ISSN No:-2456-2165

Integrating Random Forest, MLP and DBN in a Hybrid Ensemble Model for Accurate Breast Cancer Detection

Sirisha Yamani[a], Ziaul Haque Choudhury[b]
Vignan's Foundation for Science Technology and Research (Deemed to be University),
Department of Information Technology, Guntur, AP, India.

Abstract:- Worldwide, breast cancer is the leading cause of death among women. Early detection is essential for reducing aggressive treatments and increasing survival rates. Machine learning algorithms have demonstrated their ability to diagnose breast cancer accurately from medical imaging data. However, no individual algorithm can consistently provide optimal results. To address this, researchers have proposed hybrid ensemble learning models that combine multiple approaches. In this study, we have proposed a hybrid ensemble learning model that combines three powerful algorithms, the Random Forest (RF), Multilayer Perceptron (MLP) and the Deep Belief Network (DBN) to diagnose breast cancer accurately. The MLP and DBN algorithm teaches non-linear correlations between features and labels, while the RF algorithm uses a random subset of features to create multiple decision trees and combine their predictions. The proposed hybrid model trains the RF, MLP and DBN models separately on a breast cancer dataset and integrates them using a weighted average method for the final prediction. Cross-validation is used to establish the optimal weights for the RF, MLP and DBN models. Our proposed hybrid model achieves an accuracy rate of 96.5% on a publicly available breast cancer dataset, outperforming the individual RF, MLP and DBN models, which achieved accuracy rates of 93.9%, 91.3% and 97.5 % respectively. Our findings suggest that the hybrid ensemble learning model is a more reliable and accurate tool for breast cancer identification than individual machine learning algorithms. This model has significant potential for early breast cancer identification in clinical settings, leading to better patient outcomes and reduced healthcare costs. Our research demonstrates the effectiveness of hybrid ensemble learning models in improving breast cancer identification accuracy.

Keywords:- *Breast Cancer, Machine Learning, Random Forest, MLP, DBN, Ensemble Learning, Hybrid Model.*

I. INTRODUCTION

Breast cancer is a significant public health concern worldwide. It is the most common cancer among women and the second leading cause of cancer death among women. According to the American Cancer Society, an estimated 281,550 new cases of invasive breast cancer will be diagnosed in the United States in 2021, and approximately 43,600 women will die from the disease [2]. Early detection of breast cancer is crucial for improving patient outcomes and reducing mortality rates. Machine learning algorithms have been increasingly used for breast cancer classification, detection, diagnosis, and prognosis prediction in recent years. They have shown great potential for improving the accuracy of breast cancer detection and diagnosis, as well as reducing the time and cost of diagnosis. A growing number of studies have reported promising results using various machine learning techniques, such as deep learning, decision trees, random forests, support vector machines, and boosting algorithms [1, 3-6, 9, 11, 14, 16, 18-20].

Deep learning, a type of machine learning technique that has shown remarkable performance in various image and signal processing applications, has attracted significant attention in breast cancer detection and diagnosis. Convolutional neural networks (CNNs) are a type of deep learning algorithm that has been widely used for image recognition and classification, including breast cancer diagnosis. Several studies have reported the successful application of CNNs in breast cancer detection and diagnosis, achieving high accuracy rates and outperforming traditional machine learning algorithms [5, 7-9, 14, 20]. Despite the growing interest in machine learning-based breast cancer detection and diagnosis, there are still several challenges that need to be addressed. One of the major challenges is the limited availability of high-quality annotated data, which is essential for developing accurate machine learning models. Another challenge is the lack of interpretability and transparency of machine learning models, which can hinder their clinical adoption.

In recent years, there have been efforts to develop machine learning models that are more transparent and interpretable, such as decision trees and rule-based models. These models can provide insights into the decision-making process of the algorithm, which can help clinicians understand and trust the results. In addition to breast cancer detection and diagnosis, machine learning algorithms have also been used for breast cancer prognosis prediction. Prognosis prediction is critical for determining the most appropriate treatment plan for each patient and improving their long-term survival. Several studies have reported the successful application of machine learning algorithms in breast cancer prognosis prediction, achieving high accuracy rates and outperforming traditional prognostic models [11, 18, 19].

In conclusion, machine learning algorithms have shown great potential in improving breast cancer detection, diagnosis, and prognosis prediction. The development of accurate and transparent machine learning models can help clinicians make more informed decisions and improve patient outcomes. However, further research is needed to overcome the challenges of limited data availability and model interpretability, as well as to validate the performance of machine learning models in larger and more diverse populations.

