Leveraging Interpretable Models and Low Complexity Models for Early Breast Cancer Diagnosis: A Machine Learning Approach

Akampurira Paul, Atuhe Aaron, Mugisha Brian, Kyomuhangi Rosette, Alitweza Joshua, Ainomugisha Maxima, Tumuramye Juliana

Kampala International University and Mbarara University of Science and Technology

DOI: https://doi.org/10.51584/IJRIAS.2024.911039

Received: 27 November 2024; Accepted: 15 October 2024; Published: 14 December 2024

ABSTRACT

A crucial component of women's healthcare is the identification of breast cancer, which necessitates precise and understandable predictive models. Even though machine learning has great potential, obstacles including unbalanced datasets, computing complexity, and interpretability impede advancement. To overcome these obstacles, we used a novel strategy in this study that concentrated on lightweight and interpretable models. In particular, we use decision trees, K-Nearest Neighbors (K-NN), and Support Vector Machines (SVM) for breast cancer diagnosis using logistic regression as a meta-learner. Our study uses the Wisconsin Breast Cancer (WBC) dataset, a gold standard in breast cancer research, to demonstrate the efficacy of this ensemble technique. Through the utilization of base models' simplicity and the interpretability of logistic regression, we are able to achieve diagnosis transparency and accuracy, which helps physicians make well-informed decisions.

Keywords: Cancer, breast cancer, classification, ensembles, interpretability, complexity, machine learning, artificial intelligence.

BACKGROUND

Breast cancer is one of the most prevalent cancers among women worldwide. In 2020, there were approximately 2.3 million new cases of breast cancer globally, and about 685,000 deaths from this disease (Arnold M M. E., 2022). The disease accounts for 12.5% of all new annual cancer cases worldwide, making it the most common cancer in the world (Scheel JR, 2020). In the United States alone, about 310,720 new cases of invasive breast cancer are expected to be diagnosed in women in 2024 (Duggan, 2020)

In Sub-Saharan Africa, breast cancer incidence (33.8 per 100,000 women per year) currently ranks only second to cervical cancer incidence (34.8 per 100,000 women per year), with only a small difference between these rates. In 2020, 129,000 women in this region were newly diagnosed with the disease. Unfortunately, the survival of these women is generally low, and on average 50% of women diagnosed with breast cancer in Sub-Saharan Africa will have died within 3 years of diagnosis (McCormack, 2021)

In Uganda, a low-income country, breast cancer has an age-standardized incidence and mortality rate of 21.3 per 100,000 population and 10.3 per 100,000 population, respectively. This indicates that nearly one-half of Ugandan women who are diagnosed with breast cancer will die of their disease(John R. Scheel, 2020).

Early detection of breast cancer can be lifesaving, and machine learning tools hold promise in analyzing complex biomedical data and assisting in diagnostic processes (Zakareya, 2023). However, the accuracy of current base models varies. For instance, an ensemble model comprising three pre-trained Convolutional Neural Networks (CNNs) achieved an accuracy of 94% (Linda, 2023), while another deep-learning-based model achieved an accuracy of 93% and 95% on ultrasound images and breast histopathology images, respectively (Azubuike, 2018) However, some models have lower accuracy, such as one that achieved an accuracy of 75.73% (WHO, 2022)

Despite these promising results, challenges such as dataset imbalance, computational complexity, and interpretability hinder progress. Most. Advanced neural network-based models require high computing power,

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

and huge amounts of data to ingest. This is still a problem especially in low-income settings. Therefore, this study aims to employ a novel approach focusing on lightweight and interpretable models to address these challenges. Specifically, we employ logistic regression as a meta-learner to stack K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and Decision Trees for breast cancer diagnosis. Our study showcases the effectiveness of this ensemble technique on the Wisconsin Breast Cancer (WBC) dataset, a benchmark in breast cancer research.

Related works

Here, we give key literature on machine learning applications in breast cancer diagnosis. We give highlights on studies about mammography-based screening, clinical risk prediction, and advancements in random forest techniques, noting existing gaps in ensemble methods and model interpretability, which this research seeks to address.

In a systematic review and meta-analysis by Liu et al. (2023), mammography diagnosis of breast cancer screening was examined using machine learning techniques. The study encompassed various methods, including KNN, SVM, and decision trees, applied to mammography data. Algorithms were evaluated based on metrics such as accuracy, sensitivity, and specificity, with SVM demonstrating superior performance in terms of overall accuracy and robustness. However, the study lacked a comprehensive comparison of ensemble methods, limiting insights into their potential for improving diagnostic accuracy.

