

An Intelligent Model for Prediction of Breast Cancer Applying Ant Colony Optimization

Annwesha Banerjee*, Sumit Das,
Aniruddha Biswas and Avisekh Kumar Tiwari

Department of Information Technology, JIS College of Engineering, Kalyani, WB, India.

<http://dx.doi.org/10.13005/bbra/3347>

(Received: 03 October 2024; accepted: 06 March 2025)

Breast cancer remains the most prevalent cancer among women, necessitating early and accurate detection to mitigate progression and improve outcomes. Machine learning (ML) techniques are particularly promising for analyzing vast datasets and identifying potential cases. This paper introduces an ML-based model for breast cancer prediction, leveraging classifiers such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Gradient Boosting, and XGBoost. To boost the models' accuracy, Ant Colony Optimization (ACO) was applied to optimize hyperparameters. Feature selection was conducted using the SelectKBest method, enhancing model precision and reducing computation. The dataset, sourced from the UCI Machine Learning Repository, facilitated robust model training. Notably, the highest prediction accuracy achieved by this approach is 99%, with the Random Forest classifier optimized through ACO. The dataset consisting of thirty neumeric features. This research highlights the potential of integrating ML and optimization techniques to enhance disease prediction capabilities by early prediction of disease in turn a better patient outcome.

Keywords:Ant Colony Optimization (ACO), Breast Cancer Prediction, Feature Selection (SelectKBest), Hyperparameter Optimization, Machine Learning Classifiers.

Machine learning, a branch of artificial intelligence (AI), is dedicated to developing algorithms and models that empower computers to learn from data and make informed predictions or decisions without explicit programming for each task. The fundamental concept behind machine learning is to enable systems to improve over time by learning from experience, thereby enhancing accuracy and efficiency with more data exposure. One of the most impactful applications of machine learning is in healthcare, where it plays a crucial role in disease diagnosis, treatment planning, and patient outcome prediction. In

this study, we propose an optimized machine learning model specifically designed for breast cancer prediction. By leveraging data-driven learning techniques, this model aims to improve diagnostic accuracy and facilitate early detection, thereby contributing meaningfully to healthcare advancements. According to World Health Organization 2.3 million women received a breast cancer diagnosis in 2022, and 670,000 people died from the disease worldwide. In any nation on earth, breast cancer can strike women at any age after puberty, though its incidence rises with age.¹

Main Contributions of the work are as follows:

*Corresponding author E-mail: annwesha.banerjee@jiscollege.ac.in



- Ant Colony Optimization (ACO) and machine learning techniques are combined to forecast breast cancer.
- Demonstrates the efficacy of ACO in a range of machine learning applications.
- Improves generalization by optimizing classifier hyperparameters and increasing prediction accuracy.
- Combines a number of methods with sophisticated optimization to propose a comprehensive framework.

The proposed model incorporates a range of widely used classifiers, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Gradient Boosting, and XGBoost. These algorithms are chosen for their proven effectiveness in classification tasks and their ability to capture complex patterns within the data, making them well-suited for predictive modeling. After the initial model design, Ant Colony Optimization (ACO) is applied to fine-tune the hyperparameters, enhancing the model's performance by optimizing key parameters. This combination of classifiers with ACO not only boosts accuracy but also strengthens the model's reliability and robustness in identifying subtle data patterns. In swarm intelligence techniques, this algorithm is a member of the family of ant colony algorithms and is an example of a metaheuristic optimization. First put out by Marco Dorigo in his doctoral thesis in 1992.² Gradient Boosting and XGBoost are powerful ensemble techniques capable of high accuracy by successively building models to correct errors from previous iterations; Random Forest provides strong performance with resistance to overfitting; SVM is excellent at handling complex, non-linear data and high dimensional features; and KNN is straightforward and easy to understand.

The objective is to combine the capabilities of these machine learning algorithms with an optimization method that imitates ant foraging behavior to produce a more precise and effective prediction of breast cancer. Applying K Nearest Neighbor with Ant Colony Optimization produced results of 94%, Support Vector Machine with Ant Colony Optimization produced results of 96%, Gradient Boosting with Ant Colony Optimization produced results of 96%, XGBoost with Ant Colony Optimization produced results of 95.6%, and Random Forest with Ant Colony

Optimization produced the highest accuracy of 99% with a false positive rate of 0.87%. Along with that AUC score, sensitivity and specificity score of the model are 98%, 100% and 97% respectively. The data collection of breast cancer from the UCI Machine Learning Repository is a valuable resource for researchers and practitioners in the field of medical diagnostics. This dataset consists of clinical and diagnostic features extracted from digitized images of breast masses, obtained through fine needle aspirates (FNA). With a total of 569 instances, the dataset includes 30 features, encompassing characteristics such as texture, radius, perimeter, and symmetry. Each instance is labeled as either malignant or benign, providing ground truth information for training and evaluating machine learning models.

