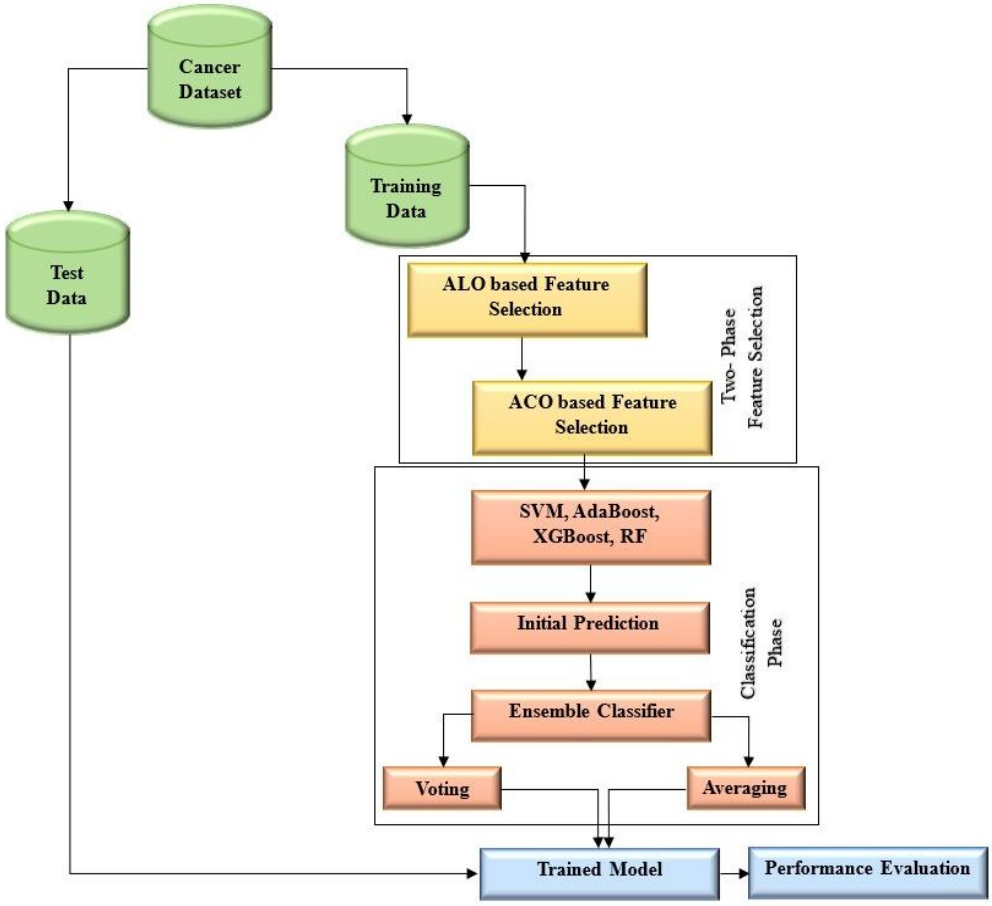

Figure 1. Workflow of the proposed model

3. RESULT AND DISCUSSION

The above-stated proposed model is evaluated on the system, including 8 GB RAM, Windows 11 OS, and an Intel i3 processor with a 2.6 GHz clock speed. The proposed model is analyzed in a Python environment using Anaconda navigator. For evaluating the performance, the proposed approach adopts five different ML-based evaluative parameters, including accuracy, precision, recall, F1-score, and specificity, which can be defined using (6)-(10) with TPO, TNE, FPO, and FNE as true positive, true negative, false positive and false negative respectively. Table 3 shows the analysis of the hybrid model using ALO and ACO feature selection mechanisms. Tables 4 and 5 represent the analysis of the model using voting and averaging ensemble techniques with ALO and ACO feature selection algorithms. Figures 2 to 6 show the performance comparison of the voting and averaging ensemble technique to the above-discussed evaluative parameters.