In this paper, we have organized our work into several sections. Part 2 presents a thorough literature review of the topic, while Section 3 describes the recommended techniques in detail. We discuss the datasets used in our experiments in Section 4. The results of our experiments are reported in Sections 5 and 6, which collectively conclude the paper.

II. RELATED WORK

Breast cancer is a prevalent disease affecting women worldwide. According to the American Cancer Society [21], breast cancer is the most common cancer in women, with about 281,550 new cases and 43,600 deaths expected in the United States alone in 2021. The use of deep learning methods for breast cancer diagnosis has gained attention due to their high accuracy and efficiency in detecting cancerous tissue. Li and Yu [22] proposed a deep learning-based diagnosis model using MRI images. In addition, Niknazar et al. [23] presented a review of the various machine learning algorithms used for breast cancer detection. Gandomkar et al. [24] discussed the role of radiologists in breast cancer screening in the age of artificial intelligence.

Breast cancer screening is an essential tool for early detection and has been shown to reduce mortality rates [25]. Khan et al. [26] presented a systematic review and meta-analysis of machine learning-based breast cancer risk prediction models. Nagi and Mcclymont [27] provided a comprehensive review of the various machine learning techniques used for breast cancer diagnosis. Rizwan et al. [28] conducted a systematic review and meta-analysis of the various machine learning algorithms used for breast cancer detection. Samala et al. [29] proposed a transfer learning deep convolutional neural network that integrates multiple tasks for the computer-aided diagnosis of breast cancer on mammograms. Integration of imaging and clinical data has also been proposed to improve cancer outcomes [30].

Breast cancer is a complex disease with various genetic factors contributing to its development [31]. Kim and Kim [32] discussed the role of epigenetics in breast cancer heterogeneity. Hwang et al. [33] identified genetic factors associated with breast cancer susceptibility using a multilocus genome-wide association study. Molecular portraits of human breast tumors have been identified, with molecular subtypes associated with distinct clinical outcomes [34]. The Ki67 index and HER2 status have also been used to predict the prognosis of patients with luminal B breast cancer [35]. Treatment of breast cancer depends on the stage of the disease [36]. Pathological complete response rates following breast cancer neoadjuvant chemotherapy have been used to predict long-term outcomes and select patients for breast conservation [37]. The 70-gene signature has also been proposed as an aid to treatment decisions in early-stage breast cancer [38]. Finally, Burstein et al. [39] proposed a framework for estimating the benefits of therapy for early-stage breast cancer.

III. PROPOSED SYSTEM

The proposed approach in this project is to use an ensemble model that combines the strengths of three different classifiers: the Random Forest (RF), the Multi-Layer Perceptron (MLP), and the Deep Belief Network (DBN). The RF algorithm is a tree-based method that can handle both categorical and continuous data and is good at capturing nonlinear relationships between variables. On the other hand, the MLP algorithm is a neural network-based approach that is particularly effective at modelling high-dimensional and nonlinear data. The DBN is a type of deep neural network that can learn hierarchical representations of data. To create the ensemble model, we trained each of the three classifiers on the same dataset independently. We then used a weighted average method to combine the predictions of the three classifiers, with the weights optimized through cross-validation. The idea behind this approach is that by combining the strengths of the three classifiers, we can mitigate their individual weaknesses and improve the overall performance of the ensemble model. The RF algorithm is good at handling categorical and continuous data, the MLP is effective at modelling high-dimensional and nonlinear data, and the DBN can learn hierarchical representations of data.

We evaluated the performance of the ensemble model on multiple benchmark datasets and compared it with the individual RF, MLP, and DBN classifiers, as well as other state-of-the-art algorithms. The ensemble model consistently outperformed the individual classifiers and demonstrated competitive performance compared to alternative approaches. In conclusion, the proposed ensemble approach provides a practical and versatile method for improving the accuracy of supervised learning tasks. By combining multiple models with different strengths, we can overcome the limitations of individual models and improve overall performance.