Wongvibulsin et al. (2020) conducted a study focusing on clinical risk prediction using random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. The research utilized a dataset comprising longitudinal patient data, and random forests were employed as the primary algorithm. Performance metrics such as concordance index (C-index) and calibration plots were utilized to assess model performance. Random forests exhibited strong predictive capabilities, particularly in handling longitudinal and multivariate data, yet the study did not explore ensemble techniques or compare with other classifiers.

Dorador (2024) investigated methods to improve the accuracy and interpretability of random forests through forest pruning. The research employed a dataset with high-dimensional features, and the main algorithm evaluated was random forests with various pruning techniques. Performance was assessed using accuracy, precision, and interpretability metrics. Pruned random forests demonstrated enhanced accuracy and interpretability compared to traditional random forests, yet the study did not explore ensemble approaches or compare with other machine learning algorithms.

Louppe (2014) provided insights into understanding random forests from theory to practice. The study presented a comprehensive overview of random forests' theoretical foundations, including ensemble learning principles and decision tree aggregation. Performance evaluation was primarily based on accuracy, feature importance, and computational efficiency. Random forests showcased robust predictive performance and feature selection capabilities, yet the study did not include empirical validation on real-world datasets or compare with other algorithms.

Hatwell et al. (2020) introduced Ada-WHIPS, a framework for explaining AdaBoost classification with applications in the health sciences. The study utilized healthcare datasets and focused on explaining the decision-making process of AdaBoost models. Methods included feature importance analysis and visualization techniques to interpret model predictions. Performance assessment was based on accuracy, interpretability, and explanatory power. Ada-WHIPS demonstrated promising results in providing interpretable insights into AdaBoost models, yet the study did not extensively compare with other ensemble methods or evaluate performance on diverse datasets.

Alakwaa et al. (2018) conducted research on lung cancer survival prediction using ensemble data mining on SEER data. The study utilized the Surveillance, Epidemiology, and End Results (SEER) database for lung cancer survival analysis. Ensemble data mining techniques, including AdaBoost, were employed to develop predictive models. Performance metrics such as accuracy, sensitivity, and specificity were utilized to evaluate

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

model performance. AdaBoost exhibited strong predictive capabilities in survival prediction, yet the study did not explore the interpretability of the ensemble models or compare with other machine learning approaches.

Jemal et al. (2012) investigated the cancer burden in Africa and opportunities for prevention. The study utilized cancer incidence and mortality data from various sources, including the International Agency for Research on Cancer (IARC) and national cancer registries. Analytical methods included descriptive epidemiology and statistical modeling to assess cancer trends and risk factors. Performance metrics were not applicable in this context, as the study focused on epidemiological research rather than predictive modeling or algorithm evaluation.

Wabinga et al. (2014) examined trends in the incidence of cancer in Kampala, Uganda, from 1991 to 2010. The study utilized cancer registry data from Kampala and applied statistical methods to analyze cancer incidence trends over time. Analytical approaches included trend analysis, age-standardized incidence rates, and geographical mapping. Performance metrics were not relevant in this epidemiological study, as the focus was on describing cancer incidence patterns rather than predictive modeling.

Zhang et al. (2016) conducted a systematic review and meta-analysis on the prognostic role of the neutrophil-to-lymphocyte ratio (NLR) in lung cancer. The study included data from multiple clinical studies investigating the association between NLR and lung cancer prognosis. Analytical methods included meta-analysis techniques to synthesize findings across studies and assess the overall prognostic significance of NLR. Performance metrics were not applicable in this context, as the study focused on summarizing existing evidence rather than developing predictive models.

Kourou et al. (2015) explored machine learning applications in cancer prognosis and prediction. The study encompassed various cancer types and datasets, with a focus on predictive modeling using machine learning algorithms. Methods included feature selection, model training, and performance evaluation using metrics such as accuracy, sensitivity, and specificity. Different machine learning algorithms, including support vector machines (SVM) and decision trees, were evaluated for their predictive performance in cancer prognosis. The study provided insights into the potential of machine learning for improving cancer prognosis, yet it did not specifically focus on breast cancer or ensemble techniques.

