Breast Cancer Diagnosis Analysis And Differentiation Of Machine Learning Algorithms

Dr. Vaka Murali Mohan(Principal and Professor),
Makloor Harshika, Paka Sai Rishitha, Bhukya Shiva Nayak, Akumarthi Dilip,
Department of CSE,
MALLA REDDY INSTITUTE OF TECHNOLOGY AND SCIENCE, Telangana ,Hyderabad.

ABSTRACT

Several empirical studies on breast cancer have used machine learning and soft computing approaches. Many writers boast that their algorithms are the best of the best whether it comes to efficiency, user-friendliness, or precision. A system for identifying malignant or normal breast tissue has been developed in this study using genetic programming and machine learning techniques. This study set out to determine the optimal training method for the algorithm to identify them. In this case, we optimized our machine learning classifiers' feature sets and parameters using genetic programming. The proposed approach was assessed using the roc curves, which stand for sensitivity, specificity, precision, and accuracy. Finding the best model automatically is possible using genetic programming, feature preprocessing methods, and classifier algorithms, as shown in this study.

1. Introduction

Worldwide, breast cancer ranks first among female cancers and is one of the leading causes of cancer-related deaths among women [1]. With the abundance of imaging options for breast cancer diagnosis, diagnostic accuracy has been greatly improved [3, 4]. This has led to early identification and treatment of the illness, which in turn has decreased mortality [2]. The breast cancer detection and classification pipeline has been developed over the past few decades utilizing data mining and machine learning. It consists of three main steps: preprocessing, feature extraction, and classification [5-7]. Multiple writers have shown how preprocessing mammography films may improve analysis and interpretation by making the peripheral areas and intensity distribution more visible [8, 9]. When diagnosing breast cancer, feature extraction is crucial for identifying benign or malignant tumors. Images' smoothness, coarseness, depth, and regularity may then be extracted using segmentation [10]. A

variety of techniques for transform-based texture analysis use the spatial frequency aspects of the variations in pixel intensity to generate a new imagery format.

Popular techniques include the wavelet transform [11], the fast Fourier transform [12], the Gabor transform [13], and the singular value decomposition [14]. Reducing the dimensionality of the feature representation is possible using principal component analysis (PCA) [15]. Automated breast cancer screening using machine learning algorithms has been the subject of many efforts. One method that uses wavelet feature extraction and fuzzy logic classification was created by Malek et al. [16]. While Sun et al. [17] investigated the matter by comparing features selection methods, Zheng et al. [18] used a K-means algorithm with a support vector machine (SVM) to diagnose breast cancer. Many earlier articles [7] have relied on clustering and classification. On page 19, Alickovi'c and Subasi introduced the rotating forest classifier and the evolutionary approach to feature extraction.
Finally, Bannaie has just finished [20] research that relies on dynamic contrast enhanced magnetic resonance imaging (DCEMRI) to extract valuable information. The preprocessing step is where the writers of this work really shine. Despite many attempts, the breast cancer diagnostic techniques described in published works are only semiautomatic. Hyperparameters are parameters that are not easily calculable from the data; this is a term used by Kuhn and Johnson [21]. Typically, in order to get the desired algorithm performance, it is necessary to fine-tune certain model parameters. As an example, the proper values for the learning rate, C, and sigma parameters used by support vector machines and neural networks, respectively, cannot be determined by any mathematical process. In the United States, nobody has yet figured out how to choose the best tuning settings for any particular model.

Machine learning has finally been commercialized as a service due to its increasing popularity. Regrettably, machine learning is still considered a niche area that sometimes need for much expertise and training. There are several processes involved in constructing a successful machine learning model, including preprocessing, feature selection, and classication. The data flow shown in Figure 1 is the result of machine learning models and pipelines. Various choices are presented at each stage of processing. At both the preprocessing and classication stages of the proposed pipeline, techniques and parameters are autonomously selected. Any competent machine learner can tell you which approach is going to be the most effective in any particular situation. Professionals in the field of machine learning may not need nearly as much time to refine the models they propose and achieve their goals. This work aims to automate the creation of machine learning models by utilizing genetic programming [22] to optimize the greatest possible combination of techniques. In Figure 2, we can see the GP algorithm in action. At every iteration, the accuracy of the pipeline's classification was evaluated. To determine the optimal pipeline, the GP algorithm underwent evolution employing the operators of selection, mutation, and crossover.

2. Materials and Methods

2.1. Dataset Used for Research.

The dataset used in this work, the Wisconsin Breast Cancer dataset, was obtained from the UCI Machine Learning Repository. Bennett [23] to detect tumors that could or might not be cancerous.

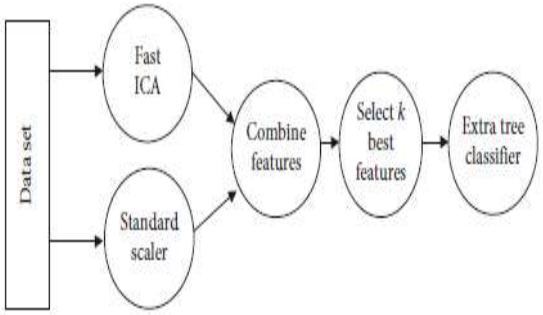


Figure 1: Example of pipeline.