The combination of various machine learning methods with the Ant Colony Optimization technique—which has received little attention in the field of breast cancer prediction, represents the work's primary contribution. The application of ACO has proven its effectiveness in different machine learning based applications found in existing literature. In a work proposed by Adamu Garba *et al.* application of ACO in Federated searching achieved significant improvised results.³ The use of ACO as a performance-enhancing technique, which raises prediction accuracy and optimizes classifier hyperparameters for improved generalization, is what makes this study novel. Unlike other models that rely on single classifiers or less complex optimization techniques, this methodology presents a comprehensive and resilient framework for breast cancer prediction by merging many algorithms and an advanced optimization technique³. In another work proposed by Duygu Yilmaz Eroglu *et al.*,⁴ the application of adaptive ACO has also proven efficiency.⁴

In section 2 few contemporary works in this field have been analyzed in details. In the section 3 and 4 the description and experimental outcomes of the proposed method have been discussed. In section 5 conclusion and the future scope of the study have been presented.

Literature Review

Numerous studies have explored machine learning approaches for cancer prediction, yielding promising results across various types of cancer. Researchers developed a breast cancer prediction

model using Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest, achieving a peak accuracy of 97.14%.⁵ Another researcher applied SVM, Decision Tree, Naïve Bayes, and KNN for breast cancer prediction, reaching a maximum accuracy of 97.13% with SVM. A Scientist proposed a model utilizing Random Forest and SVM, with SVM attaining an accuracy of 86%.⁶ Scientists targeted lung cancer prediction, employing SVM, Naïve Bayes, Decision Tree, and Random Forest; Random Forest achieved the highest accuracy of 96.8%.⁷

In colorectal cancer article, researcher applied Naïve Bayes, SVM, Random Forest, and KNN, with Naïve Bayes achieving a top accuracy of 92.83%.⁸ Another Scientist used Logistic Regression, Random Forest, SVM, and KNN for colorectal cancer, obtaining a peak accuracy of 98.2%.⁹ For brain tumor detection, gradient vanishing and overfitting issues by implementing a transfer learning ResNet-50 model with global average pooling.¹⁰ Using Deep Neural Networks (DNN) and KNN (for $k=1$ and $k=3$), the authors reported recall, accuracy, and F-measure values of 0.97, and an AUC of 0.984 with DNN. With KNN at $k=1$, they achieved a recall of 0.955 and accuracy of 0.956¹¹. Their model classified images into pituitary, meningioma, and no tumor categories, yielding a general accuracy of 92.13%.¹²

Further advancements in brain tumor detection were made by who developed an MRI-based model to localize and classify tumors.¹³ In colon cancer detection article, researcher proposed a model using CNNLeNetV2, which outperformed other configurations with an accuracy of 99.67%.¹⁴ In the realm of Covid-19 detection researcher utilized chest X-ray images neural networks, achieving an AUC score near 0.999, indicative of high diagnostic accuracy. Finally, Researchers classified lung and colon cancer images using the LC250 achieving 96.11% accuracy in training and 97.20% in validation for three categories: benign, adenocarcinoma, and squamous cell carcinoma. These studies collectively highlight the effectiveness of machine learning in disease prediction, ng robust performance across a variety of cancers through optimized classifiers and feature selection methods¹⁵.

Recent studies have introduced innovative machine learning¹⁶ models for disease prediction,

particularly in cancer and heart disease. Scientists proposed a novel convolutional neural network, EFFI-CNN, inspired by the ICDSSPLDCNN and EASPLD-CNN trials, showcasing advancements in CNN architecture¹⁷. In heart disease prediction paper, author developed a model using a concatenated ensemble classifier, achieving an accuracy of 86.89%.¹⁸ In a subsequent study, author and colleagues proposed a heart disease prediction model combining SVM-XGBoost with Particle Swarm Optimization, utilizing LASSO for feature selection, and achieving an improved accuracy of 91.8%.¹⁹

A comprehensive model presented in another study employed multiple machine learning algorithms SVM, logistic regression, decision tree classification, KNN, Gaussian Naive Bayes, and artificial neural networks reaching a high accuracy of 99%²⁰. In early-stage lung cancer detection, researcher compared several classifiers to determine the most effective technique for early diagnosis, an area particularly challenging due to subtle early symptoms. They reported an accuracy of 95.56% for SVM and 88.40% for KNN, emphasizing SVM's effectiveness in this context. researcher's research focused on feature selection methods, commonly utilizing Random Forest and Logistic Regression²¹. For lung cancer classification, this study tested Naive Bayes, SVM, and KNN classifiers, and achieved an accuracy of 81.25% with the Radial Basis Function (RBF) classifier²². These contributions underscore the evolving role of machine learning in healthcare²³. Enhanced feature selection, ensemble models, and optimized classifiers have significantly improved prediction accuracy in diseases with complex data patterns, offering promising tools for early diagnosis and effective intervention.

With classifiers like SVM, Random Forest, and KNN, several models achieve excellent accuracy in the literature on machine learning algorithms for cancer prediction, which shows impressive results across a variety of cancer types. Still, there are a number of obstacles and gaps. A single classifier or combinations of simple classifiers are used in many research, which may not necessarily produce the best prediction accuracy or generalizability. Furthermore, few studies integrate advanced optimization approaches like Ant Colony Optimization (ACO), even if

some optimize hyperparameters or select features. Furthermore, one of the biggest challenges still facing early-stage cancer diagnosis is the issue of model resilience when dealing with complicated and unbalanced information.

