# COLLATION OF MACHINE LEARNING METHODS FOR BREAST CANCER DIAGNOSIS

***Valla Shyam Kumar[1]., Jakkam Sanjana[2].,Komma Venkatasai Vaishnavi[3].,Kothi Shirisha[4].,Kunamaneni Kavya Sri[5]***

*1 Assistant Professor, Department Of CSE., Malla Reddy College Of Engineering For Women., Maisammaguda.,*

*Medchal.,Ts, India (✉vsmca15@gmail.com)*

*2, 3, 4, 5 B.Tech CSE, (19RG1A0578, 19RG1A0589, 19RG1A0592, 19RG1A0595),*

*Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India*

## ABSTRACT

***Using machine learning and soft computing methods, there has been various empirical researches addressing breast cancer. Various authors assert that their algorithms are the most efficient, user-friendly, or precise available. This research uses genetic programming and machine learning methods to build a system for determining if breast tissue is benign or cancerous. This research aimed to find the best way to train the algorithm to detect them. Here, we used genetic programming to determine the optimal feature set and parameters for our machine learning classifiers. The sensitivity, specificity, precision, accuracy, and roc curves were used to evaluate the suggested method's efficacy. This research demonstrates that by using genetic programming in conjunction with feature preprocessing techniques and classifier algorithms, the optimal model may be found automatically.***

## 1. Introduction

One of the worst killers of women across the globe, breast cancer is also the most common form of cancer in women everywhere [1]. There are a plethora of imaging methods available for many assisted breast cancer diagnosis methods have been used to improve diagnostic accuracy [3, 4], which has contributed to earlier detection and treatment of the disease as well as a decrease in mortality [2]. Preprocessing, feature extraction, and classification are the three key phases of the breast cancer detection and classification pipeline that have been established using data mining and machine learning during the last several decades [5-7]. Preprocessing mammography films, as described by various authors [8, 9], can increase the visibility of peripheral regions and intensity distribution, which in turn facilitates interpretation and analysis. Breast cancer diagnosis relies heavily on feature extraction to determine whether or not a tumor is benign or malignant. Segmentation is then used to extract attributes of images such as smoothness, coarseness, depth, and regularity [10]. The spatial frequency features of the pixel intensity fluctuations are used by a number of transform-based texture analysis algorithms to produce a new picture format.

Wavelet transform [11], rapid Fourier transform [12], Gabor transform [13], and singular value decomposition [14] are among methods that are often used. Principal component analysis (PCA) [15] may be used to minimize the number of dimensions in the feature representation. There have been several attempts to use machine learning algorithms to automate breast cancer detection. Examples include a technique developed by Malek et al. [16] that employs wavelet feature extraction and fuzzy logic classification. Zheng et al. [18] used a K-means algorithm with a support vector machine (SVM) for

breast cancer detection; whereas Sun et al. [17] examined the issue by contrasting features selection approaches. Clustering and classification have been the basis for a number of previous publications [7]. The evolutionary method for feature extraction and the rotation forest classifier were presented by Alickovi'c and Subasi [19].

Last but not least, Bannaie recently completed [20] research using the dynamic contrast enhanced magnetic resonance imaging (DCEMRI) method to glean useful data. The authors of this work make their main contribution during the preprocessing phase. Breast cancer diagnostic procedures published in the literature are semiautomatic, despite significant efforts. Parameters that cannot be readily calculated from the data are referred to as hyperparameters, a term used by Kuhn and Johnson [21]. Tuning certain model parameters is usually required to get the expected performance out of an algorithm. For example, there is no mathematical procedure to derive the correct value for the learning rate used in neural network training or the parameter C and sigma parameter used in support vector machines. The issue of how to choose the optimal tuning parameters for a given model remains unresolved in the U.S. The rising popularity of machine learning has led to its eventual commercialization as a service. Unfortunately, machine learning is still a specialized field that frequently necessitates in-depth training and knowledge. Preprocessing, feature selection, and classication are only a few of the steps involved in designing an effective machine learning model. Machine learning models and pipelines are shown as a.ow of changes on data in Figure 1. Different options are available at each processing step. The suggested pipeline uses autonomously determined procedures and parameters for both the preprocessing step and the classication stage. An experienced machine learner knows which method will work best for a given scenario. Experts in machine learning may not need as much time to fine-tune their suggested models and attain the desired results. The goal of this study is to maximize the best possible mix of strategies by applying genetic programming [22] to automate the building of machine learning models. The steps of the GP algorithm are shown in Figure 2. The precision of the pipeline's classication was measured at each iteration. The GP algorithm was evolved using the selection, mutation, and crossover operators to identify the most effective pipeline.

