International Journal of Basic and Applied Sciences, 14 (SI-1) (2025) 624-631

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/qs9zrt46
Research paper

Evolutionary Algorithm Pipeline for Improving in Prediction of Earlier Diagnosis of Breast Cancer

G. Sridevi [1] *, K. Sharmila [2]

[1] Research Scholar, Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India
[2] Research Supervisor and Associate Professor in the Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India
**Corresponding author E-mail: srimani1984@gmail.com*

Received: July 15, 2025, Accepted: July 24, 2025, Published: November 1, 2025

Abstract

The cells in the breast have been developed in terms of cancer named Breast Cancer (BC), and over the past decade, BC has been the most identified type of cancer in women. Currently, the rapid increase in the predominance of BC may lead to increased mortality worldwide. The disorder is diagnosed manually, which requires efficient expertise and time, both of which are significant in detecting cancer earlier and informing subsequent treatments. There are various Machine Learning (ML) methods that assist in making several decisions as well as performing diagnoses from the collected data, but have failed in accomplishing high prediction in diagnosing BC. This paper focuses on investigating the probability of Ensemble Learning (EL) with Evolutionary Algorithm (EA) for risk prediction and BC categorization, precisely through exact feature selection and extraction. Moreover, the key focus is on creating a consistent and precise model by associating the best features of various learning algorithms and ensembling the metadata sets. Therefore, this research focuses on analyzing methods to improve the accuracy of BC diagnosis through a proposed ensemble Artificial Neural Network (ANN) method with a Keras-based approach. Hence, the evaluated results propose a method determined through accuracy metrics, which are compared with the Boruta method and the ANN method. The accuracy of the Keras-based ensemble ANN method, at 97.37%, is higher than that of other ML methods.

Keywords: *Breast Cancer; Artificial Neural Network; Ensemble Learning; Machine Learning; Evolutionary Algorithm.*

1. Introduction

Tumors are masses that form in the breast when specific cells start growing out of control [1]. This is the fundamental symptom of BC, and the diagnosis is often defined as malignant (cancerous) or benign (non-cancerous). The possibility of transferring malignant tumors through the circulation or lymphatic systems to other areas of the body makes them particularly dangerous [2]. BC symptoms can range from tiny tumors in the breast to changes in the breast's shape or color. Early diagnosis is crucial for detecting BC, as there are numerous forms of the disease, and most of them don't initially cause symptoms. One form of cancer that develops in the region of aberrant milk-producing cells in the breast is called lobular carcinoma. People with invasive lobular carcinoma, a condition that develops in the milk-producing mammary gland lobules, may notice changes in their skin texture, breast enlargement, and thickening of the tissue around their breasts. Approximately 80% of cases, which are caused by the most common type of cancer, are invasive ductal carcinoma in situ, which is often asymptomatic and detected by mammography [3] [4]. There is strong evidence that drinking alcohol, aging, having heavy breasts, a family history of the condition, receiving radiation therapy, being fat, and being exposed to radiation all raise the risk of BC [5]. Conversely, research has discovered that extended breastfeeding, engaging in physical activity, abstaining from harmful alcohol consumption, quitting smoking, and avoiding long-term hormone use all reduce one's chance of BC [6].

BC is a significant health concern for females, accounting for one in eight deaths [7]. Around 2.3 million cases are detected each year, making it the most common cancer among humans. According to the WHO report, 95% of women died because of BC, either as the primary or secondary cause. Nonetheless, there are notable variations in BC survivability both within and across nations. In low- and middle-income countries, BC and cervical cancer account for about 80% of all cancer-related fatalities. However, for the next 20 years, it has been estimated that there would be 22 million cancer cases worldwide, an increase from 14 million cases, and that it would then progressively rise every year after that. The mortality rate for all malignancies has grown because the disease spreads to different parts of the body from where it first developed. Over 2.3 million women were impacted by the 685,000 global deaths caused by BC in 2020. BC was the most common cancer worldwide by 2020, having been identified in 7.8 million women in the five years prior. The occurrence of BC increases globally when women attain puberty; they can get BC regardless of age [8]. Moreover, 16.6% of Indonesians with a diagnosis of BC pass away [9], affecting 40.3% of the country's population. BC risk factors include inadequate nutrition, heavy drinking, and smoking. In 2030, there will be a 2% rise in the incidence of BC. By enabling patients to obtain therapy promptly, early detection of BC can considerably enhance survival chances and prognosis [10].

Several studies used different ML methods to study BC prediction. The RF and ET techniques employ Decision Trees (DT) as appropriate classifiers to achieve the final classification, which addresses the researcher's problem. The efficacy and efficiency of every algorithm's data categorization were assessed in this study [11]. To correctly identify patients with BC, the author introduced a voting classifier based on ensemble learning that integrates the stochastic gradient descent and Logistic Regression (LR) classifiers. The present investigation is also motivated by the fact that the classification and diagnosis of BC is still the exclusive domain of the ensemble classifier approach employed in the earlier research [12]. Accurately and promptly detecting various medical conditions is a significant problem facing healthcare research. The objective of this study is to propose an innovative approach to creating an EA platform that offers the flexibility required for rapid professional prototyping and a low learning curve, which is essential for most industry players and instructors. The reason behind the creation of the LEAP is the operator-centric approach to EA design. The distinguishing feature of LEAP is its ability to effectively explain the operator-based pipeline at the core of an algorithm by utilizing Python's assistance with practical functional syntax.