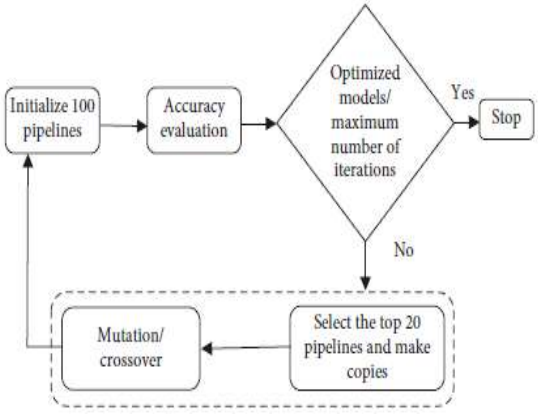


Figure 2: Flowchart of GP.

The nucleus features in this picture were extracted from digital images of a breast tumor needle aspirate [24]. There were 569 individuals analyzed using WDBC data at Wisconsin hospitals; 212 of those cases were cancer and 357 were noncancerous diseases. Each table contains the data points that were obtained from separate FNA testing. The last two features of this dataset are the indenter number and the diagnostic status. The other 30 features are the real

ones; they include the average, the margin of error, and the lower ten features of every cell's nucleus. Ten real values were considered: symmetry, area, smoothness, compactness, boundary, area, texture, concave points, and concavity.

2.2. Related Work.

An essential part of machine learning is feature selection, which involves narrowing down a collection of potential attributes to a meaningful subset. The selection of features is an essential step in developing a predictive model. A few benefits of using feature selection techniques are as follows: (a) training the machine learning algorithm becomes faster and more efficient; (b) models are simplified and made more clear; (c) models are more accurate if the correct subset is chosen; and (d) over fitting is reduced. Selecting the best subset is infamously difficult [25] because to the characteristics' potential complex interdependencies. The scientific literature has proposed a variety of breast cancer detection approaches [7, 17–20]. When classifying feature selection strategies, they are often put into the following three categories: filter, wrapper, and embedding methods [26].Due of its dependence on broad properties, the e lter approach is frequently used as a preprocessing step. There is no dependence on the mode of instruction in the mechanism for subset selection. Using ML techniques, the wrapper approach selects the most desirable attributes. The learning process directs the selection of features, as seen in Figure 3.Surrounding a feature selection and elimination technique with a "wrapper" approach is standard practice.

Embedded methods include the best features of both wrapper and filter techniques. To put them into action, algorithms are used, and many of these algorithms have their own methods for selecting features. They are in charge of the variable selection phase of learning and are specific to the learning machines in use. A data stream is shown graphically in Figure 4. Wrapper techniques were used in the experiments presented in this article.

2.3. The Proposed Method.

By joining several modules that are executed in a certain order into one unified unit, a "pipeline" is established to provide an automated machine learning process. It simplifies the process significantly while still providing access to advanced machine learning techniques via abstraction. The most prevalent kind of this is the Extract, Transform, and Load (ETL) procedure. Some examples of hyperparameters that could impact ML technique performance include neural network depth, hidden layer count, learning rate, batch size, and regularization level. The objective of this endeavor is to identify the optimal set of data transformations and machine learning algorithms to do the classification. It may be difficult to determine the optimal combination of data and machine learning algorithms. The use of genetic programming (GP) [22] to adjust the model's input data and output control parameters has been proposed as a solution to hyperparameter tuning. Using this famous evolutionary technique is necessary to find the best combination that produces the highest evaluation results. E GP randomly generates a certain number of pipes (the population). Here, we evaluate all pipelines in the population using a classification score, "fitness," that is derived from supervised models in the sickest-learn package. All of the classifiers in this research use a random selection of hyperparameters to assess their performance, with the exception of linear discriminant analysis. This study assessed a wide variety of useful approaches that may be used in the processing and analysis of breast cancer datasets in the future.

This procedure begins with a step known as "preprocessing." In this investigation, the Standard Scaler module was used to scale the features of the raw breast cancer data. Machine learning estimators often presuppose data standardization. The formula (xi-mean(x))/ stdev(x), where stdev(x) is the standard deviation, is used to transform the characteristics into a distribution compatible with the normal distribution. The *e Robust Scalar requires the

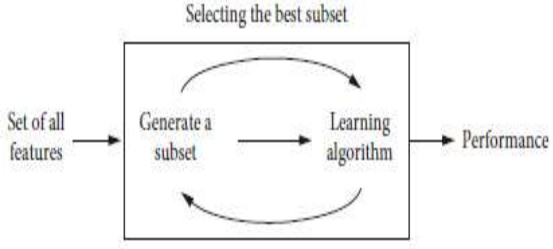


Figure 3: Wrapper methods.