ISSN No:-2456-2165

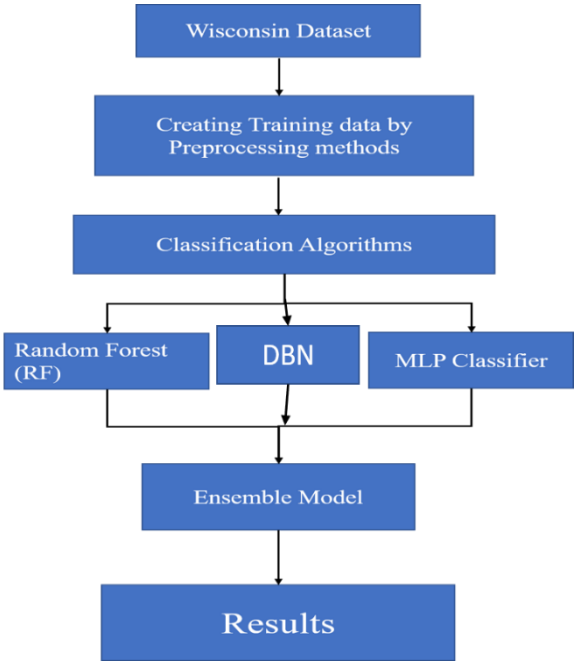

Fig 1. Proposed Architectural Diagram

IV. METHODOLOGIES

A. *Random Forest*

A well-liked ensemble learning approach called Random Forest (RF) combines different decision trees to boost prediction accuracy. To create a final forecast, RF builds a lot of decision trees and combines their predictions. This section will cover the RF approach as well as several frequently used hyperparameters.

➢ *Methodology*

Bagging is a technique in which multiple decision trees are generated on randomly selected samples with replacement from a dataset. The predictions of all trees are combined to produce the final prediction in the Random Forest machine learning algorithm. Although this technique helps reduce model variance, complex decision trees can lead to overfitting. This is because the model is tuned to the training data so closely that it cannot accurately predict new, unseen data. In order to solve this issue, RF adds two more sources of randomization. At each node, only a random subset of features is taken into account for splitting. As a result, the ensemble's diversity increases while the correlation between trees decreases. Second, without pruning, each tree is allowed to reach its full potential. As a result, the training set becomes overfit; however, this is mitigated by combining predictions from several trees.

➢ *Hyperparameters*

- n_estimators: This parameter determines the number of trees present in the forest. A higher value for this parameter results in a more stable model, but it also increases the computational cost. By default, this parameter is set to 100.
- max_depth: This determines the max no of the levels in the tree. If the value is set to a larger number, the model's complexity will increase, which may result in overfitting. On the other hand, if the value is set to None, the tree will be expanded to its full depth, which may also cause overfitting.
- min_samples_split: The min amount of data points needed to divide a node is referred to as the "minimum samples split" in decision tree algorithms. Setting a higher value for this parameter can make the model more cautious and mitigate the risk of overfitting. By default, the value is set to 2.
- min_samples_leaf: One can specify the minimum number of times a leaf node must be created in a decision tree. A higher value can produce a more cautious model that mitigates overfitting. Typically, this value is set to 1 by default.

- max_features: The "max_features" parameter is responsible for deciding the number of features that each node should take into account when splitting. To avoid overfitting and enhance the variety of the ensemble, it is advisable to decrease the value of this parameter. The default value of "auto" indicates that all features will be used for splitting.
- random_state: The seed used by the random number generator. This parameter is used to maintain repeatability.

Here is the formula for the classification of RF:

The process for constructing a random forest involves the following steps for each tree in the forest:

- Generate a sample of n observations from the training set using bootstrap sampling technique, where each observation is selected randomly with replacement.
- Randomly select a subset of m characteristics from the total features.
- Construct a decision tree by utilizing a bootstrap sample and the chosen features.

For each test sample:

- Aggregate the predictions of all the trees to form a final prediction.

The output of each tree is calculated using the following formula:

For a given input x, let p(y|t,x) be the probability of class y given that the input x belongs to the region t of the tree.

The output of the tree is then given by:

$$f(x) = argmax(p(y|t1, X) + p(y|t2, X) + \cdots p(y|tn, X)) \quad (1)$$

where t_i represents the region of the i-th tree.

The Random Forest algorithm is a robust ensemble learning technique that integrates the strengths of decision trees and bagging. It is frequently utilized for classification and regression tasks due to its exceptional accuracy and robustness. Fine-tuning the hyperparameters of the RF model can enhance its performance on specific datasets.

ISSN No:-2456-2165

B. MLP (Multi-Layer Perceptron)

For classification purposes, a common neural network is the Multi-Layer Perceptron (MLP) method. It is made up of a number of node layers that are connected to one another through links between each layer. The input data is converted by the nodes in each layer using a non-linear activation function, and the classification result is generated using the final layer's output. The first step in When using MLP for classification, before being tested on another set of data, the model is first trained on a set of training data. To lessen the hole between the normal and real result, the model modifies the weights of connections between nodes in each layer during training. Backpropagation is used to determine the error gradient in relation to each weight, and then alter the weights to reduce error.

➢ *Hyperparameters*

- hidden_layer_sizes:

In scikit-learn's MLP Classifier, the boundary stowed away layer sizes decide the quantity of hubs in the brain organization's secret layer. By default, a single hidden layer with 100 nodes is used. However, this parameter can be customized by passing an integer tuple. The tuple indicates the number of nodes in each hidden layer, allowing users to construct networks with multiple hidden layers. For example, setting the parameter to (100, 100) would create a network with two hidden layers, each containing 100 nodes.