2. Literature Review

Globally, BC is the leading cause of mortality for many people. Modern technology allows professionals to diagnose women with BC utilizing a variety of coping mechanisms in addition to traditional cancer detection techniques. A proficient doctor achieved a 79.97% correct diagnosis rate. In contrast, ML attained an accuracy of 91.1% in its predictions. A significant difficulty is choosing the best classifier among ML algorithms, which are utilized for predicting a precise diagnosis of BC. Vikas Chaurasia and SaurabhPal have used the Wisconsin Diagnostic BC(WDBC) Database to predict BC along with ML [13]. Six algorithms were examined, and their characteristics were narrowed down to 12 using statistical techniques. The models were then combined using ensemble methods. The test results yielded an accuracy of above 90% for all algorithms, indicating their good performance, particularly in the improved feature area, as indicated by the results. Among his accomplishments are the application of ensemble techniques and feature selection to increase the precision of BC prediction. Kabiraj et al. have developed a BC risk prediction framework using the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms [14]. The dataset utilized in this work was acquired from the UCI ML Repository. Using XGBoost and RF techniques, the classification strategy achieved an accuracy of 74.73% for the model. MeerjaJabbar et al. have introduced a novel ensemble technique for classifying BC data that uses a Bayesian Network and the Radial Basis function [15]. This technique outperforms previous methods with an accuracy rate of 97%.

Several variables were used to assess trial success on the Wisconsin BC Dataset (WBCD). The suggested group research can help doctors accurately diagnose tumors and assist patients in choosing their courses of therapy. Shalini and Radhika are attempting to use several ML approaches to diagnose BC [16]. ANN, SVM, Naive Bayes, DTs, and algorithms are employed, along with the UCI ML database. Consequently, a categorization accuracy of 86% was obtained. Naji et al. have tried to use ML approaches to diagnose and predict BC [17]. Five different algorithms were examined: DT (C4.5), LR, SVM, RF, and KNN. The Wisconsin BC diagnostic database was used for this analysis. The main goal is to identify the ideal algorithm for BC diagnosis. The SVM achieved the highest accuracy of 97.2% and outperformed the other classifiers, according to the findings. Essential insights are gained to upgrade BC treatment and enhance patient safety measures through research conducted in the Anaconda Python environment using the scikit-learn toolkit.

Benbrahim et al. have utilized the WBCD dataset to develop an ML-based method and evaluated the classification accuracy. Their study's conclusions showed that the NN had a remarkable 96.49% accuracy rate [18]. Ghiasi et al. have classified the types of BC by using the RF and other tree approaches. The approaches suggested aim to provide a simple and efficient graphical methodology for the categorization [19]. The WBCD provides real data covering the most critical parameters needed to build a BC diagnostic model. Alshayeji et al. have utilized the WBCD and WDBC in an ANN model without the use of selection techniques, allowing for the diagnosis and prediction of BC [20]. Based on the WBCD and WDBC datasets, the ANN performed well in classifying both malignant and benign tumors. The fourth approach used by Kwang and Eun et al., for predicting BC is the Ensemble Classifier (EC) with a voting mechanism; however, it has not received much attention from researchers. The voting classifier is an example of an ensemble technique in BC to differentiate between benign and malignant tumors. Two voting stages were developed for each of the three ML models.

The fundamental categorization model proved to be effective in several voting systems [21] [22]. Ziba et al. have stated that an article published in Nature Cancer has offered a new method for differentiating between cancer tissues and their corresponding benign counterparts [23]. Abdur et al. have suggested that the SVM classifier has been widely utilized in research to predict BC, even though few studies have relied solely on it. However, further study is needed to develop better classifiers for predicting BC [24]. Gary and Andries have leveraged Python's support for fundamental functional programming principles, significantly enhancing the readability of LEAP. Moreover, it should be noted that several other language research platforms for EA experiments (like Scala) only provide functional approaches for describing algorithms in terms of algebraic data types (such as semigroups and monoids) [25]. Although earlier studies, such as those by Shalini and Radhika [16], achieved 86 percent accuracy, they had a constraint of limited feature optimization. Other approaches, similarly, had problems with unbalanced data or suboptimal feature choices [13]. Our LEAP-ANN model directly addresses these issues by combining evolutionary feature-extraction algorithms and Keras-based ANN ensembles, thereby leading to improved accuracy and generalizability.