# 2. Materials and Methods

## 2.1. Dataset Used for Research.

The Wisconsin Breast Cancer dataset used in this study was downloaded from the UCI Machine Learning Repository. Bennett [23] to identify cancers that may or may not be malignant.

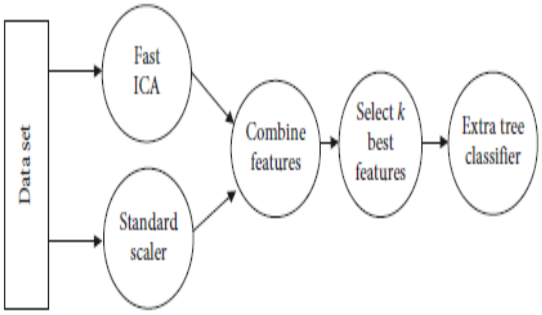



***Figure 1: Example of pipeline.***

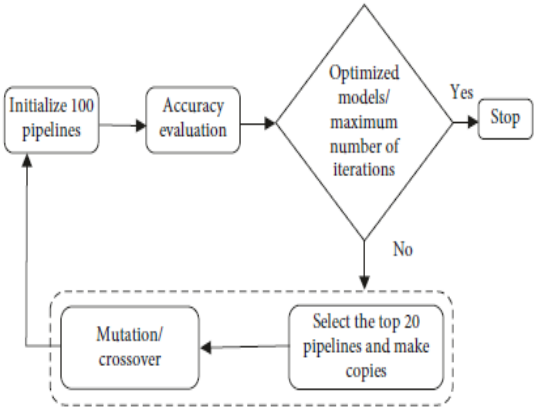



***Figure 2: Flowchart of GP.***

Features of the present image's nucleus were derived from digital photographs of a needle aspirate of a breast tumor [24]. Data from the WDBC have been studied 569 people at hospitals in Wisconsin and found 212 incidences of cancer and 357 cases of benign disease. The data points in each table are the results of individual FNA tests. The indenter number and diagnostic status are the rest two characteristics in this data collection. The remaining 30 qualities are the true ones, and they include the mean, the standard error, and the bottom 10 characteristics of the nucleus of each cell. The radius, texture, perimeter, area, smoothness, compactness, concave points, concavity, symmetry, and fractal dimension are the 10 actual values evaluated.

## 2.2. Related Work.

Feature selection is a necessary step in machine learning, in which a meaningful subset of

characteristics is selected from a pool of candidates. When building a predictive model, feature selection is crucial. Using feature selection methods has several advantages: (a) it speeds up and improves the efficiency of training the machine learning algorithm; (b) it simplifies and clarifies a model; (c) it increases a model's accuracy if the right subset is selected; and (d) it cuts down on over fitting. Because the features may have intricate dependencies on one another, picking the optimal subset is notoriously difficult [25]. There are a number of methods for detecting breast cancer that have been suggested in the scientific literature [7, 17–20]. lter, wrapper, and embedding methods are the usual buckets into which feature selection techniques are placed when categorized [26].The e lter technique is often used as a preprocessing step because to its reliance on broad characteristics. The procedure for selecting the subset does not rely on the method of instruction being used. The wrapper method picks the best characteristics by using machine learning algorithms. As can be seen in Figure 3, the learning process acts as a guide for feature selection. *It's common practice to employ a "wrapper" approach on top of a method for selecting and eliminating features.

The functionality of both filters and wrapper methods are combined in embedded methods. Algorithms are used to put these into action, and they often come equipped with their own feature selection procedures. They are unique to the particular learning machines being used and are responsible for the variable selection stage of the learning process. Figure 4 is a graphic depicting a data stream. The investigations in this paper made use of wrapper approaches.

## 2.3. The Proposed Method.

To create an automated machine learning workflow, a "pipeline" is formed by connecting many sequentially completed modules into a single unit. It gives you access to elite machine learning process by abstraction and greatly reducing the overall complexity of the procedure. Extract, Transform, and Load (ETL) processes are the most common form of this. The depth of a neural network, the number of hidden layers, the learning rate, the batch size, and the degree of regularization are all hyperparameters that may affect the performance of a machine learning method. The goal of this effort is to determine the best combination of data transformations and machine learning algorithms to carry out the categorization. It might be challenging to find the right mix of machine learning algorithm and data. Hyperparameters tuning has led to the suggestion of using genetic programming (GP) [22] to fine-tune the model's input data and output control parameters. To determine the optimal combination that yields the greatest evaluation outcomes, it is required to use this well-known evolutionary strategy. A predetermined number of pipes (the population) are generated at random using e GP. In this study, we use a classification score to assess each pipeline in the population; this score, called "fitness," is based on supervised models from the sickest-learn package. Except for linear discriminant analysis, all of the classifiers in this study use a random selection of hyperparameters to determine how well they perform. Many practical methods were evaluated in this work for use in subsequent breast cancer dataset processing and analysis stages.