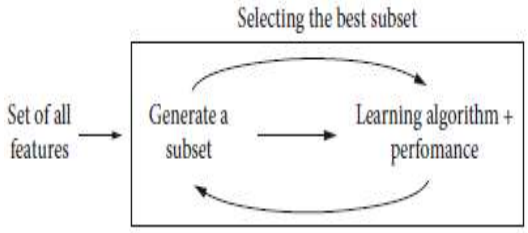


Figure 4: Embedded methods.

Eigenschaften abhängig von der Quartile-Differenz (Q3(x)-Q1(x)) zu den Quartile-Differenz (Q1(x)-Q2(x)). Each change is detailed in the scikit-learn [27] machine learning library. Selecting features is the subject of the second stage, which is detailed in Section 2.3.2. As a kind of preprocessing, feature selection is usually used before to doing any real learning. We used principal component analysis (PCA) with randomized structural vector decomposition (SVD) [28] to save the most relevant features while reducing the dataset size since no algorithm can provide reliable predictions without informative and discriminative features. The feature-selection module was built using the scikit-learn package in Python. In order to extract valuable features, all selection techniques employed too many criteria. We employed univariate feature selection, recursive feature elimination, and the elimination of characteristics with low variance as part of our work.

Third, implement a machine learning algorithm. An ensemble of ML algorithms often outperforms a single model when it comes to prediction accuracy. A model for diagnosing breast cancer may have used the winning answer from the machine learning competition, according to one reading of *is. The data set in this paper was classified using the following methods: support vector machine (SVM) [29], K-nearest neighbor (KNN) [30], decision tree (DT) [31], gradient boosting classifier (GB) [32], random forest (RF) [33], logistic regression (LR) [34], AdaBoost classifier (AB) [35], Gaussian Naive Bayes (GNB) [36], and linear discriminant analysis (LDA) [37]. Securing optimal parameters is the fourth step. Genetic programming (GP) is one example of an evolutionary algorithm (EA) that has been developed by generalizing the genetic algorithm. GP is a way to figure out what choices are best and then choose the best one. The fundamental processes of biological evolution—mutation, crossover, and selection—are the ones that GP uses to discover a solution.

Because it may be used to model systems when the necessary model structure and its key properties are unknown in advance, GP is very adaptable. By optimizing the tree-based pipelines for the classification problem, this study made use of GP, which enabled the system to seek for models from a number of different model designs. To begin, GP uses primitives such as features selection decomposition to build a certain number of pipelines. To rephrase, an iterative set of operators is used to construct machine learning pipelines, which are then evaluated in order to enhance classification accuracy. One alternative architecture of the machine learning pipeline is presented in Figure 1. The most effective parts of a machine learning pipeline are reused from generation to generation. For GP, every pipeline is a person. The following three businesses make up *e GP:

Mutation operator: changing hyperparameters or including or deleting a basic preprocessing step (e.g., Random Forest size, Standard Scalar). In its conservative model, the crossover operator presumes that 5% of individuals will mate via a 1-point random crossover. The selection operator's main objective is to identify the top twenty performers and then reproduce them. People in a population may exchange genetic data with one another thanks to the mutation or crossover operator. Figure 2 shows the different stages of GP.

3. Results

This study's models were tested using the Wisconsin Breast Cancer dataset to ensure their correctness. Breiman et al. [38] states that test partitions do not provide dependable estimates of the classification error scheme on a short dataset, while using one training and one test. For this investigation, we (*us) decided to employ a random subsample technique to lower the chance of estimating mistakes. To prevent overfitting, cross-validation is an excellent strategy to use. This led to the use of a 10-fold sample size for cross-validation on the breast cancer dataset. The inquiry included the establishment of three training trials. The first case highlighted the significance of the feature selection process. In the second experiment, we mainly looked at the classification model. The main objective of the third experiment was to combine all the prior ones into one self-regulating process. This points to the goal being the development of machine learning algorithms and optimization of programs. The first experiment

employed the open-source machine learning software WEKA to extract features based on EA. Particle swarm optimization (PSO) [39], genetic algorithm (GA) [40], evolutionary programming (EP) [22], and best first (BF) [27] were used in this process. Table 1 shows the selected attributes used in the prior search methods. The table data suggests that the feature sets used by the tested algorithms are similar to one another. Table 1: Feature-Selection Algorithm Comparison shows that 60% of the algorithms were successful, according on the results of the applied filter features.