3. Research Gap

Traditional ML techniques often employ a single model, which can lead to overfitting when dealing with complex and high-dimensional data, a significant limitation of the present studies. Furthermore, traditional methods may struggle to depict complex connections or patterns in the data. On the other hand, by combining multiple base models and utilizing their combined intelligence to enhance the anticipated accuracy and resilience, EML techniques can overcome these limitations. The accuracy and proficiency of BC diagnosis systems may be improved by improving BC recognition and categorization models with extensive EML and feature selection algorithms. This would enable earlier identification and more efficient treatment. When used for multi-class classification tasks, this design proved to be difficult. These works address challenges such as feature selection, imbalanced classes, and the complexity of ensemble approaches, providing insights into the advantages and disadvantages of various ML algorithms for BC classification. The results may contribute to the development of more accurate and reliable BC diagnostic instruments, thereby improving treatment and care for patients. Finding the most relevant and informative characteristics within the large pool of available variables requires feature selection. Selecting the most distinct attributes for the BC identification and classification model can enhance its performance and interpretability while reducing dataset complexity and minimizing the chances of overfitting. Finally, an adaptive algorithm should be used to provide a dependable way for more precise diagnosis and monitoring of BC.

4. Research Methodology

Earlier detection of BC may minimize the complexity of a patient's life, which can even reduce the mortality rate of BC. This research focuses on improving the accuracy of predicting BC in its early stages of cancer. Based on the blood test report and CT scan reports, experts have identified the value of atypical cell sizes and their characteristics, which differ from those of normal cells. The radiologist reviews the CT scan values of the patients, as listed in the report, and then discusses them with an expert. This is the basic scenario, but obtaining an appointment with the best experts has become complex, which can even prolong the time it takes to diagnose the complexity present in the cancer cells. However, automating the obtained results through various ML methods may generate some precise diagnoses, but not as accurate. Therefore, this research concentrated on improving diagnostic precision during the early stage of BC using the LEAP concept for better feature selection and extraction.

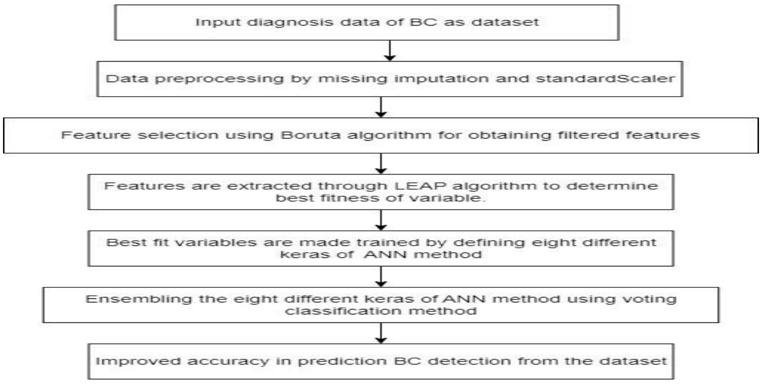

Fig. 1: Architecture of Ensembling ANN with LEAP Algorithm for Predicting BC Detection.

In preprocessing, Boruta is used for feature selection to identify relationships within a dataset, and LEAP is employed for enhanced feature extraction to discover correlations among features, as well as to optimize parameters. Moreover, the Keras-based ANN model is used in this study to identify the various layers that maximize the model's performance through a more effective ensembling technique. During the execution of the training dataset, eight different Keras-based ANN models are generated for training the data, resulting in ensemble models using a voting classifier.

5. Data Collection

The collected data was obtained from California, Irvine University ML repository as the source dataset. The collected data consist of 32 features and 569 records. The target feature is named 'diagnosis', and the independent features are radius_mean, texture_mean, concavity_mean, etc. The data has been split into an 80:20 ratio for training and testing from the BC dataset. The target variable is defined as categorical with a binary outcome, where the target is segregated into 'Benign' for non-cancerous patients and 'Malignant' for cancerous patients. The ML model is built through a training set. The results obtained for each ID with a diagnosis are considered the results for the respective patient in the training dataset. Table 1 illustrates the dataset collected for the WDBC. The dataset used in this study is the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, which is available at the UCI Machine Learning Repository. It contains 569 records about patients and 32 features extracted from digitized images of fine needle aspirates (FNA) of breast masses. This data will include 357 benign and 212 malignant cases. To ensure reproducibility, the data was divided into 20 percent testing and 80 percent training sets.

Table 1: WDBC Dataset (Sample)