$$Accuracy = \frac{TPO+TNE}{TPO+FPO+TNE+FNE} \quad (6)$$

$$Precision = \frac{TPO}{TPO+FPO} \quad (7)$$

$$Recall = \frac{TPO}{TPO+FNE} \quad (8)$$

$$F-1\ Score = \frac{TPO}{TPO+\frac{1}{2}(FPO+FNE)} \quad (9)$$

$$Specificity = \frac{TNE}{FPO+TNE} \quad (10)$$

The empirical analysis shows that the ensemble model with a voting classifier and weighted averaging outperforms all conventional ML-based classifiers in the current work. For ovarian cancer, the proposed ensemble approach with a voting classifier shows an accuracy of 98.44%, and with the weighted averaging, the model shows 96.33%. The voting and weighted averaging classifications for lung cancer show 99.08% and 97.36%, respectively. For the breast cancer dataset, the voting classifier shows a 98.06% accuracy level, and the weighted averaging classifier shows an accuracy level of 96.20% accuracy level. Similarly, the proposed approach with the voting classifier shows an accuracy level of 97.27%, and weighted averaging shows an accuracy level of 95.5%. Based on a comparison of the voting and averaging techniques, it can be observed from Tables 4 and 5 that the voting technique outperforms the averaging technique. So, as an evaluative method, the AUC value of the voting technique in contrast to different datasets is shown in Figures 7 to 10. The AUC values for the ovarian cancer, lung cancer, breast cancer, and leukemia datasets are 0.986, 0.99, 0.978, and 0.973, respectively. The analysis shows that the developed model exhibits high accuracy, sensitivity, and specificity, which indicates that the developed model can be effectively used to classify cancer and non-cancer patients. The model shows robustness and scalability across different cancer datasets. So, the developed high-performing ML-based model can be used to develop a more reliable diagnostic tool that can help clinicians make more informed decisions regarding cancer diagnosis. In order to show the efficacy, the proposed model is compared with some existing literature. Table 6 shows the performance comparison of proposed model with existing literatures in terms of accuracy. It can be clearly observed that the proposed model out performs all existing literature across all datasets.

Table 3. Performance of individual model with ALO and ACO feature selection

Dataset	Hybrid model with ALO and ACO	Accuracy	Precision	Recall	F1-score	Specificity
Ovarian	SVM	88.89	91.43	86.49	88.89	91.43
	RF	84.72	91.67	80.49	85.71	90.32
	XGBoost	86.11	82.86	87.88	85.29	84.62
	AdaBoost	87.50	87.23	93.18	90.11	78.57
Lung	SVM	87.35	89.88	90.96	90.42	80.46
	RF	86.96	90.66	91.16	90.91	76.39
	XGBoost	90.51	91.72	93.94	92.81	84.09
	AdaBoost	89.33	92.99	90.12	91.54	87.91
Breast	SVM	83.98	89.15	88.46	88.80	72.55
	RF	84.21	87.79	87.12	87.45	79.22
	XGBoost	88.63	90.16	93.75	91.92	77.22
	AdaBoost	84.32	89.15	88.46	88.80	74.55
Leukemia	SVM	84.25	87.67	85.33	86.49	82.69
	RF	87.60	86.96	90.91	88.89	83.64
	XGBoost	86.67	93.94	86.11	89.86	87.88
	AdaBoost	88.37	91.03	89.87	90.45	86.00

Table 4. Performance of voting ensemble technique with ALO and ACO feature selection

Dataset	Accuracy	Precision	Recall	F1-Score	Specificity
Ovarian	98.44	98.73	98.73	98.73	97.96
Lung	99.08	99.32	99.32	99.32	98.59
Breast	98.06	99.14	98.29	98.71	97.37
Leukemia	97.27	96.97	98.46	97.71	95.56

Table 5. Performance of averaging ensemble technique with ALO and ACO feature selection

Dataset	Accuracy	Precision	Recall	F1-Score	Specificity
Ovarian	96.33	97.10	97.10	97.10	95.00
Lung	97.36	97.42	98.69	98.05	94.59
Breast	96.20	98.29	96.64	97.46	94.87
Leukemia	95.54	95.52	96.97	96.24	93.48