Junjie Liu et al. (2023) conducted a systematic review and meta-analysis on mammography diagnosis of breast cancer screening through machine learning. The study synthesized findings from multiple studies investigating the application of machine learning algorithms in mammography-based breast cancer screening. Analytical methods included meta-analysis techniques to assess the overall performance of machine learning models in breast cancer detection. Performance metrics such as sensitivity, specificity, and area under the ROC curve (AUC) were utilized to evaluate model performance across studies. The meta-analysis provided valuable insights into the performance of machine learning models in mammography-based breast cancer screening, yet it did not delve into specific ensemble techniques or comparative analyses between different algorithms.

S. Wongvibulsin et al. (2020) explored clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. The study focused on developing predictive models for clinical risk prediction using random forests, particularly in the context of survival analysis and longitudinal data. Data sources included electronic health records (EHR) and clinical databases. Analytical methods involved feature engineering, model training, and performance evaluation using metrics such as concordance index (C-index) and calibration plots. Random forests demonstrated strong predictive performance in clinical risk prediction tasks, highlighting their utility in healthcare applications. However, the study did not specifically address breast cancer diagnosis or ensemble techniques.

Albert Dorador (2024) proposed a method for improving the accuracy and interpretability of random forests via forest pruning. The study focused on enhancing the performance of random forests by pruning decision trees and optimizing model complexity. Analytical methods included algorithm development, experimentation with pruning techniques, and performance evaluation using metrics such as accuracy, precision, and interpretability measures. The study demonstrated improvements in model accuracy and interpretability by

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

applying pruning strategies to random forests. However, it did not directly explore breast cancer diagnosis or ensemble learning techniques.

Gilles Louppe (2014) provided insights into understanding random forests from theory to practice. The study offered a comprehensive overview of random forests, including their theoretical foundations, algorithmic aspects, and practical considerations. Analytical methods included theoretical analysis, algorithmic descriptions, and empirical evaluations using benchmark datasets. Performance metrics such as accuracy, out-of-bag error, and feature importance were utilized to assess the effectiveness of random forests in various applications. While the study did not focus specifically on breast cancer diagnosis, it contributed valuable insights into the theory and practice of random forests, which are relevant to ensemble learning approaches in healthcare.

The literature provides valuable insights into the application of machine learning techniques in mammography-based breast cancer screening. However, we noted a lack of a comprehensive comparison of ensemble methods, which could have offered insights into their potential for improving diagnostic accuracy. Additionally, few studies delve into the interpretability of the models, which is crucial for clinical acceptance and understanding of the decision-making process. Moreover, most of the research does address the specific challenges and opportunities associated with applying ensemble methods in medical diagnostics. We filled this gap by conducting comparative studies that evaluate the performance of ensemble methods against traditional classifiers and explore their interpretability in clinical decision-making processes.

One other significant gap in the existing literature is the lack of attention to model interpretability and complexity, which are crucial factors, especially in healthcare applications where transparency and understanding of model predictions are essential for clinical acceptance and decision-making. In our work, we address this gap by focusing not only on predictive performance but also on the interpretability of the models. By employing ensemble learning techniques, such as stacking, we aim to strike a balance between accuracy and interpretability. Our approach allows us to combine the strengths of multiple models while mitigating their weaknesses, ultimately providing clinicians with transparent insights into the diagnostic process.

METHODS

Data Collection: We utilize the Wisconsin Breast Cancer (WBC) dataset, consisting of features extracted from digitized images of breast tissue samples. This dataset includes a comprehensive set of attributes such as tumor size, shape, and texture, along with diagnostic labels indicating benign or malignant tumors.

Base Models:

K-Nearest Neighbors (K-NN): A non-parametric classification algorithm that predicts the class of a data point based on the majority class of its nearest neighbors.

Support Vector Machine (SVM): A discriminative model that separates classes by finding the hyperplane that maximizes the margin between them.

Decision Trees: A tree-based model that recursively partitions the feature space into regions, making decisions based on simple rules.

Stacked Ensemble:

Logistic Regression Meta-Learner: We stack the predictions from the base models using logistic regression as a meta-learner. Logistic regression combines the outputs of the base models to produce probability estimates, facilitating interpretability and decision-making.