By combining several potent classifiers, such as KNN, SVM, Random Forest, Gradient Boosting, and XGBoost, this suggested study fills in the gaps in the current literature and builds a more reliable and accurate predictive model for breast cancer. This method integrates many classifiers to capture a wider range of intricate patterns in the data, in contrast to earlier research that relied on single classifiers or less sophisticated optimization techniques. Furthermore, the model's performance is improved by applying Ant Colony Optimization (ACO) for hyperparameter tweaking, which solves the optimization problem that many conventional approaches have. In comparison to current methods, this model seeks to increase diagnostic accuracy and generalization by utilizing sophisticated optimization and classifier integration, offering a more robust solution for breast cancer prediction.

MATERIALS AND METHODS

This study proposes a machine learning model for breast cancer prediction, utilizing data sourced from the UCI Machine Learning Repository. Various classifiers such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Gradient Boosting, and

XGBoost have been implemented to construct the predictive model. Each classifier was selected for its demonstrated effectiveness in classification tasks and ability to identify complex patterns in medical data. Following model construction, Ant Colony Optimization (ACO) was applied to fine-tune parameters, further enhancing model performance. The structural flow and processes of the proposed model are illustrated in the design diagram (Fig.1.), providing a clear overview of the integration of classifiers and optimization techniques. The data has been collected, then data preprocessing has been done, followed by applying different classifiers. After that ACO has been applied as shown in the Figure-1. This approach aims to improve diagnostic accuracy, contributing valuable insights to the field of machine learning-based disease prediction.

Data Collection

The dataset utilized to train the model has been collected from the UCI data repository.²⁴ This dataset is made up of clinical and diagnostic characteristics that were taken from digital images of breast mass lesions using fine needle aspirations (FNA). Thirty parameters, including texture, radius, perimeter, and symmetry, are included in the dataset, which consists of 569 examples. Machine learning (ML) models can be trained and evaluated with ground truth information provided by the labeling of each instance as benign or malignant. To guarantee the precision and applicability of the features taken from the breast mass pictures, technicians and

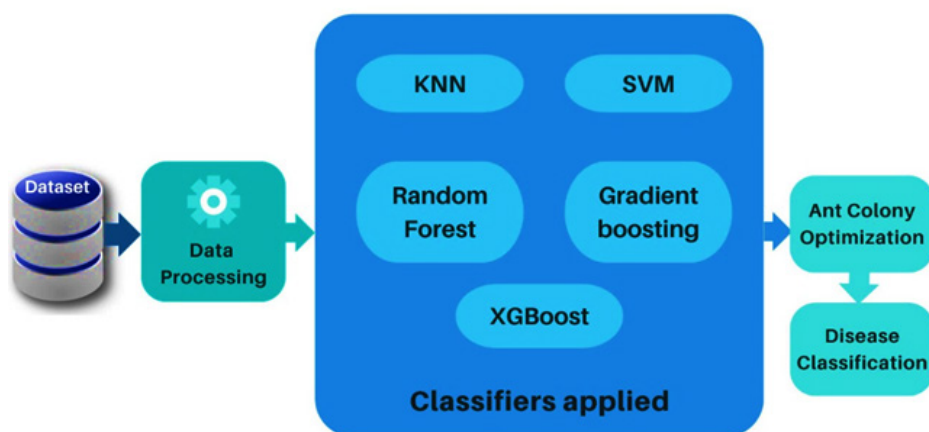


Fig. 1. Design diagram of the proposed model

medical specialists worked together during the data collecting process. The primary fields in the dataset include various attributes relevant to breast cancer diagnosis. The *Diagnosis* field categorizes cases as either malignant (M) or benign (B). Key features include *radius*, representing the mean distance from the center to points on the tumor perimeter, and *texture*, which captures the standard deviation of gray-scale values. Additional attributes consist of *perimeter* and *area*, both fundamental to tumor size and shape assessment. The dataset also includes *smoothness*, indicating the local variation in radius lengths, and *compactness*, a measure of shape regularity. Furthermore, *concavity* and *concave points* describe the extent and number of concave regions along the tumor boundary, respectively. *Symmetry* and *fractal dimension*, the latter reflecting “coastline approximation” minus one, provide further structural information. Together, these features offer a comprehensive profile for distinguishing between malignant and benign cases.

Data Pre-processing

The data pre-processing module is an acute component in the development of ML models.

Converting Categorical filed into Numerical filed

In the above-mentioned dataset, the dependent variable diagnosis is a categorical filed representing either by ‘M’ or ‘B’. This filed has been converted into numerical value applying Label Encoding. In machine learning, label encoding is a technique that transforms category data into numerical form so that models can process it. This approach is a straightforward and effective technique to manage categorical data since it gives each separate category a unique integer value. When categorical data contains an intrinsic order, label encoding is especially helpful since it maintains this order, which is desirable for some algorithms, such as tree-based models. Moreover, it consumes less memory and processing power than more sophisticated techniques like one-hot encoding.