radius_se	texture_se	perimeter_se	area_se	smoothness_se	compactness_se	concavity_se	concave points_se	symmetry_se	fractal_dimension_se
1.095	0.9053	8.589	153.4	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193
0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532
0.7456	0.7869	4.585	94.03	0.00615	0.04006	0.03832	0.02058	0.0225	0.004571
0.4956	1.156	3.445	27.23	0.00911	0.07458	0.05661	0.01867	0.05963	0.009208
0.7572	0.7813	5.438	94.44	0.01149	0.02461	0.05688	0.01885	0.01756	0.005115
0.3345	0.8902	2.217	27.19	0.00751	0.03345	0.03672	0.01137	0.02165	0.005082
0.4467	0.7732	3.18	53.91	0.004314	0.01382	0.02254	0.01039	0.01369	0.002179
0.5835	1.377	3.856	50.96	0.008805	0.03029	0.02488	0.01448	0.01486	0.005412
0.3063	1.002	2.406	24.32	0.005731	0.03502	0.03553	0.01226	0.02143	0.003749
0.2976	1.599	2.039	23.94	0.007149	0.07217	0.07743	0.01432	0.01789	0.01008
0.3795	1.187	2.466	40.51	0.004029	0.009269	0.01101	0.007591	0.0146	0.003042
0.5058	0.9849	3.564	54.16	0.005771	0.04061	0.02791	0.01282	0.02008	0.004144
0.9555	3.568	11.07	116.2	0.003139	0.08297	0.0889	0.0409	0.04484	0.01284
0.4033	1.078	2.903	36.58	0.009769	0.03126	0.05051	0.01992	0.02981	0.003002
0.2121	1.169	2.061	19.21	0.006429	0.05936	0.05501	0.01628	0.01961	0.008093
0.37	1.033	2.879	32.55	0.005607	0.0424	0.04741	0.0109	0.01857	0.005466
0.4727	1.24	3.195	45.4	0.005718	0.01162	0.01998	0.01109	0.0141	0.002085
0.5692	1.073	3.854	54.18	0.007026	0.02501	0.03188	0.01297	0.01689	0.004142
0.7582	1.017	5.865	112.4	0.006494	0.01893	0.03391	0.01521	0.01356	0.001997
0.2699	0.7886	2.058	23.56	0.008462	0.0146	0.02387	0.01315	0.0198	0.0023
0.1852	0.7477	1.383	14.67	0.004097	0.01898	0.01698	0.00649	0.01678	0.002425
0.2773	0.9768	1.909	15.7	0.009606	0.01432	0.01985	0.01421	0.02027	0.002968
0.4388	0.7096	3.384	44.91	0.006789	0.05328	0.06446	0.02252	0.03672	0.004394
0.6917	1.127	4.303	93.99	0.004728	0.01259	0.01715	0.01038	0.01083	0.001987
0.8068	0.9017	5.455	102.6	0.006048	0.01882	0.02741	0.0113	0.01468	0.002801
1.046	0.976	7.276	111.4	0.008029	0.03799	0.03732	0.02397	0.02308	0.007444

0.2545	0.9832	2.11	21.05	0.004452	0.03055	0.02681	0.01352	0.01454	0.003711
0.8529	1.849	5.632	93.54	0.01075	0.02722	0.05081	0.01911	0.02293	0.004217
0.439	1.012	3.498	43.5	0.005233	0.03057	0.03576	0.01083	0.01768	0.002967
0.6003	0.8225	4.655	61.1	0.005627	0.03033	0.03407	0.01354	0.01925	0.003742
0.8307	1.466	5.574	105	0.006248	0.03374	0.05196	0.01158	0.02007	0.00456
0.4825	1.03	3.475	41	0.005551	0.03414	0.04205	0.01044	0.02273	0.005667
0.6009	1.398	3.999	67.78	0.008268	0.03082	0.05042	0.01112	0.02102	0.003854
0.5558	0.6062	3.528	68.17	0.005015	0.03318	0.03497	0.009643	0.01543	0.003896
0.334	0.6857	2.183	35.03	0.004185	0.02868	0.02664	0.009067	0.01703	0.003817
0.4615	0.9197	3.008	45.19	0.005776	0.02499	0.03695	0.01195	0.02789	0.002665
0.286	1.019	2.657	24.91	0.005878	0.02995	0.04815	0.01161	0.02028	0.004022
0.1839	2.342	1.17	14.16	0.004352	0.004899	0.01343	0.01164	0.02671	0.001777
1.214	2.188	8.077	106	0.006883	0.01094	0.01818	0.01917	0.007882	0.001754

6. LEAP Pipeline as EA

The idea embodied by LEAP provides a pipeline that selects individuals from a series of potential parents, executes a series of operators or functions for each respective person, and then aggregates them to form a final series of offspring. To ensure that the operator chain is accepted and returns the appropriate types at each phase, while the checking of certain run-time types has also been illustrated and incorporated in Figure 2. This pipeline will produce the offspring in the following generation after just one pass. The pipeline is subsequently invoked again with these new parents as a collection, and the process gets iterated until a halting criterion is satisfied.

Fig. 2: Pipeline for LEAP Method.

The LEAP model focuses on generating antagonistic cases as the training dataset, using the WDBC dataset.

- The value of each feature is checked for null values in the dataset, and the best feature is selected using Boruta, which is considered the input for the parent data.
- The selection phase is used to choose original values from the dataset's features and set them as the threshold for determining the initial position.
- The clone phase is executed by a Modified Levy flight, which utilizes both initial velocity and initial position to determine the initial population of particles.
- The mutation phase executes the LEAP function and performs an iterative search using the victim model confidence score as the fitness function. This further updates the velocity and the position of the population particle.
- The evaluation phase performs greedy mutation with respect to the fitness score and change rate to identify the best individual in the population that satisfies the termination condition.
- The pool phase is involved in successful change, predicting the original text, and running until it reaches its maximum iteration.