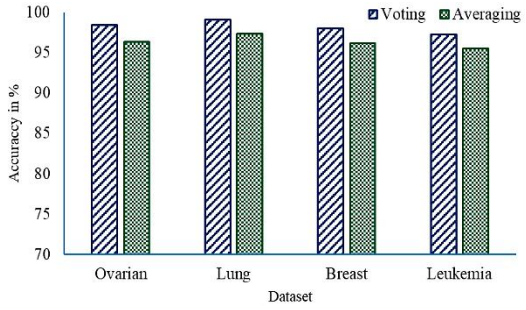

Figure 2. Accuracy comparison among voting and averaging technique

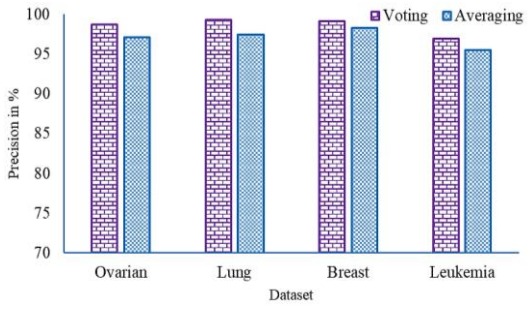

Figure 3. Precision comparison among voting and averaging technique

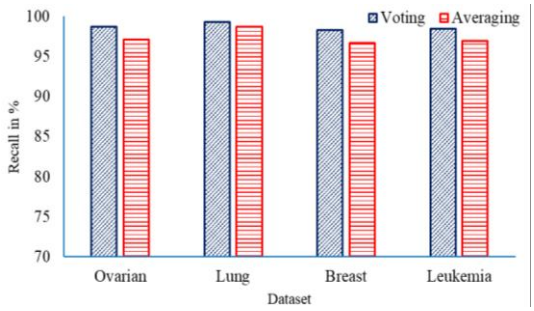

Figure 4. Recall comparison among voting and averaging technique

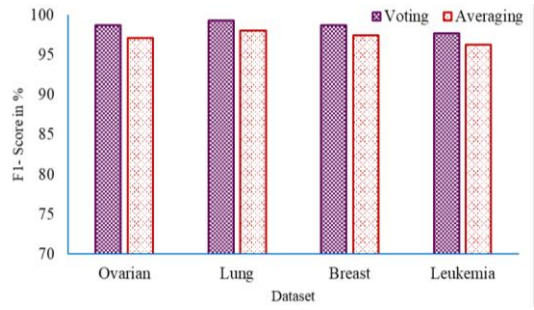

Figure 5. F1-score comparison among voting and averaging technique

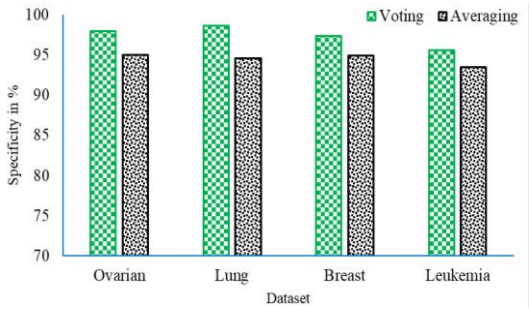

Figure 6. Specificity comparison among voting and averaging technique

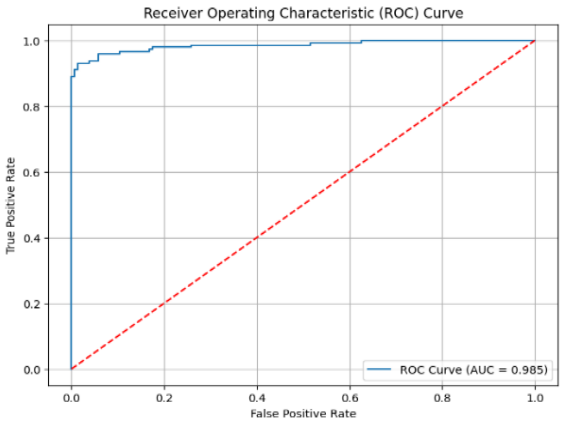

Figure 7. ROC curve for ovarian cancer dataset

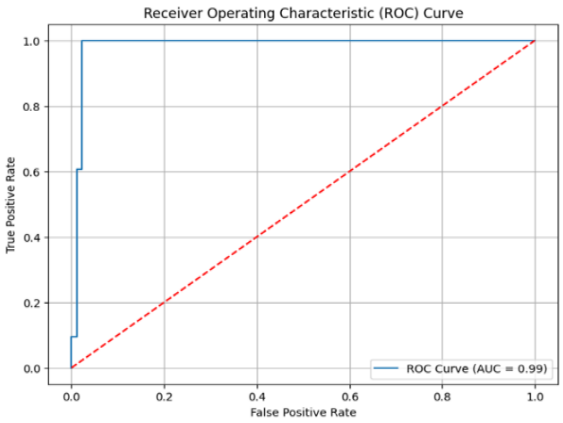

Figure 8. ROC curve for lung cancer dataset

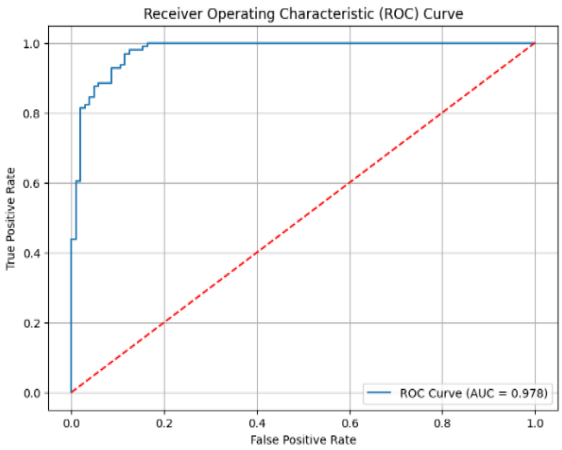

Figure 9. ROC curve for breast Cancer dataset

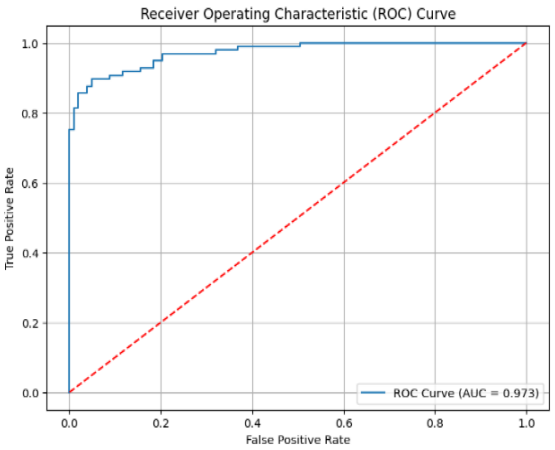

Figure 10. ROC curve for Leukemia dataset

Table 6. Accuracy comparison of the proposed model with existing work

Reference	Ovarian	Lung	Breast	Leukemia
[8]	--	--	--	92.33
[9]	95.65	93.56	84	91.36
[10]	--	--	97.2	--
[12]	--	--	--	92.7
[13]	--	--	--	86.17
Proposed	**98.44**	**99.08**	**98.06**	**97.27**

4. CONCLUSION

The current work aims the develop an ensemble-based cancer diagnosis model with ALO and ACO as the feature selection algorithm. To proposed method equips four types of machine learning classifiers: support vector machine, random forest, AdaBoost, and XGBoost with two ensemble classifiers including voting and averaging. The experimental study done in this work shows that the voting classifier outperforms the averaging classifiers with an accuracy of 98.44%, 99.08%, 98.06%, and 97.27% for the ovarian, lung, breast, and leukemia Cancer datasets, respectively. As per the ROC analysis, the AUC values of the proposed model with voting classifier are 0.985, 0.99, 0.978, and 0.973 for the ovarian, lung, breast, and leukemia cancer datasets. Our research on a model for cancer detection using machine learning suggests that this field of study and the medical community may undergo a paradigm change. Our technique improves the reliability and capacity to identify cancer early. The high accuracy, sensitivity, and specificity levels suggest that this might result in better patient outcomes. To fully grasp the intricacies of cancer, it is necessary to use multi-modal approaches that include multiple datasets, including genetic and clinical data. Ultimately, our results imply that machine learning might improve cancer detection, which would benefit society and the research community via better resource allocation, lower misdiagnosis rates, and more efficient healthcare delivery.