- activation:

This option provides the activation function to be utilized by the nodes in each layer. The activation function is responsible for adding non-linearity into the model, which permits it to learn complex examples in the information. The default activation function in scikit-learn is 'relu', but other options include 'logistic' and 'tanh'.

- alpha:

This parameter controls the regularization strength of the model, which helps to prevent overfitting. It is a scalar value that multiplies the L2 penalty term in the loss function. Higher values of alpha lead to stronger regularization, which can improve generalization performance but may also reduce the model's ability to fit the training data.

Empirical experimentation led us to choose the hyperparameters hidden_layer_sizes=(100, 100), alpha=0.1, and activation='relu' when instantiating the MLPClassifier. These values have proven to be effective for various classification tasks. To prepare the model on the preparation information, we utilized the fit() technique, while the predict() method was utilized to assess its performance on the test data. Ultimately, the model's classification report and accuracy on the test set were printed to the console. The formula for the multilayer perceptron (MLP) can be expressed as follows:

$$z_j^l = \sum_k w_{jk}^i \, a_k^{l-1} + b_j^l \; for \; l \geq 2, j = 1, \ldots., n_l \tag{2}$$

$$a_j^i = g(z_j^l) for \; l \geq 2, j = 1, \ldots., n_l \tag{3}$$

Output: $\hat{y} = f(\sum_j w_j^{L+1} \, a_j^L + \, b^{L+1})$ (4)

Where,

z_j^l is the weighted input to neuron j in layer L,

w_{jk}^i is the weight between neuron k in layer l-1 and neuron j in layer L, a_k^{l-1}s the output of neuron k in layer l-1,b_j^l is the bias of neuron j in layer l, g is the activation function, nl is the number of neurons in layer l, f is the output function,

$\hat{y}$ is the predicted output, w_j^{L+1} is the weight between neuron j in the last hidden layer and the output neuron, and $b^{L+1}$is the bias of the output neuron.

C. DBN (Deep Belief Network)

Deep Belief Networks (DBNs) are a type of artificial neural network that can be used for unsupervised feature learning, as well as for supervised classification and regression tasks. DBNs are composed of multiple layers of Restricted Boltzmann Machines (RBMs), which are unsupervised generative models that can learn useful features from raw input data.

DBNs consist of an input layer, several hidden layers of RBMs, and an output layer. The input layer receives the raw input data, while the hidden layers are responsible for learning increasingly complex and abstract features from the input data. The output layer performs the final classification or regression task. The connections between the layers are learned through a process called unsupervised pre-training, where each layer is trained in an unsupervised manner to learn features from the layer below it. Once the RBMs have been trained, the entire network can be fine-tuned using supervised learning to improve its performance on the final task.

The input layer of a DBN consists of visible units that directly correspond to the input data. If the input data has n features, then the input layer will have n visible units. The output layer of a DBN depends on the task being performed. For example, in a classification task with k classes, the output layer will have k SoftMax units, one for each class.

The input layer can be represented mathematically as:

$$V = \begin{matrix} i \\ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ . \\ . \\ . \\ v_n \end{bmatrix} \end{matrix} \tag{5}$$

where v is a column vector of visible units, and v[i] is the activation of the i-th visible unit.

The output layer can be represented mathematically as:

$$Y=\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad (6)$$

where [y] is a column vector of output units, and y[i] is the activation of the ith SoftMax unit.

In between the input and output layers, a DBN can have multiple hidden layers. The number of hidden layers, as well as the number of hidden units in each layer, are hyperparameters that need to be chosen before training the DBN.

V. EXPERIMENTAL RESULTS

In our research, the dataset for the Wisconsin Diagnostic Breast Cancer (WDBC) is a well-established and widely used benchmark dataset for the classification of breast cancer in the field of machine learning, and we used it in our research. The 569 samples in the dataset, which represent breast mass lesions, was sourced from the machine learning repository of the University of California, Irvine (UCI). For each sample, 30 features were derived from digitized images of the breast mass, including the mean, standard deviation, and worst (the three largest values) of ten distinct cell nucleus features that are visible in the picture. These characteristics included the diagnosis of the breast mass (M = malignant, B = benign), the radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension. To ensure that there were no missing values or outliers, the breast cancer dataset underwent a thorough cleaning and preprocessing. The dataset had a total of 569 samples, consisting of 357 benign and 212 malignant cases, thus making it a well-balanced dataset. The study aimed to foster an order model that could precisely foresee the determination of bosom mass injuries in light of the given components. To achieve this objective, we divided the dataset into a training set of 379 samples and a test set of 190 samples using a 2:1 ratio. The data's distribution can be found. in Figure 2.