2.3.1. The first phase is called "preprocessing." The raw breast cancer data was processed in this study to scale the characteristics using the Standard Scaler module. Many estimators used in machine learning assume that data has been standardized. Using the formula (xi-mean(x))/ stdev(x), where stdev(x) is the standard deviation, it converts the characteristics to a Gaussian distribution. In order to do its transformations, *e Robust Scalar uses the

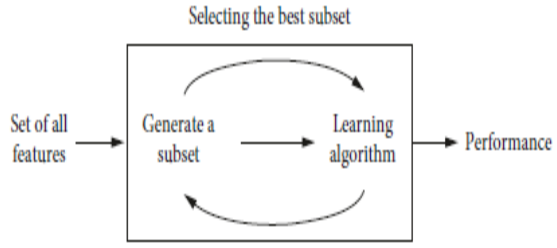



***Figure 3: Wrapper methods.***

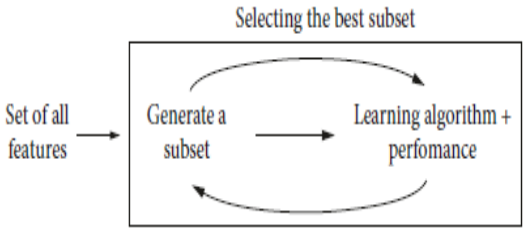



***Figure 4: Embedded methods.***

Characteristics based on the ratio of the quartile differences (Q3(x)-Q1(x)) to the quartile differences (Q1(x)-Q2(x)). The scikit-learn library [27] for machine learning contains all the modifications that were applied. The second phase, which entails choosing features, is described in Section 2.3.2. Typically, feature selection is used before any actual learning is done as a form of preprocessing. However, without informative and discriminative

features, no algorithm can make accurate predictions; thus, we implemented PCA with randomized SVD [28] to retain the most important features while shrinking the dataset. The Python scikit-learn package was used to construct the feature-selection module. There were too many criteria used by all selection procedures for useful feature extraction. As part of our effort, we eliminated features with low variance, used univariate feature selection, and recursively eliminated features.

Step 3: Apply an Algorithm for Machine Learning. In most cases, the prediction performance of an ensemble of machine learning algorithms exceeds that of a single model. One interpretation of *is as a machine learning competition is that the winning answer was put to use in a model for detecting breast cancer. 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] were used to classify the given data set in this paper. Fourth, perfecting the parameters. The genetic algorithm has been generalized to create evolutionary algorithms (EAs), one of which is known as genetic programming (GP). GP is a methodology for evaluating options and picking the most suitable one. GP finds an answer by mimicking the basic mechanisms of biological evolution (mutation, crossover, and selection).

GP's adaptability originates from the fact that it may be used to model systems for which neither the required model structure nor its important attributes are known in advance. In this research, GP was used to optimize the tree-based pipelines for the classification issue, allowing the system to search for models from a variety of alternative model architectures. First, GP creates a set number of pipelines using primitives like features selection decomposition. In other words, machine learning pipelines are generated by an evolving series of operators and then assessed to improve classification accuracy. The machine learning pipeline shown in Figure 1 is only one possible configuration. Machine learning pipelines are iterated; the best of each generation are used to inform the next. An individual of GP, each pipeline is treated as such. *e GP is composed of the following three companies:

Operator for mutation: adjusting hyperparameters or adding/removing a simple preprocessing step (e.g., Standard Scalar, random forest size). The crossover operator makes the conservative assumption that 5% of individuals will mate with each other through a random 1-point crossover. The primary goal of the selection operator is to choose the best twenty people and then replicate them. The crossover or mutation operator allows members of a population to share genetic information with one another. Figure 2 depicts the successive GP phases.

## 3. Results

The Wisconsin Breast Cancer dataset was used to check the accuracy of the models developed in this research. One training and test, as stated by Breiman et al. [38], on a small dataset, test partitions do not provide reliable estimates of the classification error scheme. To reduce the possibility of estimating errors, we (*us) opted to use a random sub sampling approach in this study. Over fitting is something you want to avoid, thus cross-validation is a great tool for doing so. So, the breast cancer data set was subjected to cross-validation with a 10-fold sample size. Three training trials were set up as part of the investigation. In the first example, it was the feature selection procedure that was of primary importance. The categorization model was the primary focus of the second trial. The third experiment's primary goal was to integrate all the previous ones into a single, self-controlling procedure. What this means is that programmatic optimization and creation of machine learning algorithms was the target. Particle swarm optimization (PSO) [39], genetic algorithm (GA) [40], evolutionary programming (EP) [22], and best first (BF) [27] were used to extract features based on EA in the first experiment using the open-source machine learning software WEKA. Table 1 displays the chosen characteristics used in the previous search strategies. From the data in the table, we can infer that the feature sets used by each of the evaluated algorithms are roughly comparable. Based on the outcomes of the implemented filter features, we were able to deduce that only 60% of the algorithms in Table 1: Feature-Selection Algorithm Comparison was effective.