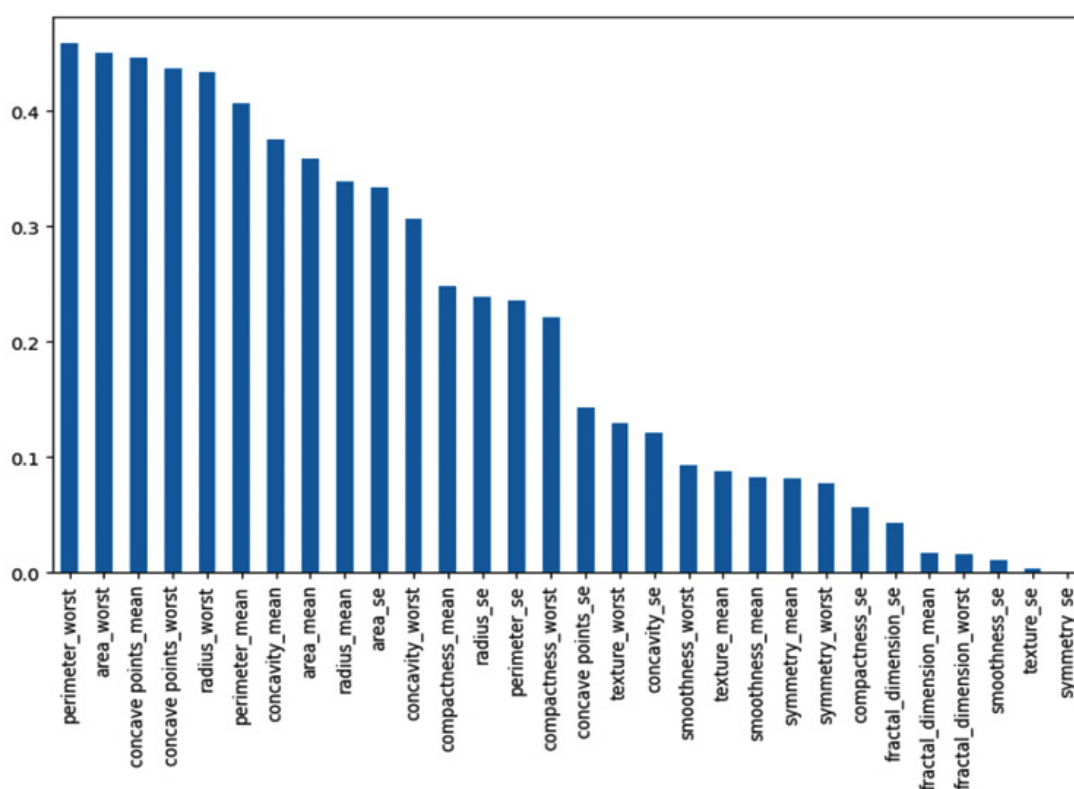


Fig. 2. Feature importance applying SelectKBest

Feature Selection

Another important part of building any machine learning model is feature selection. All the existing features are doing not contribute equally in decision making and sometimes presence of unnecessary feature may degrade the model's performances. In this work SelectKBest has been applied for the feature selection.

SelectKBest, a popular machine learning feature selection strategy, helps enhance model performance by choosing the best "K" features according to their results. This approach determines which aspects are most relevant to the target variable by calculating a score of each feature and then choosing the features with highest scores. SelectKBest minimizes the dimensionality of the data by emphasizing the most significant features. This reduces noise and the chance of overfitting, resulting in more precise and broadly applicable models. SelectKBest is very helpful when working with huge datasets because it can

considerably speed up the training process by minimizing the number of features. Additionally, it simplifies the generated models, which facilitates management and interpretation. SelectKBest's adaptability to many machine learning issues stems from its capacity to operate with multiple statistical tests according on the type of data. Its popularity is further enhanced by its simplicity of implementation, which makes it possible to incorporate it effortlessly into machine learning workflows for effective and efficient feature selection.

Applying SelectKBest top 18 features have been selected for training the model. The feature importance applying SelectKBest has been shown in the Figure 2. It has been identified that perimeter_worst has the highest priority where symmetry_se has the lowest priority.

Applying Classifiers

Multiple classifiers have been applied for prediction of breast cancer as follows:

Table 1. Pseudo code of KNN Classifier

```
Knn_clf=KNN_Classifier(No of neighbours, p,weights)
Knn_clf.fit(x_train_selected,y_train)
y_prediction=Knn_clf.predict(x_test,y_prediction)
accuracy=accuracy_score(y_test,y_prediction)
print(accuracy)
```

Table 2. Pseudo code of SVC Classifier

```
Svc_clf=SVC(C=100,gamma='scale',kernel='liner',random_state=42)
Svc_clf.fit(x_train_selected,y_train)
y_prediction=Svc_clf.predict(x_test_selected)
accuracy=accuracy_score(y_test,y_prediction)
print(accuracy)
```

Table 3. Pseudo code of SVC Classifier

```
Rf_clf=RandomForestClassifier(max_depth,min_samples_leaf,min_samples_split,n_estimators,random_state)
Rf_clf.fit(x_train_selected,y_train)
y_prediction=Rf_clf.predict(x_test_selected)
accuracy=accuracy_score(y_test,y_pred)
print(accuracy)
```

Table 4. Pseudo code of XGBoost Classifier

```
Xgb_clf=xgb.XGBClassifier(learning_rate,n_estimators,random_state)
Xgb_clf.fit(x_train_selected,y_train)
y_prediction=Xgb_clf.predict(x_test_selected)
accuracy=accuracy_score(y_test,y_prediction)
```

KNN Classifier

The KNN is a simple learning algorithm used for classification and regression tasks especially for instance based. It classifies data points based on the majority vote of their KNN in the feature-space. The designing of KNN model has been shown in the Table-1 below.