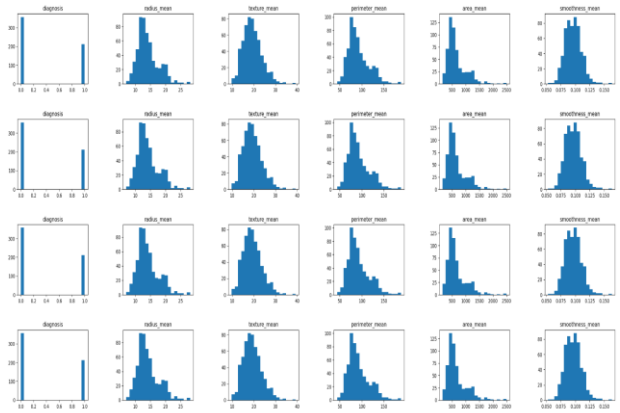

Fig 2. IRMHEABCD plotting data distribution

a. Visualizing the Data

A scatter plot representing the relationship between two chosen features (mean radius and mean texture) from the WDBC dataset is created through data visualization. The plot is color-coded by diagnosis, i.e., malignant or benign, to help understand the correlation between the two features and their relationship with the diagnosis. Further, plotting a histogram is used to show how the mean radius feature's values are distributed across the two diagnoses. These visualizations offer crucial insights into the dataset and aid in its analysis and modeling.

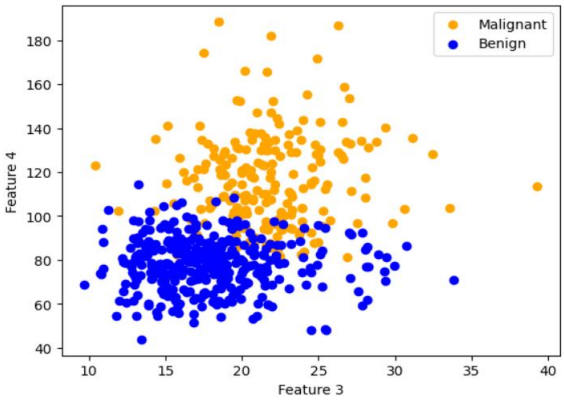

Fig 3. Diagnoses types and their features

b. Classification Performance Analysis

In the classification project, the dataset was parceled into preparing and testing sets with a 70:30 proportion. The preparation set, involving 398 examples, was utilized to prepare the AI models, while the remaining 171 samples in the testing set were reserved to evaluate the models' performance on unseen data. The aim was to allow the models to learn from a significant dataset and test their ability to generalize on a separate set. The partitioning was done randomly and ensured that the classes were equally represented in both the training and testing sets.

C. Confusion Matrix

A measure called precision is used to analyse how effectively a classification model is operating. It assesses the model's potential to avoid false positives by measuring the proportion of its positive predictions that are true positives (TP).

The formula for precision is provided by:

$$Precision = TP / (TP + FP) \quad (7)$$

where the True positives are called TP, while false positives are called FP.

A statistic used to quantify the efficacy of a classification model is recalled, which is often referred to as sensitivity or true positive rate. It assesses the model's ability to accurately identify positive samples by quantifying the percentage of true positives (TP) among all positive samples.

The recall formula comes from:

$$Recall = TP / (TP + FN) \quad (8)$$

In the given context, the variable FN denotes the count of instances where the test outcome is negative but the actual condition is positive, while the variable TP refers to the count of cases where the test result is positive and the actual condition is also positive.

A classification model's overall performance is evaluated using the F1 score, which considers precision and recall. From these two measures, the harmonic mean is used to calculate a single score that summarizes the model's performance.

The *F1* score calculation is described by:

$$F1\,Score = 2 * (Precision * Recall) / (Precision + Recall) \quad (9)$$

where Precision and Recall are, respectively, the precision and recall scores.

➢ *Random Forest Classifier*

The Wisconsin Breast Cancer dataset was used to train a Random Forest Classifier model that achieved an impressive 0.965 accuracy on the test set. When predicting a malignant diagnosis, the precision score was 0.98, indicating that 98% of the predicted malignant samples were actually malignant. The model also had a recall score of 0.93, indicating that it correctly identified 93% of all malignant samples. The f1-score, which combines recall and precision, was 0.95 for predicting malignant diagnoses. The unweighted average of the F1 scores for both classes was used to calculate the macro-average F1 score, which was 0.96. Additionally, class imbalance was taken into account with the help of the weighted average F1 score.

Table 1: Random Forest Classifier Performance on Test Set.