REFERENCES

[1] A. Panigrahi *et al.*, "En-MinWhale: an ensemble approach based on MRMR and whale optimization for cancer diagnosis," *IEEE Access*, vol. 11, pp. 113526–113542, 2023, doi: 10.1109/ACCESS.2023.3318261.

[2] M. Khalsan *et al.*, "A survey of machine learning approaches applied to gene expression analysis for cancer prediction," *IEEE Access*, vol. 10, pp. 27522–27534, 2022, doi: 10.1109/ACCESS.2022.3146312.

[3] B. Ma, F. Meng, G. Yan, H. Yan, B. Chai, and F. Song, "Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data," *Computers in Biology and Medicine*, vol. 121, Art. no. 103761, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103761.

[4] S. Osama, H. Shaban, and A. A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: a comprehensive review," *Expert Systems with Applications*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118946.

[5] B. Sahu, A. Panigrahi, S. K. Rout, and A. Pati, "Hybrid multiple filter embedded political optimizer for feature selection," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, Hyderabad, India, 2022, pp. 1-6, doi: 10.1109/ICICCSP53532.2022.9862419.

[6] A. Pati, M. Parhi, B. K. Pattanayak, B. Sahu, and S. Khasim, "CanDiag: fog empowered transfer deep learning based approach for cancer diagnosis," *Designs*, vol. 7, no. 3, 2023, doi: 10.3390/designs7030057.

[7] A. Panigrahi, S. Bhutia, B. Sahu, M. G. Galety, and S. N. Mohanty, "BPSO-PSO-SVM: an integrated approach for cancer diagnosis," *Lecture Notes in Electrical Engineering*, vol. 905, pp. 571–579, 2022, doi: 10.1007/978-981-19-2177-3_53.

[8] R. M. Aziz, "Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data," *Medical & Biological Engineering & Computing*, vol. 60, no. 6, pp. 1627–1646, Jun. 2022, doi: 10.1007/s11517-022-02555-7.
[9] N. Nekouie, M. Romoozi, and M. Esmaeili, "A new evolutionary ensemble learning of multimodal feature selection from microarray data," *Neural Processing Letters*, vol. 55, no. 5, pp. 6753–6780, 2023, doi: 10.1007/s11063-023-11159-7.
[10] M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science*, vol. 191, pp. 487–492, 2021, doi: 10.1016/j.procs.2021.07.062.
[11] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: an ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, May 2020, doi: 10.1016/j.future.2019.12.033.
[12] N. Tavasoli, K. Rezaee, M. Momenzadeh, and M. Sehhati, "An ensemble soft weighted gene selection-based approach and cancer classification using modified metaheuristic learning," *Journal of Computational Design and Engineering*, vol. 8, no. 4, pp. 1172–1189, Jul. 2021, doi: 10.1093/jcde/qwab039.
[13] L. Sun, X.-Y. Zhang, Y.-H. Qian, J.-C. Xu, S.-G. Zhang, and Y. Tian, "Joint neighborhood entropy-based gene selection method with fisher score for tumor classification," *Applied Intelligence*, vol. 49, no. 4, pp. 1245–1259, Apr. 2019, doi: 10.1007/s10489-018-1320-1.
[14] A. K. Shukla, P. Singh, and M. Vardhan, "Gene selection for cancer types classification using novel hybrid metaheuristics approach," *Swarm and Evolutionary Computation*, vol. 54, May 2020, doi: 10.1016/j.swevo.2020.100661.
[15] L. Meenachi and S. Ramakrishnan, "Metaheuristic search based feature selection methods for classification of Cancer," *Pattern Recognition*, vol. 119, Nov. 2021, doi: 10.1016/j.patcog.2021.108079.
[16] Z. Zhu, Y.-S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognition*, vol. 40, no. 11, pp. 3236–3248, Nov. 2007, doi: 10.1016/j.patcog.2007.02.007.
[17] S. Mirjalili, "The ant lion optimizer," *Advances in Engineering Software*, vol. 83, pp. 80–98, May 2015, doi: 10.1016/j.advengsoft.2015.01.010.
[18] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016, doi: 10.1016/j.neucom.2016.03.101.
[19] H. M. Zawbaa, E. Emary, and B. Parv, "Feature selection based on antlion optimization algorithm," in *2015 Third World Conference on Complex Systems (WCCS)*, Nov. 2015, pp. 1–7, doi: 10.1109/ICoCS.2015.7483317.
[20] H. Peng, C. Ying, S. Tan, B. Hu, and Z. Sun, "An improved feature selection algorithm based on ant colony optimization," *IEEE Access*, vol. 6, pp. 69203–69209, 2018, doi: 10.1109/ACCESS.2018.2879583.
[21] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MLACO: a multi-label feature selection algorithm based on ant colony optimization," *Knowledge-Based Systems*, vol. 192, Mar. 2020, doi: 10.1016/j.knosys.2019.105285.
[22] B. M and S. V, "Fault-type coverage based ant colony optimization algorithm for attaining smaller test suite," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3, pp. 507–519, Sep. 2020, doi: 10.11591/ijai.v9.i3.pp507-519.
[23] H. E. Kiziloz, "Classifier ensemble methods in feature selection," *Neurocomputing*, vol. 419, pp. 97–107, Jan. 2021, doi: 10.1016/j.neucom.2020.07.113.
[24] G. Sahoo *et al.*, "Breast cancer relapse disease prediction improvements with ensemble learning approaches," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 1, pp. 335–342, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp335-342.
[25] T. E. Mathew, "A cost sensitive SVM and neural network ensemble model for breast Cancer classification," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 11, no. 2, Apr. 2023, doi: 10.52549/ijeei.v11i2.3934.