| 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 |

The EA shared 80% of its properties with the other approaches. Relevance was determined for a given feature using filtering approaches, which focus just on that characteristic. Using a relevance score to filter out irrelevant information. In addition, a tuning parameter was necessary for every search technique used. Since there is no one "best" features selection technique, as stated by Yong et al. [25], we believe that the prediction accuracy of the applied classifiers is enhanced by combining multiple feature extraction methods. We confirmed that the selected model's performance was enhanced by characteristics extracted using hybrid approaches. The second study examined the performance of several well-known supervised learning algorithms used to classify the issue. Different metrics are used to assess the quality of a machine learning algorithm's proposed model. Accuracy, area under the curve, confusion matrix, and precision-recall were utilized as performance measures. The number of misclassified samples may be inferred from the accuracy (ACC) metric, which measures the classifier's accurate prediction. Specifically, it means

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

Where TP, FP, TN, and FN are the projected true positive, false positive, true negative, and false negative rates.

The following are the definitions of the remaining confusion matrix-based metrics:

$$\mathrm{recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \quad (2)$$

$$\mathrm{F1} = 2 \times \frac{(\mathrm{precision} \times \mathrm{recall})}{(\mathrm{precision} + \mathrm{recall})}.$$

Receiver operator characteristic (ROC) graphs [41] were used to depict the correlation between sensitivity (recall) and specificity in addition to the aforementioned measures. The LR, LDA, K-neighbors' classifier, DT classifier, GNB, RF classifier, additional trees classifier, AB, and GB models were used in this experiment, and their respective ROC curves show the performance of the learning technique without accounting for class distribution or error overheads. True positives and false positives are the x and y axes of the ROC space, as was previously explained. The ROC curve is a statistical measure that averages performance at several cutoff points. *Classification models that score below the diagonal of the ROC graph are regarded as being poorer than random guessing. A perfect classifier would have a true positive rate of 1 and a false positive rate of 0. This would place it in the upper left corner of the graph. The area under the ROC curve (AUC) is a measure of how well a classification model performs. It is shown that applied models can provide more precise predictions than *us. The results of the nine computational models are compared in Figures 5-13. We observed that GNB had a higher mean ROC of 77% in our trial. Classification error schemes based on a single train and test partition are not accurate estimators, as stated by Breiman et al. [38]. To reduce the potential for bias in the estimate process, we chose to use a random sub sampling approach. The breast cancer data set was subjected to cross-validation with a ten-fold sample size. We used five-fold cross-validation to improve the clarity of the ROC values. In the previous experiment, we used the default settings for the entire machine learning classifiers' input parameters and found that the LR, LDA, and GNB algorithms provided the best fit.

## 4. Discussion

Experiments show that combining features selection approaches results in better accuracy performance, proving our hypothesis. Because of this, the genetic programming strategy was suggested. As a means of building the predetermined number of conduits. As a result, a variety of machine learning techniques were used to automate the process of selecting the optimal pipeline. Therefore, the proposed method is viewed as a potential means to pick the right algorithm and fine-tune hyperparameters for optimal model performance .the hyperparameters regulate the complexity of the selected model and are used in model selection but are not directly learnt by the classifiers. *There are many hyperparameters settings to choose from; therefore it's important to do your research. Model parameters in machine learning techniques are chosen arbitrarily. Therefore, many researchers resort to manual adjustments. In a nutshell, the prediction performance of the learning algorithm and the complexity of the model are both impacted by the choice of the control parameters. The goal of this effort was to investigate the hyperparameters issue. *e experiment had three

distinct phases. The first experiment compared two of the most well-known evolutionary algorithms, PSO and GA, for selecting characteristics. This investigation demonstrated that 80% of the targeted characteristics were shared Many adjustable parameters were typical of evolution-based methods. There is a risk of author bias in method selection since it is sometimes impossible to claim familiarity with all established techniques. We employed a refined control setting to counteract this slant.