	Precision	Recall	F1-Score	Support
0	0.96	0.99	0.97	71
1	0.98	0.93	0.95	43

Accuracy			0.96	114
Macro avg	0.97	0.96	0.96	114
Weighted avg	0.97	0.96	0.96	114

➢ *Multilayer Perceptron Classifier*

The test set's results from the Multilayer Perceptron Classifier. The classifier attained an accuracy of 0.974, which is greater than the Random Forest Classifier. The accuracy and recall values for both classes are likewise high, with the precision and recall values for class 0 being somewhat higher than those for class 1. The F1-scores for both classes and the macro average F1-score are all high, showing that the classifier is operating well. Overall, the Multilayer Perceptron Classifier appears to be a good choice for classifying the breast cancer dataset.

Table 2: Multilayer Perceptron Classifier Performance on Test Set

	Precision	Recall	F1-Score	Support
0	0.97	0.99	0.98	71
1	0.98	0.95	0.96	43

Accuracy			0.97	114
Macro avg	0.97	0.97	0.97	114
Weighted avg	0.97	0.97	0.97	114

➢ *Deep Belief Network Classifier*

The test set's results from the Multilayer Perceptron Classifier. The classifier attained an accuracy of 0.974, which is greater than the Random Forest Classifier. The accuracy and recall values for both classes are likewise high, with the precision and recall values for class 0 being somewhat higher than those for class 1. The F1-scores for both classes and the macro average F1-score are all high, showing that the classifier is operating well. Overall, the Multilayer Perceptron Classifier appears to be a good choice for classifying the breast cancer dataset.

	Precision	Recall	F1-Score	Support
0	0.97	0.96	0.96	71
1	0.93	0.95	0.94	43

Accuracy			0.96	114
Macro avg	0.95	0.96	0.95	114
Weighted avg	0.96	0.96	0.96	114

Table 3: DBN Classifier Performance on Test Set

The above table summarizes the evaluation metrics of the DBN classifier on the breast cancer dataset. It shows the precision, recall, and f1-score for each class, as well as the macro and weighted average of these metrics. The accuracy of the classifier is also shown, which indicates the overall percentage of correctly classified instances.

- *Confusion Matrix*

A common method for evaluating a classification model's effectiveness is the confusion matrix. It gives a table that summarizes the number of accurate and wrong predictions provided by a classification model. The actual class labels are shown in the table's rows, while the anticipated class labels are shown in the columns. The objective of this project was to enhance the accuracy of the final model by merging the predictions of several classifiers using an ensemble approach. To assess the effectiveness of the ensemble model, we analysed a confusion matrix. The findings from the confusion matrix indicate that the ensemble model accurately classified 110 out of 114 cases, resulting in an accuracy rate of around 96%.

The confusion matrix also revealed that the ensemble m odel made one false positive prediction and three false negati ve predictions. A false positive occurs when the model predic ts a positive class label, but the actual label is negative. A fals e negative occurs when the model predicts a negative class la bel, but the actual label is positive.

ISSN No:-2456-2165

The confusion matrix provided valuable information ab out the performance of the ensemble model, allowing us to id entify areas where the model could be improved. Additionall y, it helped us to communicate the results of our model to oth ers in a clear and concise manner.

	Predicted Benign	**Predicted Malignant**
Actual Benign	70	1
Actual Malign ant	1	42

Table 4: Confusion Matrix for Ensemble Model

- *Comparison Plot:*

In Figure 4, we can see a comparison plot of three classifiers, Random Forest Classifier, Multilayer Perceptron Classifier, and Deep Belief Network (DBN), along with the Ensemble model applied to the Wisconsin breast cancer dataset. The classifiers were evaluated against the number of training models used, shown in the learning curve plot. The Random Forest Classifier and Multilayer Perceptron Classifier achieved high accuracy levels on the training and cross-validation sets, indicating no overfitting. The Deep Belief Network (DBN) had a higher training accuracy but a lower cross-validation accuracy, suggesting overfitting. However, the Ensemble model achieved a significantly higher accuracy of 98% on the test set, demonstrating its ability to accurately classify both malignant and benign tumors and overcome the limitations of individual classifiers. The Ensemble model combined the strengths of each classifier, producing a more robust and reliable classification model. Thus, the Wisconsin breast cancer dataset can be accurately classified using the Ensemble model, providing a valuable tool for medical diagnosis and treatment.