In all LEAP algorithms, the pipe() is utilized as their basic design pattern, and the code of the LEAP snippet has been exemplified by toolzpipe() for generating offspring as a simplistic GA. The randomly generated individual sequence is initiated by parents chosen using a subsequent operator through binary selection in a tournament. The parent is made to clone due to no further disturbance in the population position and velocity by the next phase operation. Further, the clone is made mutated as well as evaluated to identify the best fitness of population particles. Finally, the offspring pool is the last operation is progresses until the iteration process matches, and it also performs as offspring parents. It is returned as the parent for the subsequent generation. In this research, the process is made to iterate for 50 generations. Establishing the initial position and velocity of the particles to search is the primary goal of population initialization. Since the goal of an adversary test method for delivering inputs may mislead the model into producing false predictions with high confidence, LEAP leverages the victim's model confidence for achieving the particle fitness. The optimization algorithm can look for parameters that are most probable to be classified by employing confidence as the fitness function. LEAP uses the fitness score to determine the initial position.

It's possible to express mathematically the LEAP integration with Keras ANN. Consider an input feature matrix X ∈ Rn x d, the evolutionary algorithm optimizes the feature subset X' as:

$$\mathrm{X}' = \arg\max_{\mathrm{S}\subseteq\mathrm{X}} \mathrm{f}(\mathrm{S}) \tag{1}$$

where f(S) is the classification accuracy of the ANN. The ANN forward pass is defined as:

$$\mathrm{y} = \sigma(\mathrm{WX}' + \mathrm{b}) \tag{2}$$

where W and b are weight and bias parameters optimized during training, and σ is the activation function. Pseudocode for the LEAP pipeline with ANN included is presented in Algorithm 1.

Algorithm 1

```
LEAP-ANN algorithm for feature extraction
Input: Tori: Original text, pop size: population number for each iteration; max iters: Max iterations.
Output: Tadv: test case for Adversarial.
Step 1: Initialize Tpop←Levy-Initialization(Tori) and if Tadv in Tpop, then return Tadv
Step 2: end if
Step 3: gBest=max{Tpop};
Step 4: lBest=copy{Tpop};
Step 5: when it doesn't exceed max iteration and adaptively set the inertia weight ω;
Step 6: for n in pop size do and Updating the particle n velocity and position;
 end for
Step 7: Current population evaluation and Greedy-Mutation with respect to the change rate
Step 8: for n in pop size do
if fit(n)>fit(lBest) then
lBest=popn;
end if
end for
Step 9: if fit(lBest)>fit(gBest) then
gBest=lBest;
end if
Step 10: Current population evaluate and return
Working of the ANN architecture
```

Because of ANNs' early prevails, there was a consensus that humans are imminently speaking with truly intelligent machines. A network of artificial neurons that can calculate any logical statement can be constructed. One of the most fundamental ANN architectures is the perceptron. It is based on a Threshold Logic Unit (TLU), a somewhat different type of artificial neuron. Rather than being binary on/off, values are accounted in terms of input as well as output counts, and all input connections are assigned to it. Rather than using binary on/off values, the TLU first calculates the input linear functions. Each input connection is assigned its weight. The TLU has initially computed with its input linear functions shown in equation 1.

$$z = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b \tag{3}$$

Where,
X_1, x_2 = Input features
W_1, W2 = Weight of the input features
b = bias

For conventional linear binary classification, one TLU is sufficient. It calculates a linear function from its inputs and outputs the positive class if the result is greater than a threshold. All TLUs have arranged for a single layer, each connected to all inputs, which make up a perceptron. This kind of layer is referred to as a dense layer or a fully linked layer. The input layer is made up of the inputs. Furthermore, the TLU layer is referred to as the output layer since it generates the final outputs. This perceptron is a multilabel classifier since it can simultaneously classify cases into three distinct binary classes. It can be applied to multiclass classification as well. The fully connected layer is illustrated in equation 2.

$$h_{w,b} \quad (x) = \varphi(XW + b) \tag{4}$$

Where,
X = Features as input matrix
W = Connected weights of the weight matrix
'b = Bias vector per neutron
φ = Activation function

Algorithm for ensemble kernel classifier
Input: Filtered features of the dataset
Output: Ensemble model with improved accuracy
Step 1: Let the selected feature data be considered for the training dataset and initiate to wrapped ANN model.
Step 2: Eight different kernels of the ANN model were generated as wrappers in the training model with 70 epochs and, batch size of 32.
Step 3: Ensembling kernel classification of the ANN method using a voting classifier with soft type voting.
Step 4: Return with the best kernel through a voting classifier for improved accuracy

7. Result and Discussion

The analysis and experiment results for the purpose of detecting breast cancer with our suggested method. To evaluate the efficacy of our approach, we created the experimental setting and examined measures including sensitivity, accuracy, F1 score, and specificity. The F1 score offers an accurate assessment, whereas accuracy assesses the overall classification correctness results, sensitivity assesses its capability in recognizing positive cases, and specificity assesses the capacity for recognizing negative instances. This proposed model can predict and evaluate with other current techniques in the BC diagnosis field, which provides important insights into the effectiveness of the research strategy. A high-end computer with eight cores, 64 GB RAM, and a hard drive with 100 GB was used for our experiments. To take advantage of the feature selection model with Boruta and feature extraction using LEAP and ensembling-based ANN is involved in this research. The environment involves necessary dependencies and libraries specifically for the Keras classifier. Utilizing a compressed environment plays an essential role in conducting the experiments proficiently and facilitates the analysis, and validates the proposed approach in diagnosing BC. The proposed LEAP with the ensemble ANN method is compared with other methods, namely Boruta with the ensemble ANN method and Boruta with the LGBM classifier method. The confusion matrix of LEAP with the ensemble ANN method is shown in Figure 3.