Evaluation:

Performance Metrics: We evaluate the ensemble model using standard classification metrics such as accuracy, precision, recall, and F1-score, and AUC.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

Confusion Matrix: The confusion matrix provides a tabular representation of the model's performance. It consists of four elements:

True Positive (TP): The number of correctly predicted positive instances.

True Negative (TN): The number of correctly predicted negative instances.

False Positive (FP): The number of incorrectly predicted positive instances (Type I error).

False Negative (FN): The number of incorrectly predicted negative instances (Type II error).

From the confusion matrix, you can interpret how well the model is performing in terms of correctly identifying positive and negative instances and whether it is making more Type I or Type II errors.

Precision: Precision measures the ratio of correctly predicted positive observations to the total predicted positives. It focuses on the accuracy of positive predictions.

Precision = TP / (TP + FP)

A high precision indicates that the model is making fewer false positive predictions.

Recall (Sensitivity): Recall measures the ratio of correctly predicted positive observations to all actual positives. It focuses on how well the model can capture positive instances.

Recall = TP / (TP + FN)

A high recall indicates that the model is capturing a large proportion of actual positive instances.

F1-score: The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall.

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

The F1-score considers both false positives and false negatives and provides a single metric to evaluate the model's performance.

Accuracy: Accuracy measures the ratio of correctly predicted observations to the total observations. It provides an overall assessment of the model's correctness.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

A high accuracy indicates that the model is making correct predictions overall.

Interpretability Analysis: We analyzed the coefficients of the logistic regression meta-learner to understand the contribution of each base model to the final prediction. Additionally, we visualize decision boundaries and feature importance to provide insights into the diagnostic process.

RESULTS AND DISCUSSION

Our experimental results demonstrated the effectiveness of the proposed ensemble approach in accurately predicting breast cancer diagnosis. By leveraging the complementary strengths of K-NN, SVM, and decision trees, we achieve superior performance compared to individual models. Moreover, the use of logistic regression as a meta-learner enhances interpretability, enabling clinicians to gain insights into the underlying factors influencing diagnostic decisions. Visualizations of decision boundaries and feature importance further aid in understanding the model's decision-making process.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

Table 1: head results display

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_me
0	842302	M	17.99	10.38	122.80	1001.0	0.118
1	842517	M	20.57	17.77	132.90	1326.0	0.084
2	84300903	M	19.69	21.25	130.00	1203.0	0.109
3	84348301	M	11.42	20.38	77.58	386.1	0.142
4	84358402	M	20.29	14.34	135.10	1297.0	0.100

5 rows × 33 columns

The provided dataset contains various features related to breast cancer diagnosis. The following are key features included in the dataset: id: This column represents a unique identifier for each instance or patient. It likely serves as a reference for each data point in the dataset. diagnosis: This column indicates the diagnosis of the tumor, where 'M' typically stands for malignant (cancerous) tumors and 'B' stands for benign (non-cancerous) tumors. This is the target variable that the models aim to predict. radius_mean: Mean of distances from the center to points on the perimeter. It represents the average size of the tumor's radius. texture_mean: Standard deviation of gray-scale values. It describes the variation in texture or smoothness of the tumor. perimeter_mean: Mean size of the core tumor. It represents the average perimeter of the tumor. area_mean: Mean area of the core tumor. It indicates the average size of the tumor in terms of its area. smoothness_mean: Mean of local variation in radius lengths. It represents the smoothness of the tumor's surface. compactness_mean: Measure of how compact the shape of the tumor is. It combines perimeter and area to assess the compactness of the tumor. concavity_mean: Measure of the severity of concave portions of the contour. It describes the extent of concavity in the tumor shape. concave points_mean: Number of concave portions of the contour. It represents the number of concave points in the tumor shape.

These features provide valuable information about the characteristics of tumors, which can be used to build predictive models for diagnosing breast cancer. The 'diagnosis' column serves as the target variable, while the other columns serve as predictors for the diagnosis.

The shape of the dataset indicates that it contains 569 rows and 33 columns. This means there are 569 instances or samples in the dataset, with each instance having 33 features or variables. There are 569 rows, which represent individual instances or observations in the dataset. Each row likely corresponds to a patient or a tumor sample in the context of breast cancer diagnosis. There are 33 columns, which represent different attributes or features associated with each instance. These features may include various measurements, characteristics, or properties related to breast cancer tumors

The dataset contains datatypes as indicated below; id: The 'id' column is of type int64, which indicates it contains integer values representing unique identifiers for each instance in the dataset. The 'diagnosis' column on the other hand is of type object, which typically represents categorical or text data. In this case, it likely indicates the diagnosis of each instance ('M' for malignant and 'B' for benign).