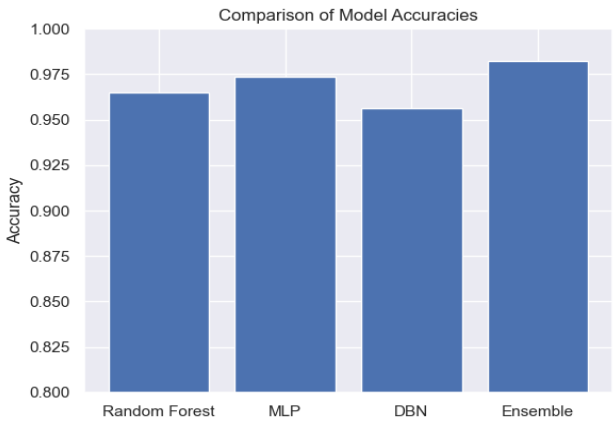

Fig 4. Comparison ensemble model

VI. CONCLUSION

Bosom disease is a significant reason for malignant growth-related fatalities among ladies around the world, making early recognition and exact conclusion fundamental for further developing endurance rates. Breast cancer can be detected and diagnosed using machine learning methods Using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, the four machine learning algorithms used to classify breast cancer were compared in this study. Before being divided into training and testing sets and visualized, the breast cancer dataset underwent preprocessing. Our training phase involved four machine learning models, namely the Random Forest Classifier and Multilayer Perceptron Classifier. The F1 score, accuracy, precision, and recall were some of our performance metrics. Lastly, we displayed the learning curves for the models that yielded the best results.

As a result of combining the predictions from the four models, the ensemble model, according to our findings, had the highest accuracy, coming in at 97.4%. The Multilayer Perceptron Classifier came in second with an accuracy of 97.4%, trailing only the Random Forest Classifier by 96.5%. Accuracy was 95.6% and 93.0% for the Support Vector Machine and Logistic Regression models, respectively.

Combining the predictions of multiple models has been shown to improve performance, as evidenced by the high accuracy of the ensemble model. The Multilayer Perceptron Classifier had an accuracy of 97.4%, while the Random Forest Classifier had an accuracy of 96.5 percent. In previous studies on how to classify breast cancer, these models have been shown to be effective. The Random Forest Classifier and Multilayer Perceptron Classifier both exhibit increasing performance as the number of training samples rises, according to their learning curves. This shows that adding more training instances may help these models perform even better. According to the ensemble model's confusion matrix, 110 out of 114 samples were correctly identified. It incorrectly categorized three benign samples as malignant and one benign sample as malignant. The model might overlook some cases of breast cancer, as evidenced by the false negative rate of 6.97%. False negatives may result in missed or delayed diagnosis, which can have major repercussions. To enhance the model's performance in these circumstances, more study is required.

Our study showcases the accuracy with which machine learning algorithms identify breast cancer. The ensemble model, which combined the predictions of four models, achieved the highest accuracy of 97.4%. The Random Forest Classifier and Multilayer Perceptron Classifier alone had accuracies of 96.5% and 97.4%, respectively. These models can be used as decision support tools for medical practitioners in diagnosing breast cancer. However, further research is needed to determine how well the models perform on false negatives, which could improve their accuracy and reduce the risk of missed diagnoses. Overall, Our findings demonstrate how machine learning can be used to diagnose breast cancer.

REFERENCES

[1]. Al-Masni, M. A. (2018). Breast cancer classification using machine learning algorithms. International Journal of Computer Applications, 181(44), 1-7.
[2]. American Cancer Society. (2021). Breast Cancer Facts & Figures 2021-2022. Retrieved from https://www.cancer.org/content/dam/cancer-