Support Vector Classifier (SVC)

The Support Vector Classifier (SVC) is a supervised machine learning algorithm widely applied in classification tasks. Its primary goal is to identify the optimal hyperplane that separates different classes within the feature space, ensuring maximum margin between the classes. By maximizing this margin, SVC improves classification accuracy and generalizes effectively to new data. The design process and functional architecture of SVC is illustrated in Table-2, demonstrating how the algorithm delineates the decision boundary between classes. This approach is particularly valuable in cases where a clear

separation between classes is crucial for accurate predictions.

Random Forest

Random Forest is an ensemble learning technique that builds multiple decision trees during the training process and makes predictions based on the majority vote of these trees. This method employs bagging (Bootstrap Aggregating) and random feature selection, which help mitigate overfitting and enhance model accuracy. By averaging the results from diverse trees, Random Forest achieves robust performance and generalizes well to new data. Table 3 illustrates the design of the K-Nearest Neighbor (KNN) model, showcasing its approach to classification based on the proximity of data points in the feature space. Both Random Forest and KNN offer distinct, effective strategies for handling classification tasks.

XGBoost

XGBoost, or Extreme Gradient Boosting, is a highly optimized library built to enhance the

Table 5. Pseudo code of Gradientboosting Classifier

```

Gb_clf=GradientBoostingClassifier(learning_rate,max_depth,n_estimator, random_state)
Gb_clf.fit(x_train_selected,y_train)
y_prediction=Gb_clf.predict(x_test_selected)
accuracy=accuracy_score(y_test,y_prediction)

```

Table 6. Pseudo code of Gradientboosting Classifier

```

Aco=aco(fitness_function,dimensions=X.shape[1],colony_size,max_iter)
best_solution=Aco.Optimize()

```

Table 7. Overall experimental Outcomes of the proposed work

Algorithm	Best Parameter	Accuracy	AUC	Sensitivity	Specificity
KNNWith ACO	{'n_neighbors':7,'p':1, 'weights': 'distance'}	94%	0.9231	0.9859	0.8604
SVC with ACO	{'C':100, 'gamma': 'scale', 'kernel': 'linear'}	96%	0.9464	0.9859	0.9069
Rando mForestwith ACO	{'max_depth':20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators':50}	99%	0.9883	1	0.9767
Gradient Boosting Classifier with ACO	{'learning_rate': 0.1,'max_depth':3, 'n_estimators':200}	96%	0.951	0.9302	0.9718
XG Boost with ACO	{'learning_rate': 0.1, 'n_estimators':100}	95.60%	0.951	0.9718	0.9302

gradient boosting framework. It is specifically engineered for efficient, scalable implementation of gradient boosting algorithms, renowned for delivering exceptional performance and accuracy. The design and architecture of the XGBoost model are illustrated in Table-4.

Gradient Boosting

Gradient Boosting is a machine learning approach that constructs an ensemble of weak learners, often decision trees, in a sequential manner. In this process, each new learner is added to the ensemble to reduce the loss function,