Search algorithm	Number of selected attributes	Numbers
PSO	1, 9, 10, 16, 21, 23, 24, 25, 26, 27, 30, 31	12
Evolutionary search	1, 3, 9, 10, 11, 15, 23, 24, 25, 26, 27, 29, 30	13
Genetic algorithm	1, 7, 9, 10, 16, 21, 23, 24, 25, 26, 29, 30	12
Best first	1, 4, 9, 10, 16, 21, 23, 25, 26, 27, 29, 30	12

Eighty percent of the EA's characteristics were common to all of the methods. By using filtering procedures, which zero down on a certain characteristic, the relevance of a particular feature was ascertained. Removing superfluous data using a relevance score. Also, each search method has to have its own tuning parameter. We think that integrating various feature extraction approaches enhances the prediction accuracy of the applied classifiers, as there is no one "best" features selection methodology (as mentioned by Yong et al. [25]). We verified that the chosen model's performance was improved by traits retrieved using hybrid methods. In the second research, the effectiveness of many popular supervised learning algorithms in problem classification was tested. To determine how good a machine learning algorithm's model is, many measures are used. Results were evaluated using precision-recall, confusion matrix, area under the curve, and accuracy. It is possible to deduce the amount of misclassified samples from the accuracy (ACC) metric, which evaluates the precision of the classifier's prediction. More especially, it denotes

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{FP} + \mathrm{FN} + \mathrm{TP} + \mathrm{TN}}, \quad (1)$$

The predicted rates of true positives, false positives, true negatives, and false negatives are denoted as TP, FP, TN, and FN, respectively.

The remaining metrics that are based on confusion matrices are defined below:

$$\begin{aligned} \mathrm{recall} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\ \mathrm{precision} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\ \mathrm{F1} &= 2 \times \frac{(\mathrm{precision} \times \mathrm{recall})}{(\mathrm{precision} + \mathrm{recall})}. \end{aligned} \quad (2)$$

Along with the previously described metrics, receiver operator characteristic (ROC) diagrams [41] were used to illustrate the relationship between sensitivity (recall) and specificity. The ROC curves showcase the effectiveness of the learning approach in this experiment, disregarding class distribution and error overheads. The models used include LR, LDA, K-neighbors', DT, GNB, RF, extra trees, AB, and GB. As mentioned before, the x and y axes of

the ROC space represent true positives and false positives, respectively. Statistically speaking, the ROC curve averages performance over several cutoff points. *Classification models are considered to be worse than random guessing if their scores fall below the diagonal of the ROC curve. The ideal true positive and false positive rates for a classifier would be 1 and 0, respectively. Consequently, it would be positioned in the top left corner of the graph. An indicator of a classification model's efficacy is the area under the ROC curve (AUC). Results demonstrate that applied models outperform *us in terms of prediction accuracy. Figures 5–13 provide a comparison of the outcomes produced by the nine computational models. In our study, we found that GNB had a better mean ROC, coming in at 77%. Breiman et al. [38] argued that accurate estimators cannot be achieved by classification error approaches that rely on a single train and test partition. We opted for a random subsample strategy to lessen the possibility of bias in the estimation procedure. A ten-fold sample size was used for cross-validation on the breast cancer dataset. The ROC results were made more understandable by using five-fold cross-validation. In the prior trial, we discovered that the LR, LDA, and GNB algorithms gave the greatest match when we left all of the machine learning classifiers' input parameters at their default values.

4. Discussion

We found experimental evidence supporting our hypothesis that merging features selection methodologies improves accuracy performance. The genetic programming approach was proposed as a result. To construct the specified amount of conduits. So, to automate choosing the best pipeline, several machine learning approaches were used. Consequently, the suggested approach is seen as a possible way to choose the appropriate algorithm and optimize the model's hyperparameters for maximum yield.Model selection relies on hyperparameters, which control the chosen model's complexity but aren't directly learned by classifiers.Do your homework since there are a lot of hyperparameters options to choose from. Machine learning approaches use arbitrary model parameters. As a result, a lot of researchers end up making modifications by hand. To sum up, the control parameters may affect both the learning algorithm's prediction performance and the model's complexity. Examining the hyperparameters problem was the intended objective of this endeavor.There were three separate parts to the experiment. In the first test, two popular evolutionary algorithms—PSO and GA—were pitted against one another in terms of feature selection. Based on the results of this study, it was found that 80% of the specified traits were common. Typical of approaches based on evolution, there were many modifiable parameters. Since it can be difficult to pretend to be well-versed in every single known methodology, authors run the danger of introducing bias into method selection. To offset this bias, we used a precision control setting.

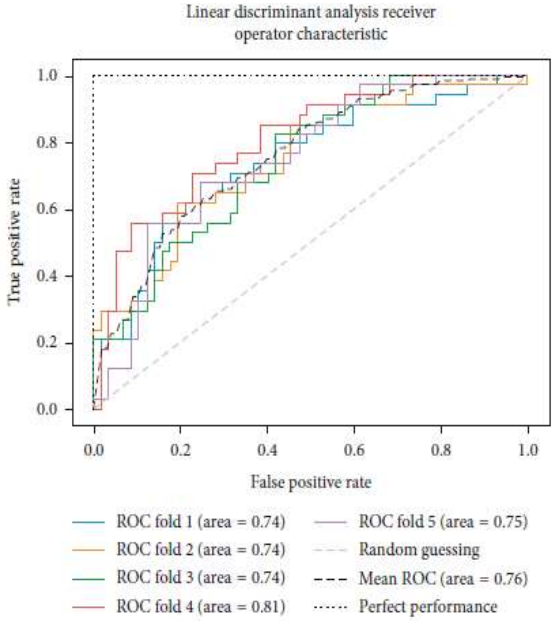


Figure 5: ROC curve for LDA.