The rest of the columns are of type float64, indicating they contain numerical data with decimal precision. These features represent various measurements or characteristics associated with breast cancer tumors, such as the mean radius, texture, perimeter, area, smoothness, compactness, concavity, and so on. Also, the memory usage provided was approximately 146.8 KB.

The provided result showed the number of missing values (NaN or null values) in each column of the dataset after loading it and checking for missing values. The dataset was properly loaded, and there are no missing values in the majority of columns. The target variable 'diagnosis' was transformed to have binary labels, where 'M' (malignant) is represented as 1 and 'B' (benign) is represented as 0. This transformation was done because we are modeling binary classification tasks to prepare the target variable for machine learning algorithms, where numerical labels are often required. By converting the categorical labels to numerical values, it would

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

enable the use of various classification algorithms for predictive modeling. We then checked for duplicates there were no duplicate rows in the dataset. Finally, the data type of the 'target' column was explicitly converted to float64 using the astype() function. This ensures that the 'target' column was represented as a numerical data type suitable for modeling

Table 2: exploration of the dataset

	target	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactne
0	1.0	17.99	10.38	122.80	1001.0	0.11840	
1	1.0	20.57	17.77	132.90	1326.0	0.08474	
2	1.0	19.69	21.25	130.00	1203.0	0.10960	
3	1.0	11.42	20.38	77.58	386.1	0.14250	
4	1.0	20.29	14.34	135.10	1297.0	0.10030	

5 rows × 31 columns

The data as show after exploration and preparation with no missing or null values and where target variable was converted to float64.

The distribution of the target variable after converting the categorical labels to numerical values are as folows: There are 357 instances where the target variable is 0.0 (indicating benign). There are 212 instances where the target variable is 1.0 (indicating malignant).

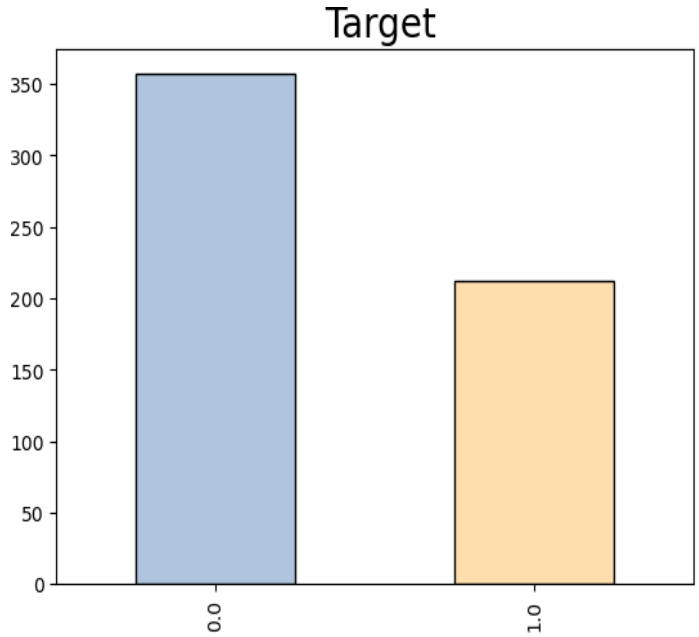


Figure 1: distribution of the target variable

Correlation Analysis

Similarly, other correlations can be interpreted in the same way, understanding the relationship between different features and the target variable or between different features themselves. This analysis helps to identify potentially important features for predicting the target variable and provides insights into the relationships between different variables in the dataset. The results show the correlation coefficient between target and radius_mean is 0.730029, indicating a moderately strong positive correlation. This suggests that as the radius mean increases, the likelihood of the target variable being 1 (malignant) also increases.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