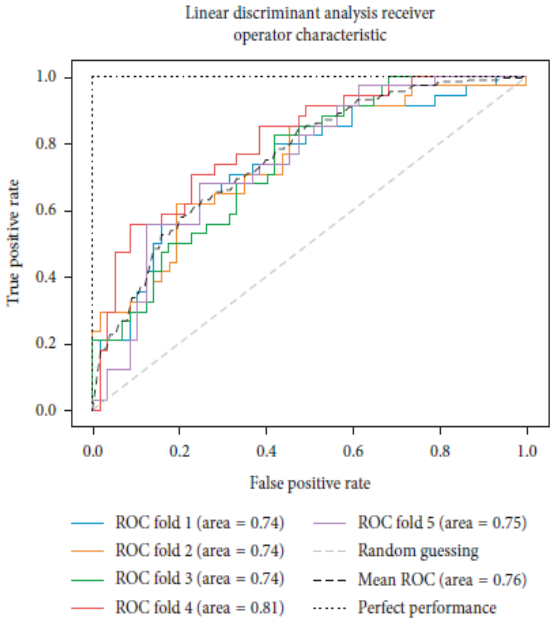



***Figure 5: ROC curve for LDA.***

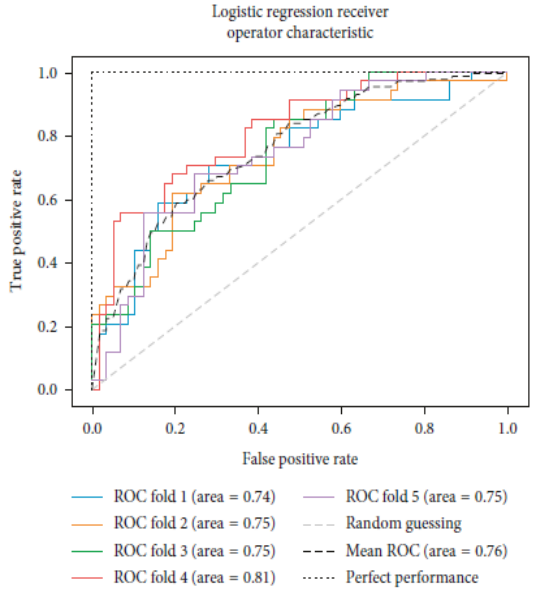



***Figure 6: ROC curve for LR.***

Given an algorithm with n parameters, the resulting configuration space is a hypercube of n dimensions. We considered using this space to implement a straightforward approach for identifying useful characteristics. Consequently, PCA was selected for dimensionality reduction in the breast cancer dataset. On the one hand, there was good reason to standardize feature selection, but on the other, the old methods needed a few parameters. The authors in [42] also made similar decision, arguing that feature extraction and feature selection have the benefits of (a) maintaining the data's interpretability and (b) increasing its discriminative potential. This is evident from the placement of the courses in Figure 14.

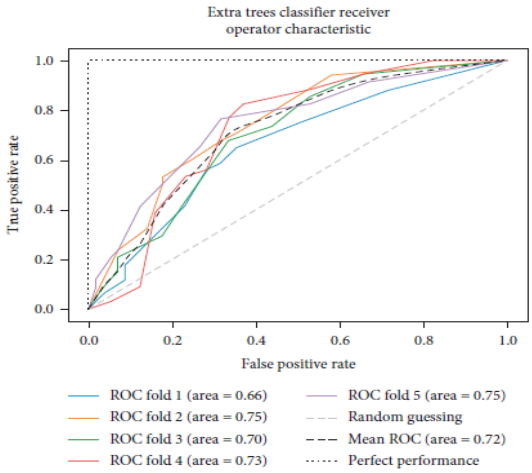



***Figure 7: ROC curve for ET.***

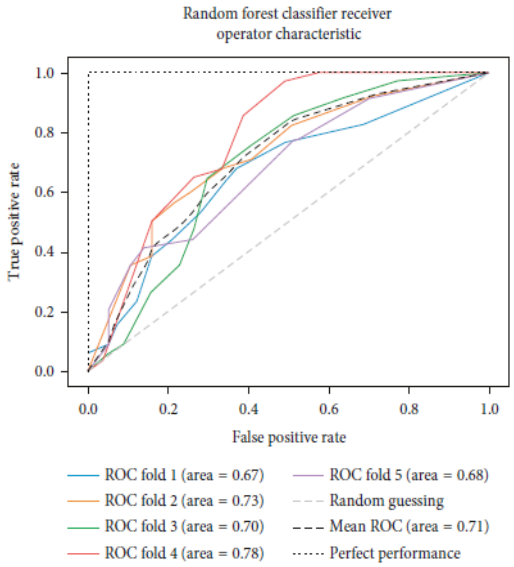



***Figure 8: ROC curve for RF.***

Standardized feature selection and principal component analysis provide linearly separable data.