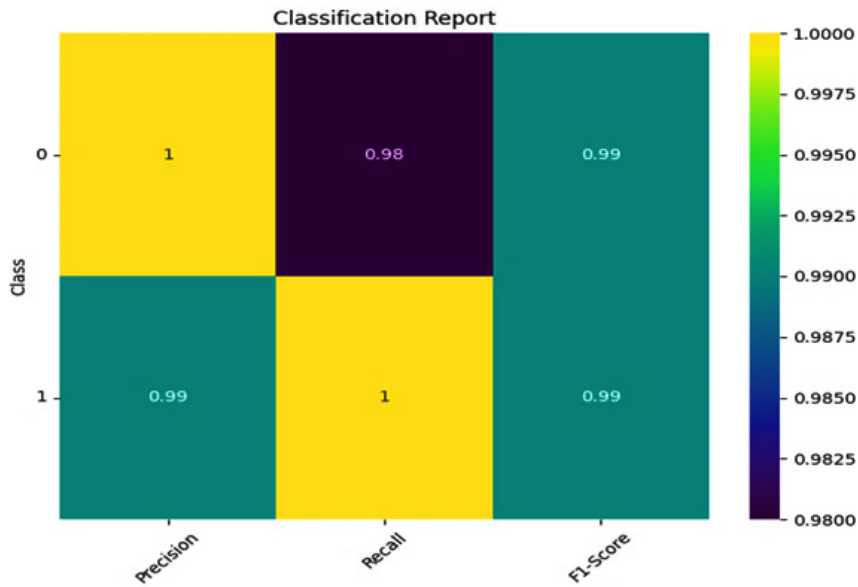


Fig. 3(a). Confusion Matrix of the Model utilizing Random Forest and ACO

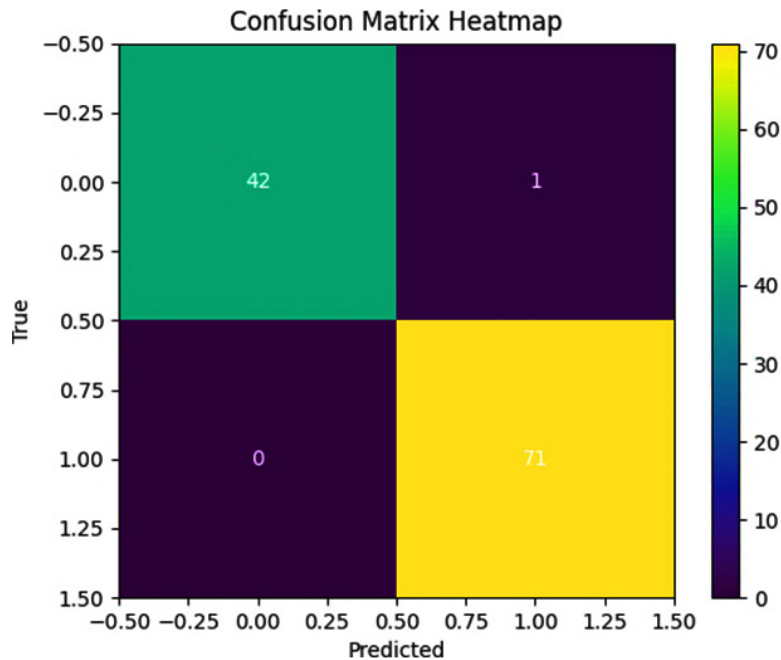


Fig. 3(b). Classification report of the Model utilizing Random Forest and ACO

iteratively enhancing the model's prediction accuracy. The structure of the Gradient Boosting model is illustrated in Table-5.

Mode Optimization

For performance optimization Ant Colony Optimization has been applied. Some ant species' foraging strategies serve as the foundation for ant colony optimization, or ACO. When these ants locate a good path, they mark it for other ants in the colony to follow by leaving pheromone markers

on the ground. An analogous approach is used in ant colony optimization to address optimization issues. Within the context of swarm intelligence techniques, this algorithm belongs to the family of ant colony algorithms and represents a few metaheuristic optimizations. The first algorithm was first presented by Marco Dorigo in 1992 in his PhD thesis.^{25,26} Its goal was to find the best path through a graph by mimicking the actions of ants as they navigated a path among their colony and the

Table 8. Comparative Analysis

Works	Observations
Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques ⁵	In this study, lung cancer prediction was explored using a combination of machine learning models, including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest. Among these methods, the highest prediction accuracy achieved was 97.14%.
Prediction of Survival Rate from Non-Small Cell Lung Cancer using Improved Random Forest ⁷	Another approach for lung cancer prediction employed Support Vector Machine (SVM), Naïve Bayes, Decision Tree, and Random Forest models. The highest accuracy recorded in this combination was 96.8%, showcasing the effectiveness of ensemble methods for disease prediction.
A novel transfer learning approach for the classification of histological images of colorectal cancer ⁸	For colorectal cancer prediction, a blend of Naïve Bayes, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbor (KNN) algorithms was applied. This approach yielded a top accuracy of 92.83%, highlighting the models' effectiveness in handling colorectal cancer datasets.
Classification using deep learning neural networks for brain tumors ¹⁰	In this analysis, Deep Neural Networks (DNN), K-Nearest Neighbor (KNN) with parameters $k=1$, $k=1$ and $k=3$, Linear Discriminant Analysis (LDA), and Sequential Minimal Optimization (SMO) were applied to predict cancer. The model achieved an Area Under the Curve (AUC) of 0.97, indicating high discriminatory power for classification.
Diagnosing COVID-19 Infection in chest X-Ray Images using Neural Network ¹⁴	A prediction model for COVID-19 diagnosis was developed using Support Vector Machine (SVM) and a Neural Network. This model achieved an impressive Area Under the Curve (AUC) of 0.99, reflecting its strong capability for accurate COVID-19 prediction.
Heart Disease Prediction Using Concatenated Hybrid Ensemble Classifiers ¹⁸	For heart disease prediction, a concatenated ensemble classifier was implemented, which combines multiple classifiers to improve accuracy. This method resulted in an overall accuracy of 86.89%, demonstrating the utility of ensemble approaches for complex health condition prediction.
Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques ²¹	Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms were applied for disease prediction, achieving a peak accuracy of 95.56%. This indicates the models' suitability for robust prediction in medical datasets.
Breast Cancer Prediction Using Machine Learning Techniques ²⁷	To enhance the prediction of lung cancer, this research applied Support Vector Machine (SVM), Decision Tree, Naïve Bayes, and K-Nearest Neighbor (KNN) algorithms. This combination of models reached a peak accuracy of 97.13%, demonstrating strong predictive potential in lung cancer diagnosis.
Our Proposed Model	A comprehensive model for breast cancer prediction incorporated K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Gradient Boosting, XGBoost, and Ant Colony Optimization algorithms. The highest accuracy achieved was 99%, underscoring the effectiveness of this diverse ensemble for breast cancer diagnosis.