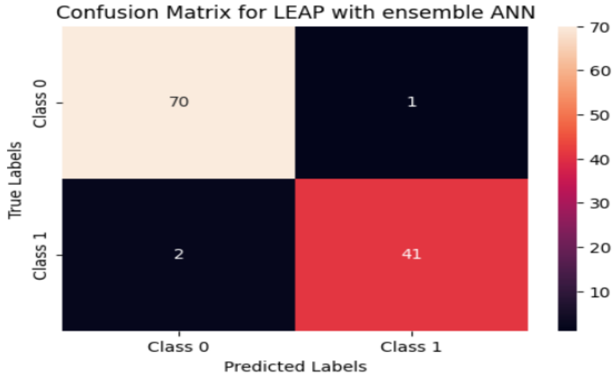

Fig. 3: Confusion Matrix for LEAP with Ensemble ANN.

Similarly, this dataset is used for Boruta with LGBM as well as Boruta with the ensemble ANN method. Table 2 illustrates the performance evaluated through the confusion matrix metrics like accuracy, precision, recall, and the positive instance and negative instance can be identified through sensitivity and specificity of the BC diagnosis status.

Table 2: Confusion Matrix Metric for Various Classification Methods

Classification Methods	Confusion Matrix metrics				
	Accuracy (%)	Precision (%)	Recall (%)	Sensitivity	Specificity
Boruta with LGBM Classifier	95.61	95.83	97.18	0.9718	0.9302
LEAP with Ensemble ANN	97.37	97.22	98.59	0.9859	0.9535
Boruta with Ensemble ANN	94.74	95.77	95.77	0.9577	0.9302

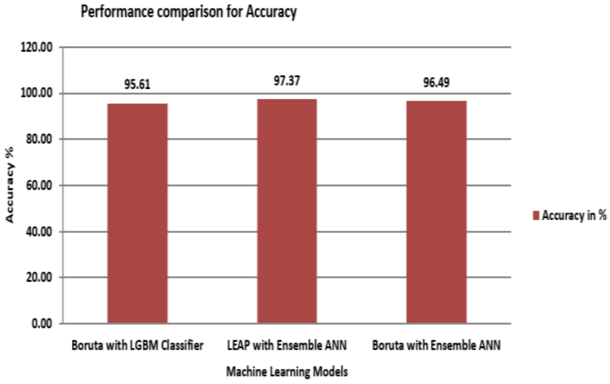

Fig. 4: Accuracy Performance for Various Feature Extraction-Based ML Methods.

Figure 4 illustrates that the performance of the proposed LEAP with the ensemble ANN method has a high accuracy of 97.37%. The performance of the proposed LEAP with the ensemble ANN method ensures a high percentage in accuracy, while compared to the other two, namely Boruta with the ensemble ANN method and Boruta with the LGBM method, are 95.61% and 94.74% respectively. Therefore, the proposed LEAP with the ensemble ANN method has better prediction in detecting BC.

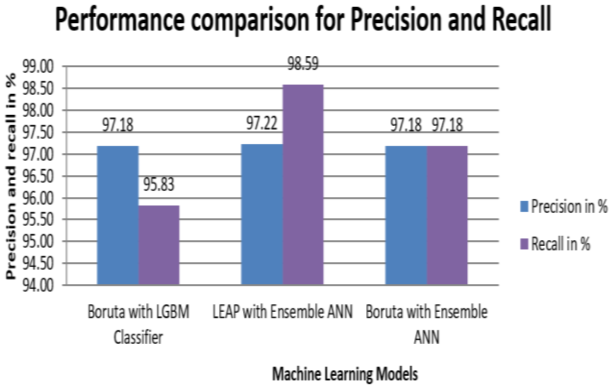

Fig. 5: Precision and Recall Performance for Various Feature Extraction-Based ML Methods.

Figure 5 illustrates the performance of the proposed LEAP with the ensemble ANN method has a high recall (true positive rate) and precision (False positive rate) are 98.59% and 97.22% respectively. The performance of the proposed LEAP with the ensemble ANN method ensures a high percentage in true positive rate, while compared to the other two, namely Boruta with the ensemble ANN method and Boruta with LGBM classifier, are 97.18% and 97.18% as well as 95.83% and 97.18% respectively. Therefore, the proposed LEAP with the ensemble ANN method has better prediction in detecting BC.

8. Limitations

The proposed LEAP-ANN model has some limitations, despite being more accurate. Evolutionary algorithms are affected by the scale of datasets, making them computationally complex and potentially causing a scalability problem. Also, there is a problem with the performance on highly imbalanced datasets, because overfitting to the majority class might take place. Next time, studies, it would be wise to address these areas and make the model more robust.