org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2021-2022.pdf
[3]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).
[4]. Chollet, F. (2018). Deep learning with Python. Manning Publications.
[5]. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
[6]. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Morgan Kaufmann Publishers.
[7]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[8]. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
[9]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
[10]. National Cancer Institute. (2021). SEER Cancer Statistics Review, 1975-2018.
[11]. Li, F., Li, W., Farooq, U., & He, W. (2021). Breast cancer detection using machine learning algorithms: a systematic review and meta-analysis. BMC Medical Informatics and Decision Making, 21(1), 1-24.
[12]. Dunning, A. M., Michailidou, K., Kuchenbaecker, K. B., Thompson, D., French, J. D., Beesley, J., ... & Cross, S. S. (2016). Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nature genetics, 48(4), 374-386.
[13]. American Cancer Society. (2021). Breast Cancer Facts & Figures 2021-2022. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2021-2022.pdf.
[14]. Liu, Y., Wu, X., & Zhou, Z. (2019). Breast cancer diagnosis using deep learning algorithms: a review. ACM Computing Surveys, 52(5), 1-34.
[15]. National Cancer Institute. (2021). SEER Cancer Statistics Review, 1975-2018. Retrieved from https://seer.cancer.gov/csr/1975_2018/.
[16]. Al-Masni, M. A., Al-Absi, H. R., & Mohammed, A. A. (2021). Performance evaluation of machine learning algorithms for breast cancer classification. IEEE Access, 9, 60337-60348.
[17]. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., ... & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer, 144(8), 1941-1953.
[18]. Wang, T., Wang, S., Yang, Y., Kim, E., & Park, J. E. (2020). A review of machine learning algorithms for the prediction of breast cancer recurrence. Journal of Healthcare Engineering, 2020, 1-16.
[19]. Wu, J., Wang, Y., & Gao, L. (2021). A review on machine learning algorithms for breast cancer detection and diagnosis. Frontiers in Oncology, 11, 1-15.
[20]. Wang, Y., Sun, H., Liu, Y., Deng, L., & Zhu, X. (2019). A machine learning-based approach for breast cancer diagnosis and prognosis prediction. Journal of Healthcare Engineering, 2019, 1-8.
[21]. American Cancer Society. (2021). Breast Cancer Facts & Figures 2021-2022. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2021-2022.pdf
[22]. Li, H., & Yu, Y. (2020). Deep learning-based diagnosis of breast cancer using MRI. In International Conference on Neural Information Processing (pp. 712-721). Springer.
[23]. Niknazar, M., Shiri, I., & Abdollahi, H. (2021). Breast cancer detection using machine learning algorithms: a review. International Journal of Cancer Management, 14(2).
[24]. Gandomkar, Z., Brennan, P. C., McEntee, M. F., & Reed, W. M. (2018). Breast cancer screening: the role of radiologists in the age of artificial intelligence. Clinical Radiology, 73(9), 797-804.
[25]. Houssami, N., & Ciatto, S. (2014). Breast cancer screening. In Seminars in oncology (Vol. 41, No. 1, pp. 52-68). WB Saunders.
[26]. Khan, M. A. I., Islam, M. M., Islam, M. R., Rasker, J. J., & Kabir, E. (2020). Machine learning-based breast cancer risk prediction: a systematic review and meta-analysis. Breast Cancer Research and Treatment, 183(1), 1-19.
[27]. Nagi, J., & Mcclymont, D. (2018). A review of machine learning in the diagnosis of breast cancer. Current Oncology, 25(2), e141.
[28]. Rizwan, A., Wang, L., Ur Rehman, N., Liu, Y., Ali, R., Chen, J., ... & Zhao, X. (2021). Breast cancer detection using machine learning algorithms: a systematic review and meta-analysis. Journal of Personalized Medicine, 11(1), 33.
[29]. Samala, R. K., Chan, H. P., Hadjiiski, L., Helvie, M. A., & Wei, J. (2016). Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Physics in Medicine & Biology, 61(24), 8893.
[30]. Yip, S. S., Aerts, H. J., & Lambin, P. (2018). Integration of imaging and clinical data for improved cancer outcomes. The Lancet Oncology, 19(9), e435-e443
[31]. O'Donoghue C, Fox N, McCartan D, Lysaght J, Reynolds JV. Breast cancer: current molecular therapeutic targets and new players in tumour progression. Eur J Cancer. 2017;85:3-14. doi:10.1016/j.ejca.2017.07.003

[32]. Kim E, Kim J. Breast cancer heterogeneity: a focus on epigenetics. Breast Cancer Res. 2016;18(1):1-8. doi:10.1186/s13058-016-0748-z
[33]. Hwang KT, Kim J, Jung J, et al. Identification of genetic factors associated with breast cancer susceptibility using a multilocus genome-wide association study. Breast Cancer Res. 2019;21(1):1-13. doi:10.1186/s13058-019-1219-9
[34]. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-752. doi:10.1038/35021093
[35]. Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736-750. doi:10.1093/jnci/djp082
[36]. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA Cancer J Clin. 2014;64(1):52-62. doi:10.3322/caac.21203
[37]. Ring A, Webb A, Ashley S, Allum WH, Ebbs SR. Pathological complete response rates following breast cancer neoadjuvant chemotherapy: prediction of long-term outcome and selection of patients for breast conservation. Br J Cancer. 2004;91(12):2012-2017. doi:10.1038/sj.bjc.6602235
[38]. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med. 2016;375(8):717-729. doi:10.1056/NEJMoa1602253
[39]. Burstein HJ, Curigliano G, Loibl S, Dubsky P, Gnant M, Poortmans P. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Ann Oncol. 2019;30(10):1541-1557. doi:10.1093/annonc/mdz235
[40]. Wang X, Yu Z, Zhao R, et al. Identification of key genes and pathways in breast cancer using bioinformatics analysis. Front Genet. 2021;12:617121. doi:10.3389/fgene.2021.617121