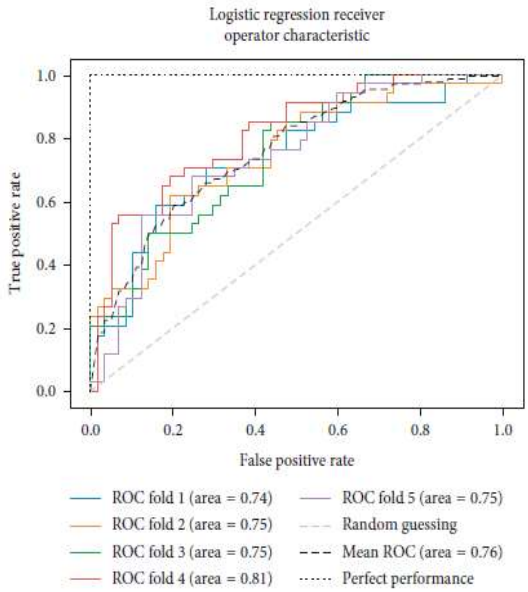


Figure 6: ROC curve for LR.

The resultant configuration space of an algorithm with n parameters is an n-dimensional hypercube. We contemplated introducing a simple method for determining valuable traits here. As a result, principal component analysis (PCA) was used to reduce dimensionality in the breast cancer dataset. While standardizing feature selection was warranted, the previous techniques required a handful of parameters. The authors in [42] came to a similar conclusion, stating that feature extraction and selection help keep the data interpretable and increase its discriminative power. The courses' location in Figure 14 makes this very clear.

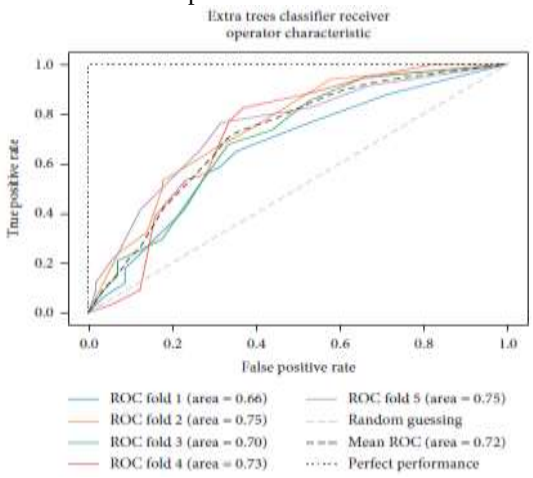


Figure 7: ROC curve for ET.

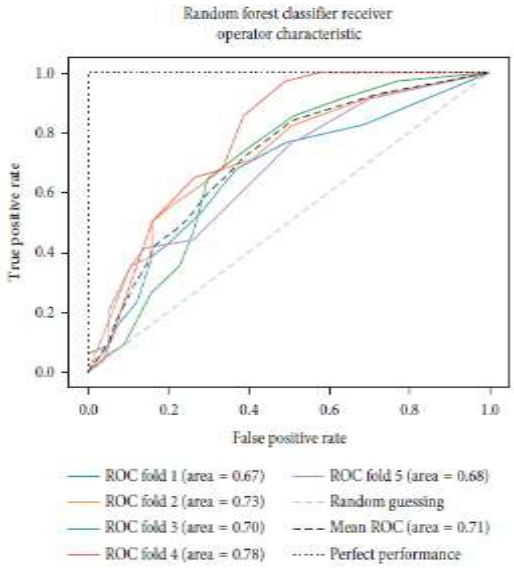


Figure 8: ROC curve for RF.

Principal component analysis and standardized feature selection provide data that can be linearly separated. Second, how can the researcher choose the most effective machine learning algorithm? Many things are considered while choosing a machine learning algorithm, but two of the most common are complexity and accuracy. On the other hand, accuracy is the only concern of many consumers. Consequently, there are writers who claim their algorithms outperform the ones that have been published before. It is common practice to need extensive training and careful selection of hyperparameters when attempting to maximize the performance of a machine learning system.

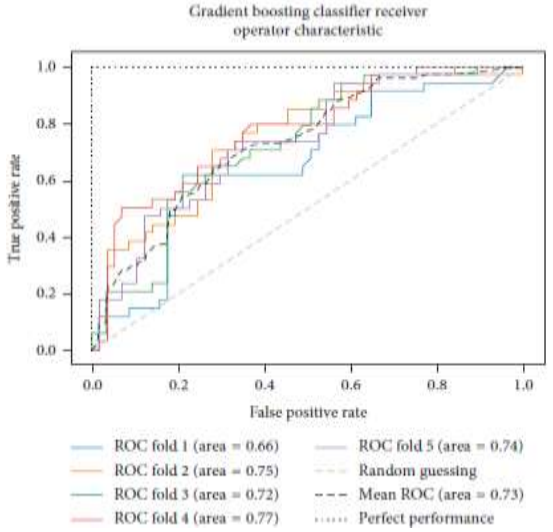


Figure 9: ROC curve for GB.