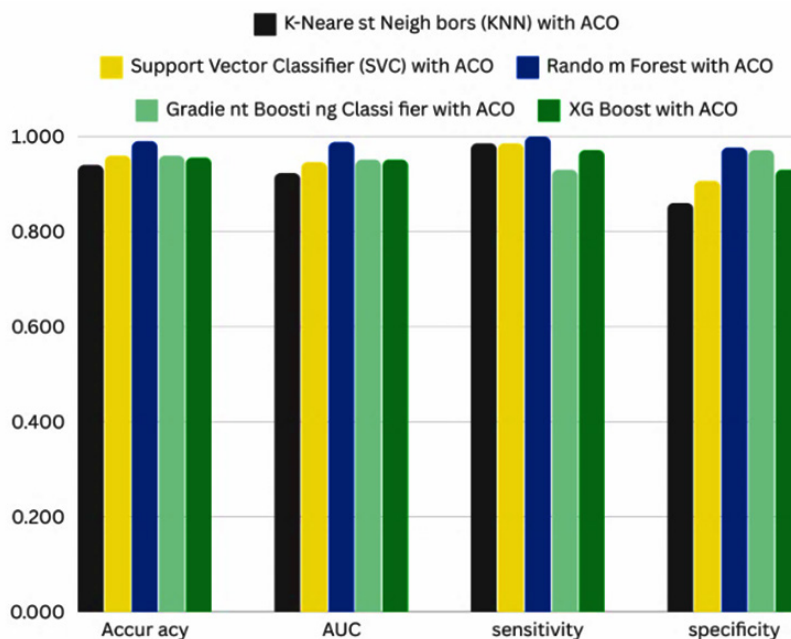


Fig. 4. Overall outcome metrics

food source. Sample code snapshot of designing ACO has been represented in Table-6 below

RESULTS

In the proposed model Ant Colony Optimization has been applied over KNN, SVM Classifier, Gradient Boosting, Random Forest, and XGBoost. It has been identified that Random Forest Classifier with Ant Colony Optimization has achieved highest accuracy of 99%. The details observations of investigational results have been shown in the Table 7 below.

DISCUSSION

Combining Random Forest (RF) with Ant Colony Optimization (ACO) to achieve 99% accuracy in predictive modeling represents a major machine learning success, with several important consequences as shown in Figure 4. First, the model's consistent performance across a range of data examples demonstrates its resilience. This indicates that the model can generalize well to new data, which increases its dependability for real-world applications. Furthermore, the model's

high prediction accuracy gives decision-makers great confidence and guarantees that the model's outputs may be relied upon for important decisions in a variety of industries, including banking and healthcare. The model is useful for real-time and high-throughput environments because to its efficiency and scalability, which are shown by its capacity to handle massive datasets and give correct predictions in reasonable timescales. finally, the high accuracy has significant ramifications for decision-making procedures.

By providing more dependable early detection, individualized therapy, and improved predictive analytics, Random Forest with ACO models with 99% accuracy have the potential to completely transform clinical practice in the healthcare industry. The model has the potential to detect diseases like cancer early with high accuracy, which could result in better outcomes and earlier therapies. Additionally, it would help physicians minimize human error, optimize resources, and customize treatment approaches for each patient. Furthermore, Random Forest combined with ACO may improve decision-making and overall healthcare efficiency by more correctly predicting patient outcomes.

It can maximize results in areas such as fraud detection, resource allocation, and customized suggestions, finally resulting in better informed, significant, and impacting decisions.

The comparative analysis of the proposed work with few existing works have been shown in the Table 8.

CONCLUSION

Machine learning is a convenient tool for identification since it analyses large amounts of data. AnML based breast cancer expectation model is presented in this study. There have been several classifiers utilized, including KNN, SVM, Gradient Boosting, Random Forest, and XGBoost. To enhance performance, ACO has been applied to these models. In conclusion, the model's superior learning and generalization abilities are demonstrated by its 99% accuracy with Ant Colony Optimization with Random Forest. It gives decision-makers an effective instrument to make highly confident and precise data-driven decisions, opening up new possibilities and promoting improvements across a range of industries.

ACKNOWLEDGEMENT

Authors would like to thank JIS college of Engineering, JISGROUP, for providing all kinds of R&D resources and encouraging for R&D activities.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest

The authors do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials.

Permission to reproduce material from other sources

Not Applicable.

Authors' Contribution

Annwesha Banerjee: Conceptualization, Methodology, Writing – Original Draft; Aniruddha Biswas and Avisekh Kumar Tiwari: Data Collection, Analysis, Writing – Review & Editing; Sumit Das: Visualization, Supervision, Project Administration.

REFERENCES

1. Breast cancer- World Health Organization- <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (Accessed on 23.01.2025)
2. A. Colomi, M. Dorigo et V. Maniezzo, Distributed Optimization by Ant Colonies, actes de la première conférence européenne sur la vie artificielle, Paris, France, Elsevier Publishing, 134-142, 1991.
3. Garba, A., Khalid, S., Aleryni, A., Ullah, I., Tairan, N.M., Shah, H. and Mumin, D. 2024. Utilizing Ant Colony Optimization for Result Merging in Federated Search. *Engineering, Technology & Applied Science Research*. 14, 4 (Aug. 2024), 14832–14839. DOI:<https://doi.org/10.48084/etasr.7302>.
4. Yilmaz Eroglu, D., & Akcan, U. (2024). An Adapted Ant Colony Optimization for Feature Selection. *Applied Artificial Intelligence*, 38(1). <https://doi.org/10.1080/08839514.2024.2335098>
5. Islam M, Haque M, Iqbal H, Hasan MM, Hasan M, Kabir MN. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Comput Sci*. 2020;1:290. doi:10.1007/s42979-020-00305-w
6. Banerjee N, Das S. Prediction Lung Cancer– In Machine Learning Perspective. In: *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*. ; 2020:1-5. doi:10.1109/ICCSEA49143.2020.9132913
7. Nanda P, Duraipandian N. Prediction of Survival Rate from Non-Small Cell Lung Cancer using Improved Random Forest. In: *2020 International Conference on Inventive Computation Technologies (ICICT)*. ; 2020:93-97. doi:10.1109/ICICT48043.2020.9112558
8. Ohata EF, Chagas JVS das, Bezerra GM, Hassan MM, de Albuquerque VHC, Filho PPR. A novel transfer learning approach for the classification of histological images of colorectal