The researcher's second problem is deciding which machine learning algorithm is best? When deciding on a machine learning algorithm, many factors, including accuracy and complexity, are typically taken into account. Many users, however, focus only on precision. Subsequently, some authors assert that the performance of their algorithms surpasses that of previously reported algorithms. In order to get the most performance out of a machine learning approach, hyperparameters selection and a lot of training are usually necessary.

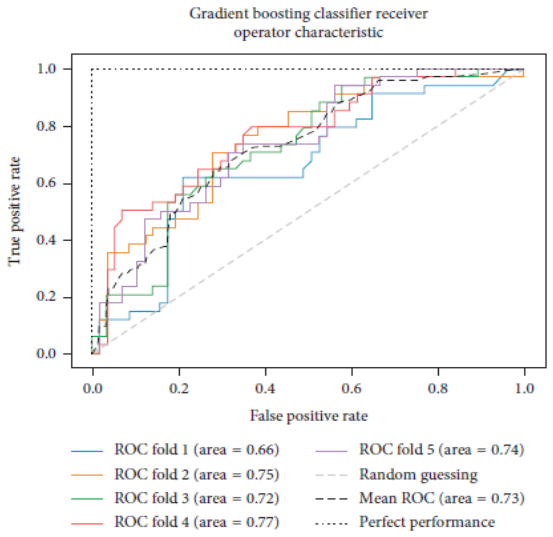



***Figure 9: ROC curve for GB.***

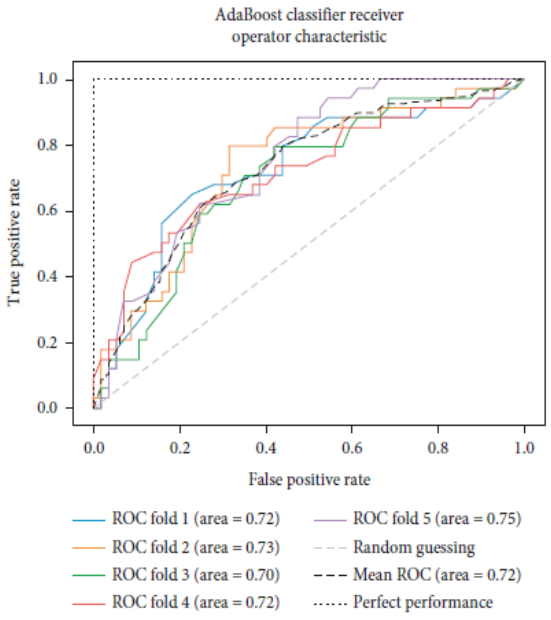



***Figure 10: ROC curve for AB.***

The "No Free Lunch" theorem [[Wolpert and Macready]] states that no algorithm is universally applicable. As a result, there are a lot of methods that need to be tested to specific issue before choosing a victor. We evaluated several different machine learning algorithms, including KNN, SVM classification, DT, RF, AB, GB, GNB, LDA, quadratic discriminant analysis, LR, and extras classifier. Accuracy and log-loss were both helpful measures for this experiment. Useful machine learning's accuracy comparison is shown in Figure 15, while the log-loss is shown in Figure 16. Table 2 suggests that AdaBoost is the most accurate classifier, with a success rate of 98.24%.