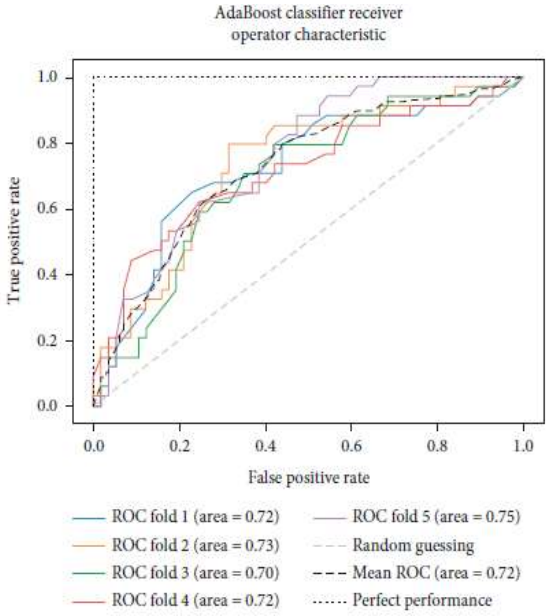


Figure 10: ROC curve for AB.

No algorithm may be said to have universal applicability, according to the "No Free Lunch" theorem (Wolpert and Macready, 2018). Therefore, several approaches must be evaluated with respect to a given problem before a winner can be declared. Our evaluation of machine learning algorithms included DT, RF, AB, GNB, LDA, quadratic discriminant analysis, LR, and features classifier, among others. For this experiment, accuracy and log-loss were useful metrics to keep in mind. Figure 16 shows the log-loss, while Figure 15 shows the comparison of useful machine learning's accuracy. According to Table 2, AdaBoost has the highest success rate of any classifier at 98.24%.

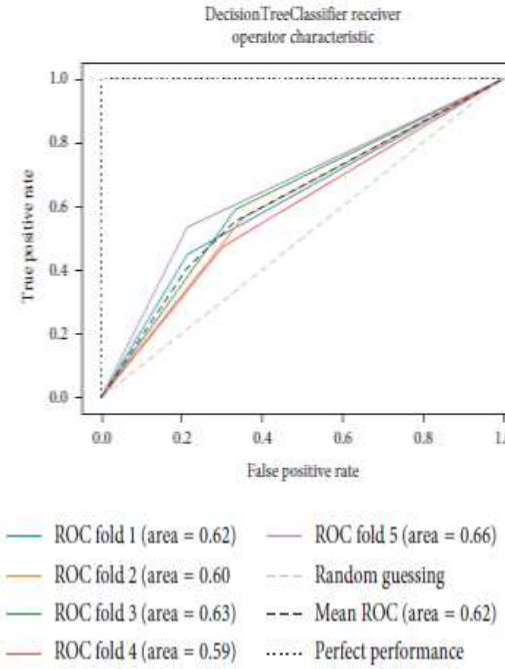


Figure 11: ROC curve for DT.

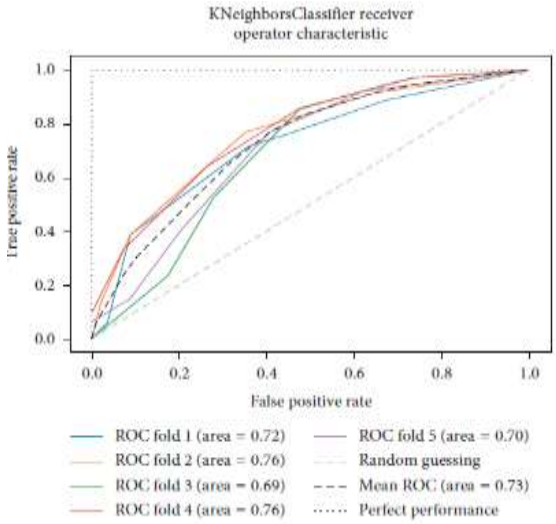


Figure 12: ROC curve for KNN.

But this is incorrect since the AdaBoosting classifier has a log-loss measure of 0.39, as shown in Table 3. The log-loss is widely acknowledged for providing a more accurate representation of the model. We found that the GB classifier, the RF classifier, and the extra tree classifier were the three best performers according to the data. Figure 17 shows the estimate's variability as a function of the standard deviation of the average accuracies. You can observe a big difference between the training set and the cross-validation set in the picture, which shows the accuracy curves. Consequently, the accuracy curve confirms the previous results. Setting the control parameters of a practical model has a significant impact on its performance, as mentioned before. This is why we made an effort to automate the whole process, beginning with feature selection and ending with classification. As seen in Section 2, a great deal of flexibility in configuration was achieved by use of GP by means of the pre-existing modules. The control parameters were adjusted with every randomly produced structural component. Using the Standard Scaler to standardize the input data, the RFE to reduce the quantity of features, and the logic regression for classification, one may develop a random-building model. The key control parameters of the chosen algorithms were fine-tuned using GP approaches in this endeavor. Once the GP settings (population size, generation numbers, etc.) are initialized, no human intervention is necessary thereafter. With so many options for features,