- cancer. *J Supercomput.* 2021;77(9):9494-9519. doi:10.1007/s11227-020-03575-6
9. Li H, Lin J, Xiao Y, Colorectal Cancer Detected by Machine Learning Models Using Conventional Laboratory Test Data. *Technol Cancer Res Treat.* 2021;20:15330338211058352. doi:10.1177/15330338211058352
 10. Amin J, Sharif M, Haldorai A, Yasmin M, Nayak RS. Brain tumor detection and classification using machine learning: a comprehensive survey. | Complex & Intelligent Systems | EBSCOhost. doi:10.1007/s40747-021-00563-y
 11. Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM. Classification using deep learning neural networks for brain tumors. *Future Comput Inform J.* 2018;3(1):68-71. doi:10.1016/j.fcij.2017.12.001
 12. Khan AH, Abbas S, Khan MA, Intelligent Model for Brain Tumor Identification Using Deep Learning. *Appl Comput Intell Soft Comput.* 2022;2022(1):8104054. doi:10.1155/2022/8104054
 13. Av^oar E, Salçin K. Detection and classification of brain tumours from MRI images using faster R-CNN. *Teh Glas.* 2019;13(4):337-342. doi:10.31803/tg-20190712095507
 14. Zaki SM, Jaber MM, Kashmoola MA. Diagnosing COVID-19 Infection in Chest X-Ray Images Using Neural Network. *Baghdad Sci J.* 2022;19(6):1356-1356. doi:10.21123/bsj.2022.5965
 15. Thapa BKH Himel Chand. Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images. Seventh Sense Research Group. Accessed September 30, 2024. <https://dev.ijcttjournal.org/archives/ijctt-v68i10p104>
 16. Das S, Saha T, Nath I, Dipansu M. Exploring Machine Learning Methods for Developing a Predictive System for Parkinson's Disease. *Biosci Biotechnol Res Asia.* 2024;21:569-582. doi:10.13005/bbra/3248
 17. Research Scholar, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, Andhra Pradesh 522510, India., Ponnada VT, Srinivasu DrSVN, Professor, Computer science and engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh 522601, India. Efficient CNN for Lung Cancer Detection. *Int J Recent Technol Eng IJRTE.* 2019;8(2):3499-3503. doi:10.35940/ijrte.B2921.078219
 18. Majumder AB, Gupta S, Singh D, Heart Disease Prediction Using Concatenated Hybrid Ensemble Classifiers. *Algorithms.* 2023;16(12):538. doi:10.3390/a16120538
 19. Majumder AB, Gupta S, Majumder S, Singh D. A Heart Disease Prediction Model using Merged XGBoost-SVM Classifier and Particle Swarm Optimization. In: *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI).* ; 2024:241-248. doi:10.1109/ICMCSI61536.2024.00042
 20. Bhandari A, Majumder AB, Das S. An Intelligent System for Prediction of Lung Cancer Under Machine Learning Framework. In: Sharma N, Goje AC, Chakrabarti A, Bruckstein AM, eds. *Data Management, Analytics and Innovation.* Springer Nature; 2024:27-43. doi:10.1007/978-981-97-3242-5_3
 21. Abdullah D, Abdulazeez A, Sallow A. Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques. *Qubahan Acad J.* 2021;1:141-149. doi:10.48161/qaj.v1n2a58
 22. Das S, Koley S, Saha T. Machine Learning Approaches for Investigating Breast Cancer. *Biosci Biotechnol Res Asia.* 2023;20:1109-1131. doi:10.13005/bbra/3163
 23. Das S, Nath I, Dey D, An Intelligent Approach for Detrimental Emotion Classification and Healthcare Management. In: *2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCGST).* ; 2024:1-5. doi:10.1109/ICCGST60741.2024.10717562
 24. William Wolberg OM. Breast Cancer Wisconsin (Diagnostic). Published online 1993. doi:10.24432/C5DW2B
 25. Colomi A, Dorigo M, Maniezzo V. *Distributed Optimization by Ant Colonies.*; 1991.
 26. Dorigo M. Optimization, learning and natural algorithms. *Ph Thesis Politec Milano.* Published online 1992. Accessed September 30, 2024. <https://cir.nii.ac.jp/crid/1573950400977139328>
 27. Apoorva V, Yogish HK, Chayadevi ML. Breast Cancer Prediction Using Machine Learning Techniques. In: Atlantis Press; 2021:348-355. doi:10.2991/ahis.k.210913.043