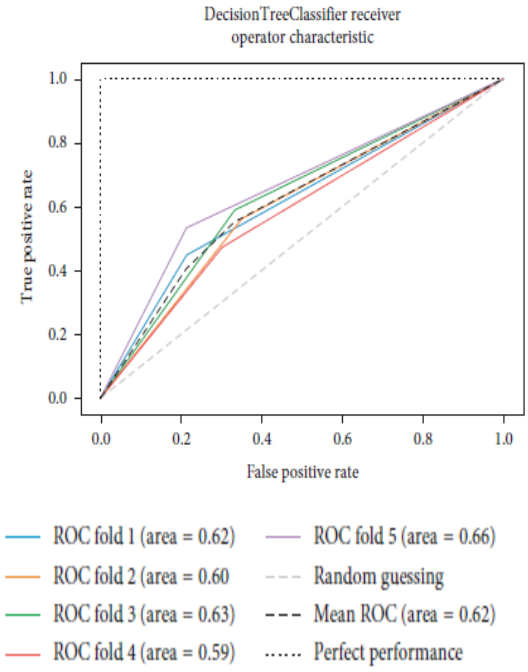



***Figure 11: ROC curve for DT.***

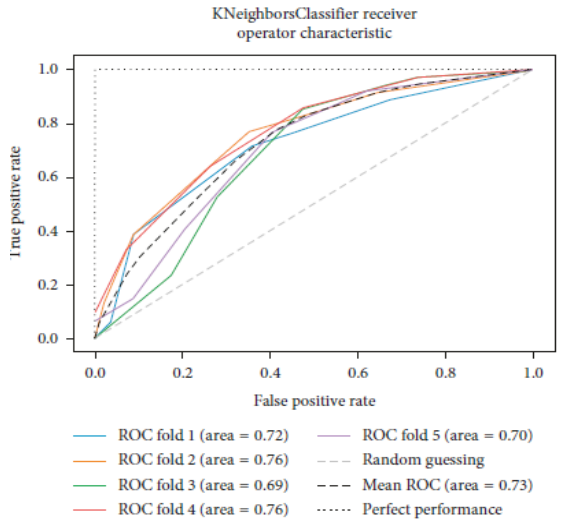



***Figure 12: ROC curve for KNN.***

However, this is not right, as Table 3 shows that the AdaBoosting classifier has a log-loss measure of 0.39. The log-loss provides a more precise picture of the model, which is well recognized. Based on the data, we identified three top performers: the GB classifier, the RF classifier, and the additional tree classifier. The standard deviation of the average accuracies is used to illustrate the variance of the estimate in Figure 17. The accuracy curves in the training set and the cross-validation set differ greatly, as seen in e figure. As a consequence, the findings up to this point are supported by the accuracy curve. As was previously indicated, a practical model's performance is very sensitive to the settings of its control parameters. For this reason, we attempted to mechanize everything from feature selection through categorization. Because of GP, we were able to create a wide variety of configurations using the preexisting modules, as shown in Section 2. The settings of control were fine-tuned for each component of the structure that was generated arbitrarily. To create a random-building model, one may use the Standard Scaler module to normalize the input data, the RFE module to minimize the amount of features, and the logic regression module for classification. In this effort, GP methods were used to fine-tune the relevant control parameters of the selected algorithms. The only time a human is required is during the initialization of the GP parameters (population size, generation numbers, etc.). For the many features to choose from,

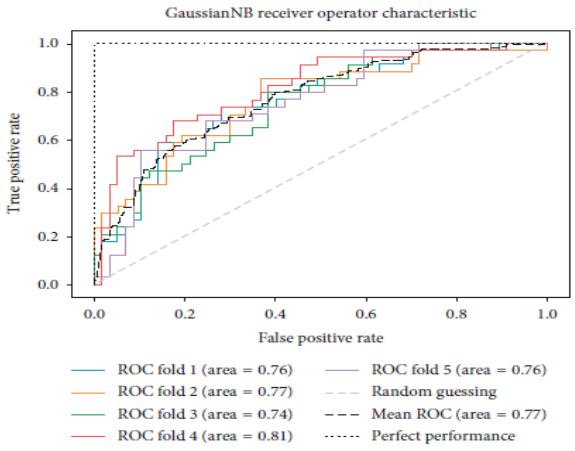



***Figure 13: ROC curve for GNB.***

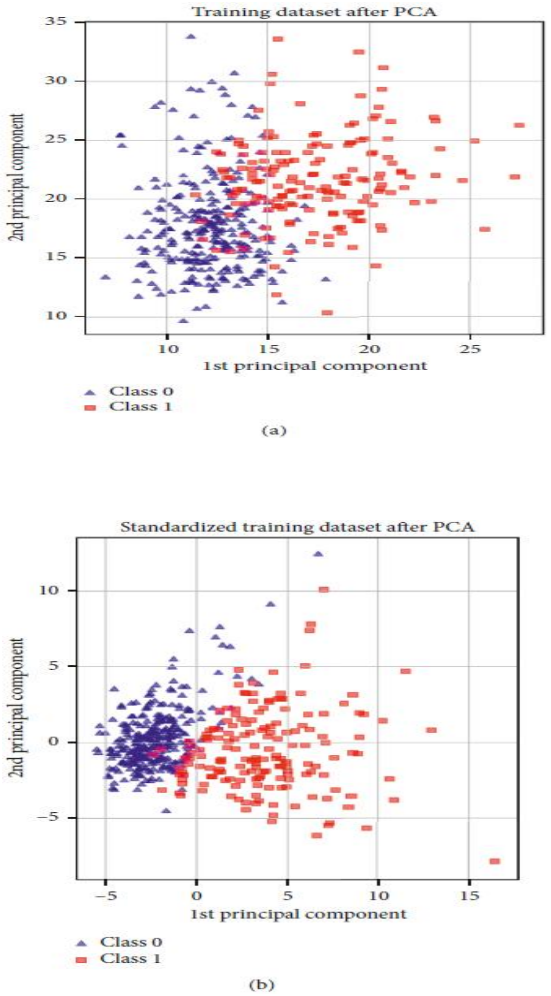



***Figure 14: Combining feature extraction.***

In order to reduce the time spent on evaluations, it is best to choose a small number of algorithms for each methodology and classification method. We've

already covered the reasoning for our method selection. We maintained the model with the highest accuracy after comparing it to the results of the randomly chosen approaches. An ensemble of methods was formed by sequentially combining the techniques used; the resulting model included the MaxAbsSclaer operator for the preprocessing stage, the polynomial features operator for feature selection, and the gradient boosting classier as the model for supervised classification. The validation accuracy obtained was 98.24%.