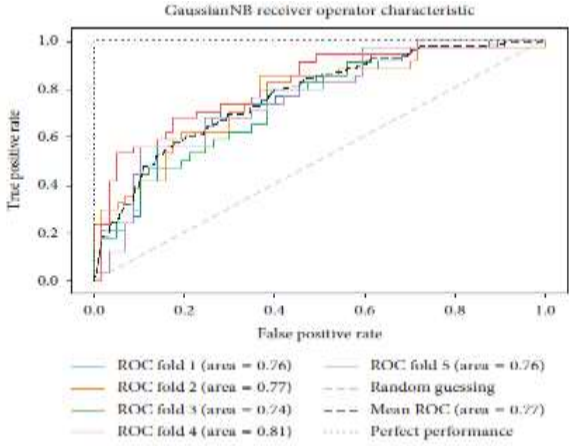


Figure 13: ROC curve for GNB.

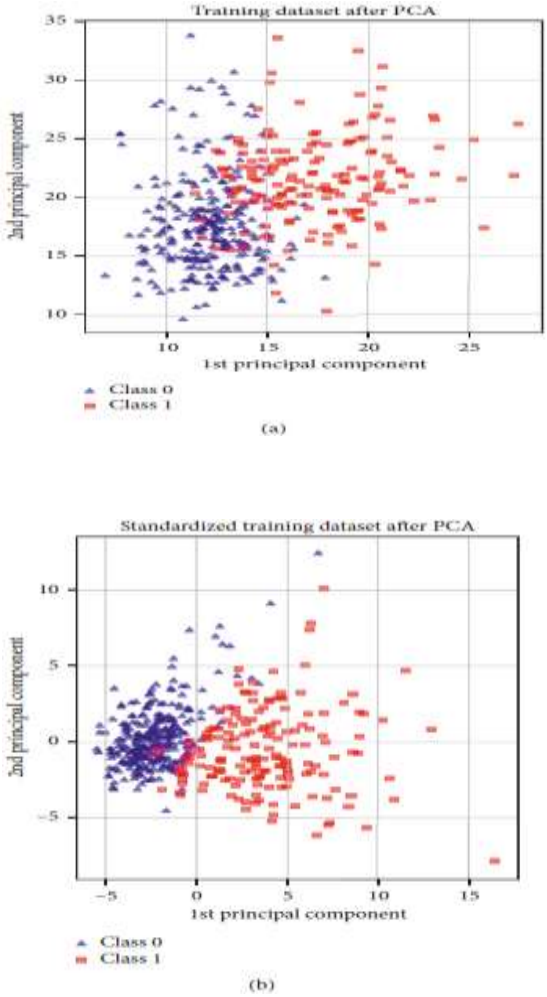


Figure 14: Combining feature extraction.

Selecting a limited set of algorithms for use with each methodology and classification approach might help keep assessment time down. We have already discussed why we chose this particular approach. Following a comparison with the outcomes of the randomly selected methods, we preserved the model with the most accuracy. The MaxAbsSclaer operator was used for preprocessing, the polynomial features operator for feature selection, and the gradient boosting classier as the model for supervised classification. After systematically integrating the approaches employed, an ensemble of methods was constructed. An accuracy of 98.24% was achieved throughout validation.

Vol 13 Issue 04,Nov 2023
ISSN NO: 0364-4308

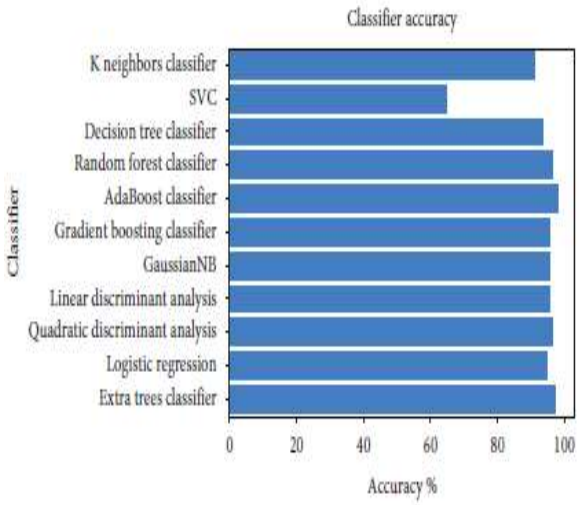
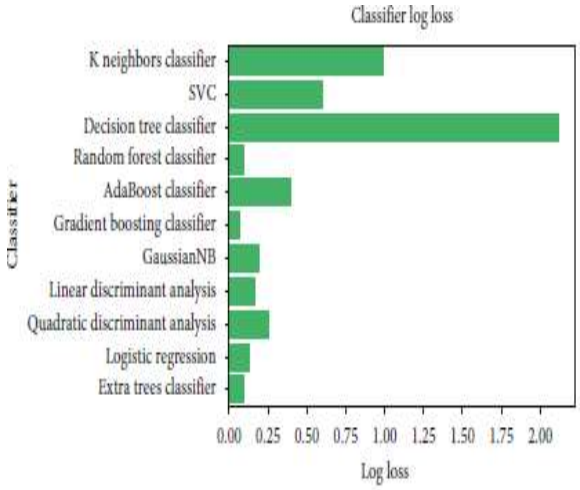


Figure 15: Comparison of classifier accuracy.