9. Conclusion

There are several research studies have been conducted in recent years to diagnose BC at an early stage. This is quite difficult for researchers in selecting the best approach for an efficient diagnostic model, even though a lot of effort has been put into this subject area. This study mainly focuses on improving the precision rate in detecting BC by modifying the feature-based activity and ensembling concept to the ML method. Based on the analysis, there are various improvements in accuracy for BC diagnosis through the proposed Bourta with a base ensemble ANN method. The evaluation of the proposed method is determined through accuracy metrics by comparing with Boruta with ANN method and the ANN method. The accuracy of the proposed method at 97.37% is higher than other ML methods. Hence, the high classification rates in LEAP with the ensemble ANN method have the potential for assisting healthcare industries to make more accurate BC diagnosis decisions. In future works, the image processing feature extraction is considered with hybrid DL methods for diagnosing BC earlier, as well as comparing with traditional DL methods. Future research will aim to expand the LEAP-ANN framework to explore multimodal medical data, including mammograms, genomic sequences, and blood-based biomarkers. In addition, further evaluation with future studies using the method in low-resource settings with highly imbalanced datasets will further offer feasibility and benefit to its practicality.

References

[1] Park, M. Y. *et al.* Function and application of flavonoids in the breast cancer. *Int. J. Mol. Sci.* 23, 7732 (2022). https://doi.org/10.3390/ijms23147732.

[2] Uddin, K. M. M., Biswas, N., Rikta, S. T. & Dey, S. K. Machine learning-based diagnosis of breast cancer utilizing feature optimization technique. Comput. Methods Progr. Biomed. Update 3, 100098 (2023). https://doi.org/10.1016/j.cmpbup.2023.100098.

[3] A. Sharma and P. K. Mishra, "Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis," International Journal of Information Technology (Singapore), vol. 14, no. 4, pp. 1949–1960, Jun. 2022, https://doi.org/10.1007/s41870-021-00671-5.

[4] S. S. Yadav and S. M. Jadhav, "Thermal infrared imaging based breast cancer diagnosis using machine learning techniques," Multimed Tools Appl, vol. 81, no. 10, pp. 13139–13157, Apr. 2022, https://doi.org/10.1007/s11042-020-09600-3.

[5] A. Alzu'bi, H. Najadat, W. Doulat, O. Al-Shari, and L. Zhou, "Predicting the recurrence of breast cancer using machine learning algorithms," Multimed Tools Appl, vol. 80, no. 9, pp. 13787–13800, Apr. 2021, https://doi.org/10.1007/s11042-020-10448-w.

[6] S. Rani, M. Kaur, and M. Kumar, "Recommender system: prediction/diagnosis of breast cancer using hybrid machine learning algorithm," Multimed Tools Appl, vol. 81, no. 7, pp. 9939–9948, Mar. 2022, https://doi.org/10.1007/s11042-022-12144-3.

[7] Afolayan, J.O.; Adebiyi, M.O.; Arowolo, M.O.; Chakraborty, C.; Adebiyi, A.A. Breast cancer detection using particle swarm optimization and decision tree machine learning technique. In Intelligent Healthcare; Chakraborty, C., Khosravi, M.R., Eds.; Springer Nature: Singapore, 2022; pp. 61–83. https://doi.org/10.1007/978-981-16-8150-9_4.

[8] Guida, F.; Kidman, R.; Ferlay, J.; Schüz, J.; Soerjomataram, I.; Kithaka, B.; Ginsburg, O.; Vega, R.B.M.; Galukande, M.; Parham, G.;et al. Global and regional estimates of orphans attributed to maternal cancer mortality in 2020. Nat. Med. 2022, 28, 2563–2572. https://doi.org/10.1038/s41591-022-02109-2.

[9] W.Wang, R. Jiang, N. Cui, Q. Li, F. Yuan, and Z. Xiao, ''Semi-supervised vision transformer with adaptive token sampling for breast cancer classification,'' Frontiers Pharmacol., vol. 13, Jul. 2022, Art. no. 929755. https://doi.org/10.3389/fphar.2022.929755.

[10] Z.-R. Jiang, L.-H. Yang, L.-Z. Jin, L.-M. Yi, P.-P. Bing, J. Zhou, and J.-S. Yang, ''Identification of novel cuproptosis-related lncRNA signatures to predict the prognosis and immune microenvironment of breast cancer patients,'' Frontiers Oncol., vol. 12, Sep. 2022, Art. no. 988680. https://doi.org/10.3389/fonc.2022.988680.

[11] M. Umer, M. Naveed, F. Alrowais, A. Ishaq, A. A. Hejaili, S. Alsubai, A. A. Eshmawi, A. Mohamed, and I. Ashraf, ''Breast cancer detection using convoluted features and ensemble machine learning algorithm,''Cancers, vol. 14, no. 23, p. 6015, Dec. 2022. https://doi.org/10.3390/cancers14236015.

[12] B. He, Y. Zhang, Z. Zhou, B. Wang, Y. Liang, J. Lang, H. Lin, P. Bing, L. Yu, D. Sun, H. Luo, J. Yang, and G. Tian, ''A neural network framework for predicting the tissue-of-origin of 15 common cancer types based on RNA-Seq data,'' Frontiers Bioeng. Biotechnol., vol. 8, p. 737, Aug. 2020. https://doi.org/10.3389/fbioe.2020.00737.