BIOGRAPHIES OF AUTHORS

Pinakshi Panda completed her Master of Computer Applications (M.C.A.) degree at Biju Patnaik University of Technology (BPUT) in Rourkela, India, in 2012. Subsequently, she earned a Master of Technology (M.Tech.) degree in computer science and informatics from Siksha 'O' Anusandhan (SOA) University, Bhubaneswar, India, in 2014. She is pursuing a Doctor of Philosophy (Ph.D.) degree in computer science and engineering at C. V. Raman Global University (CGU), Bhubaneswar, India. She has accumulated more than three years of professional experience in the academic field. She can be contacted at email: pinakshipanda@gmail.com.

Sukant Kishoro Bisoy received a Doctor of Philosophy (Ph.D.) degree, a Master of Technology (M.Tech.) degree, and a bachelor's degree in computer engineering in 2017, 2003, and 2000, respectively. He presently occupies the professor and dean role within the Department of Computer Science and Engineering at C. V. Raman Global University in India. The current research interests of the individual encompass the fields of neuro-robotics, machine learning, cloud computing, and software defined networks. He has been designated as a Margdarshak by the All-India Council for Technical Education/Ministry of Human Resource Development. He has over 100 publications in SCIE, Scopus journals, and conferences. He can be contacted at email: sukantabisoyi@yahoo.com.

Amrutanshu Panigrahi is currently working as an assistant professor at the Deptartment of Computer Science and Engineering, Siksha O Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. He obtained M.Tech. in information technology from the College of Engineering and Technology, Govt. of Odisha, and a B.Tech. from BPUT Odisha in 2014 and 2011, respectively. He completed his Ph.D. in computer science and engineering with Siksha 'O' Anusandhan University in 2023. His research interests include blockchain, ML, and DL. He has over 50 publications in SCIE, Scopus journals, and conferences. He can be contacted at amrutansup89@gmail.com.

Abhilash Pati is currently working as an assistant professor in the Department of Computer Science and Engineering, FET-ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India. He completed his B.Tech. and M.Tech. in computer science and engineering with the BPUT, Odisha, India, in 2009 and 2012, respectively, and his Ph.D. in computer science and engineering with Siksha 'O' Anusandhan University, Bhubaneswar, India in 2023. His research interests include the IoT, fog computing, ML, and DL, and he has more than 50 publications in SCIE, ESCI, Scopus indexed journals and/or conferences to his credit. He can be contacted at er.abhilash.pati@gmail.com.