Log loss classifier comparison, shown in Figure 16.

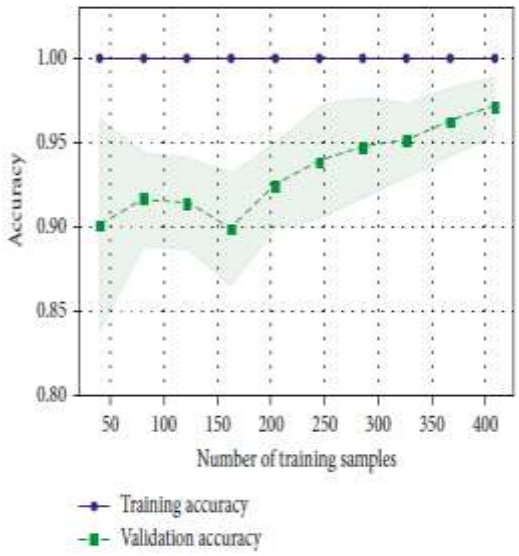


Figure 17: Validation accuracy.

Table 2: F1Measurements for breast cancer results.

	GB	DT	RF	GBN	SVM	KNN	AB	LDA	QDA	LR	ET
Benign (%)	96.69	95.36	97.37	96.69	70.72	95.42	98.67	96.73	0.97.26	96.10	98.01
Malignant (%)	93.51	90.91	94.74	95.51	0	86.84	97.44	93.33	95.12	91.89	96.10
Average (%)	95.57	93.80	96.45	95.57	51.10	91.11	98.23	95.33	96.51	94.63	97.34

Table 3: Logloss measure result for breast cancer results.

	GB	DT	RF	GBN	SVM	KNN	AB	LDA	QDA	LR	ET
Log-loss (%)	0.06	2.12	0.09	0.29	0.59	0.992	0.39	0.16	0.25	0.15	0.09

5. Conclusions

This study takes on the problem of autonomous breast cancer detection by using a machine learning approach. Numerous studies made use of the breast cancer dataset. With the right parameters, the three most popular evolutionary algorithms provide the same outcomes in the first trial. The idea that combining several feature selection methods at once improves accuracy was tested in the second experiment. Using machine learning, we automated the creation of a supervised classifier in the final experiment. The hyperparameters problem is tough for ML algorithms to solve, therefore we tried the GP method. The recommended approach zeroed focused on the optimal setup for the current issue. The code for each and every experiment was written in Python. After evaluating an ensemble of techniques from a complete machine learning strategy, we found that the suggested approach produced considerable results. However, we encountered much greater time consumption than what was originally expected. Ultimately, it seems that the proposed model is well-suited for discovering the control parameters of machine learning algorithms and for automated breast cancer screening.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," International Journal of Cancer, vol. 136, no. 5, pp. 359–389, 2014.
[2] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, "Variability in radiologists' interpretations of mammograms," New England Journal of Medicine, vol. 331, no. 22, pp. 1493–1499, 1994.
[3] N. Mao, P. Yin, Q. Wang et al., "Added value of radiomics on mammography for breast cancer diagnosis: a feasibility study," Journal of the American College of Radiology, vol. 16, no. 4, pp. 485–491, 2019.
[4] H. Wang, J. Feng, Q. Bu et al., "Breast mass detection in digital mammogram based on gestalt psychology," Journal of Healthcare Engineering, vol. 2018, Article ID 4015613, 13 pages, 2018.
[5] A. J. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer Informatics, vol. 2, pp. 59–77, 2006.
[6] G. Valvano, G. Santini, N. Martini et al., "Convolutional neural networks for the segmentation of microcalcification in mammography imaging," Journal of Healthcare Engineering, vol. 2019, Article ID 9360941, 9 pages, 2019.
[7] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Systems with Applications, vol. 36, no. 2, pp. 3240–3247, 2009.
[8] D. Narain Ponraj, M. Evangelin Jenifer, P. Poongodi, and J. Samuel Manoharan, "A survey of the preprocessing techniques of mammogram for the detection of breast cancer," Journal of Emerging Trends in Computing and Information Sciences, vol. 2, no. 12, pp. 656–664, 2011.
*[9] A. P. Charate and S. B. Jamge, "*e preprocessing methods of mammogram images for breast cancer detection," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 1, pp. 261–264, 2017.*
[10] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," IEEE Transactions on Image Processing, vol. 9, no. 4, pp. 561–576, 2000.
[11] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on Image Processing, vol. 1, no. 2, pp. 205–220, 1992.
[12] G. Carter, C. Knapp, and A. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," IEEE Transactions on Audio and Electroacoustics, vol. 21, no. 4, pp. 337–344, 1973.