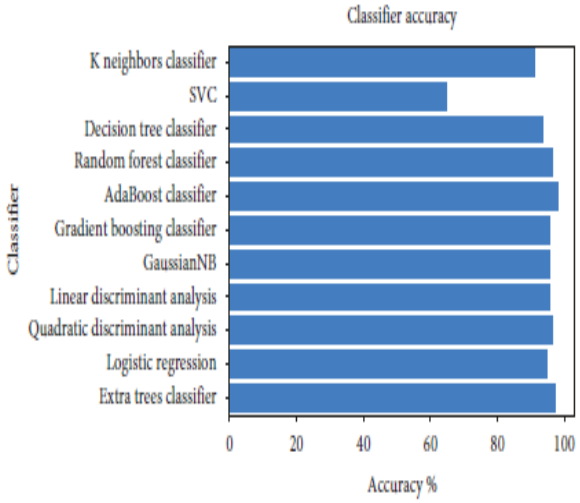
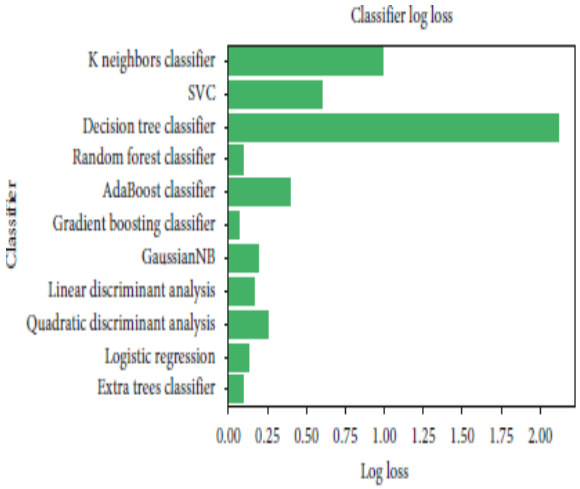



***Figure 15: Comparison of classifier accuracy.***




***Log loss classifier comparison, shown in Figure 16.***

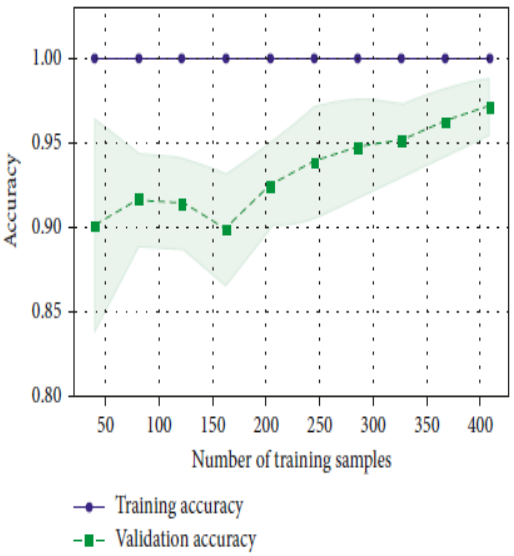



***Figure 17: Validation accuracy.***

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 | 78.72 | 93.42 | 98.67 | 96.73 | 0.97.26 | 96.10 | 98.01 |
| Malignant (%) | 93.51 | 90.91 | 94.74 | 93.51 | 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 | 0.19 | 0.59 | 0.992 | 0.39 | 0.16 | 0.25 | 0.13 | 0.09 |

## 5. Conclusions

This investigation employs a machine learning method to address the challenge of autonomous breast cancer diagnosis. The breast cancer dataset was used in a variety of investigations. In the first experiment, we showed that, with proper setup, the three most widely used evolutionary algorithms produce identical results. The second experiment investigated the hypothesis that using several feature selection techniques simultaneously boosts accuracy. In the last experiment, we automatically designed a supervised classifier using machine learning. We used the GP technique to try to fix the hyperparameters issue, which is difficult to tackle for ML algorithms. The suggested method determined which configuration best suited the problem at hand. Every single experiment was coded in Python. While the proposed approach yielded significant results

through the evaluation of an ensemble of approaches from a comprehensive machine learning technique, we ran into significantly higher time consumption than was initially anticipated. At the end of the day, the suggested model seems to be well-suited for both automated breast cancer detection and determining the control parameters of machine learning algorithms.

## 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.*