[13] Chaurasia, V. & Pal, S. Applications of machine learning techniques to predict diagnostic breast cancer. SN Comput. Sci. 1, 1–11(2020). https://doi.org/10.1007/s42979-020-00296-8.

[14] Kabiraj, S. et al. Breast cancer risk prediction using XGBoost and random forest algorithm. In 2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020. https://doi.org/10.1109/ICCCNT49239.2020.9225451.

[15] Jabbar, M. A. Breast cancer data classification using ensemble machine learning. Eng. Appl. Sci. Res. 48, 65–72 (2021).

[16] Shalini, M. & Radhika, S. Machine learning techniques for prediction from various breast cancer datasets. In 2020 6th International Conference on Bio Signals, Images, and Instrumentation, ICBSII 2020.

[17] Naji, M. A. et al. Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Comput. Sci. 191, 487–492 (2021). https://doi.org/10.1016/j.procs.2021.07.062.

[18] Benbrahim, H.; Hachimi, H.; Amine, A. Comparative study of machine learning algorithms using the breast cancer dataset. In Proceedings of the Advanced Intelligent Systems for Sustainable Development (AI2SD'2019), Marrakech, Morocco, 8–11 July 2019; Springer: Berlin/Heidelberg, Germany, 2020; pp. 83–91. https://doi.org/10.1007/978-3-030-36664-3_10.

[19] Ghiasi, M.M.; Zendehboudi, S. Application of decision tree-based ensemble learning in the classification of breast cancer. Comput. Biol. Med. 2021, 128, 104089. https://doi.org/10.1016/j.compbiomed.2020.104089.

[20] Alshayeji, M.H.; Ellethy, H.; Abed, S.; Gupta, R. Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. Biomed. Signal Process. Control 2022, 71, 103141. https://doi.org/10.1016/j.bspc.2021.103141.

[21] Kwang Ho Park, ErdenebilegBatbaatar, YongjunPiao, NiponTheera-Umpon, Keun Ho Ryu, Deep learning feature extraction approach for hematopoietic cancer subtype classification, Int. J. Environ. Res. Public Health 18 (4) (2021) 2197. https://doi.org/10.3390/ijerph18042197.

[22] Eun Young Park, Myungsun Yi, HyeSook Kim, Haejin Kim, A decision tree model for breast reconstruction of women with breast cancer: a mixed method approach, Int. J. Environ. Res. Public Health 18 (7) (2021) 3579. https://doi.org/10.3390/ijerph18073579.

[23] ZibaKhandezamin, MarjanNaderan, Mohammad JavadRashti, Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier, J. Biomed. Inform. 111 (2020) 103591. https://doi.org/10.1016/j.jbi.2020.103591.

[24] AbdurRasool, ChayutBunterngchit, Luo Tiejian, MdRuhul Islam, QiangQu, Qingshan Jiang, Improved machine learning-based predictive models for breast cancer diagnosis, Int. J. Environ. Res. Public Health 19 (6) (2022) 3211. https://doi.org/10.3390/ijerph19063211.

[25] Gary Pampar_a and Andries P Engelbrecht. Evolutionary and swarm-intelligence algorithms through monadic composition. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, pages 1382{1390, 2019. https://doi.org/10.1145/3319619.3326845.

[26] Mirabi, M., & Lahti, E. (2025). Graph signal processing-based interference suppression for multi-user massive MIMO systems. Electronics, Communications, and Computing Summit, 3(2), 56–61.

[27] Veerappan, S. (2025). Digital Twin Modeling for Predictive Maintenance in Large-Scale Power Transformers. National Journal of Electrical Machines & Power Conversion, 39-44.

[28] Abdullah, D. (2024). Design and implementation of secure VLSI architectures for cryptographic applications. Journal of Integrated VLSI, Embedded and Computing Technologies, 1(1), 21-25.

[29] Prasath, C. A. (2023). The role of mobility models in MANET routing protocols efficiency. National Journal of RF Engineering and Wireless Communication, 1(1), 39-48.

[30] Sadulla, S. (2025). IoT-enabled smart buildings: A sustainable approach for energy management. National Journal of Electrical Electronics and Automation Technologies, 1(1), 14–23.

[31] Rahim, R. (2025). Lightweight speaker identification framework using deep embeddings for real-time voice biometrics. National Journal of Speech and Audio Processing, 1(1), 15–21.

[32] Surendar, A. (2025). Lightweight CNN architecture for real-time image super-resolution in edge devices. National Journal of Signal and Image Processing, 1(1), 1–8.

[33] Abbas, M. A., Hatem, T. M., Tolba, M. A., & Atia, M. (2023). Physical Design of Speed Improvised Factor in FPGA Applications. Journal of VLSI Circuits and Systems, 5(1), 61–66. https://doi.org/10.31838/jvcs/05.01.09.

[34] Marwedel, R., Jacobson, U., & Dobrigkeit, K. (2025). Embedded systems for real-time traffic management: Design, implementation, and challenges. SCCTS Journal of Embedded Systems Design and Applications, 2(1), 43–56.