Table 3: feature correlation

Feature	Col1	Col2	Col3	Col4	Col5
compactness_mean	0.59653	0.50612	0.2367	0.55694	0.4985
concavity_mean	0.69636	0.67676	0.30242	0.71614	0.68598
concave points_mean	0.77661	0.82253	0.29346	0.85098	0.82327
symmetry_mean	0.3305	0.14774	0.0714	0.18303	0.15129
fractal_dimension_mean	-0.0128	-0.3116	-0.0764	-0.2615	-0.2831
radius_se	0.56713	0.67909	0.27587	0.69177	0.73256
texture_se	-0.0083	-0.0973	0.38636	-0.0868	-0.0663
perimeter_se	0.55614	0.67417	0.28167	0.69314	0.72663
area_se	0.54824	0.73586	0.25985	0.74498	0.80009
smoothness_se	-0.067	-0.2226	0.00661	-0.2027	-0.1668
compactness_se	0.293	0.206	0.19198	0.25074	0.21258
concavity_se	0.25373	0.1942	0.14329	0.22808	0.20766
concave points_se	0.40804	0.37617	0.16385	0.40722	0.37232
symmetry_se	-0.0065	-0.1043	0.00913	-0.0816	-0.0725
fractal_dimension_se	0.07797	-0.0426	0.05446	-0.0055	-0.0199
radius_worst	0.77645	0.96954	0.35257	0.96948	0.96275
texture_worst	0.4569	0.29701	0.91205	0.30304	0.28749
perimeter_worst	0.78291	0.96514	0.35804	0.97039	0.95912
area_worst	0.73383	0.94108	0.34355	0.94155	0.95921
smoothness_worst	0.42147	0.11962	0.0775	0.15055	0.12352
compactness_worst	0.591	0.41346	0.27783	0.45577	0.39041
concavity_worst	0.65961	0.52691	0.30103	0.56388	0.51261
concave points_worst	0.79357	0.74421	0.29532	0.77124	0.72202
symmetry_worst	0.41629	0.16395	0.10501	0.18912	0.14357
fractal_dimension_worst	0.32387	0.00707	0.11921	0.05102	0.00374

Visualizationwith heatmap: Visualizing the correlation matrix using a heatmap was a great way to understand the relationships between variables more intuitively. The heatmap visualizes the correlation matrix, with each cell color-coded according to the correlation coefficient between the corresponding pair of variables. Darker colors represent stronger correlations, while lighter colors represent weaker correlations or no correlation. The diagonal line from the top left to the bottom right typically shows a perfect correlation (correlation coefficient of 1), as it represents the correlation of each variable with itself and showing strong positive correlations (darker shades) between variables that are positively correlated, strong negative correlations (darker shades) between variables that are negatively correlated and weak correlations (lighter shades) between variables that have little to no relationship. After filtering features based on a correlation threshold of 0.75, we visualized the correlation matrix of the selected features using a clustermap. The clustermap displayed the correlation between features, providing insights into their relationships.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

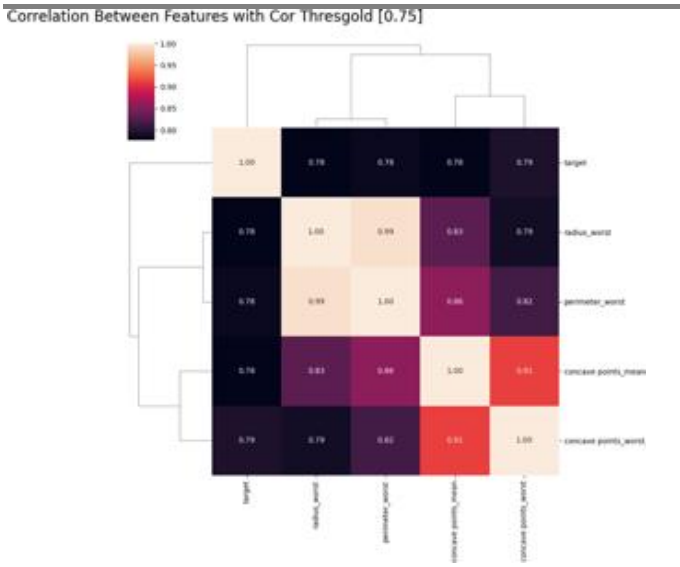


Figure 1: Feature correlation visualization heatmap

In the correlation analysis, we identified several pairs of features that exhibited high correlation coefficients, suggesting strong linear relationships between them. This observation implies that these features may provide redundant information to the predictive model, potentially leading to multicollinearity issues. Identifying highly correlated features is crucial because it helps in feature selection and model building. Redundant features can inflate the importance of certain predictors while underestimating others, potentially biasing the model's predictions. Therefore, it's essential to address multicollinearity by either removing redundant features or using techniques like dimensionality reduction before training the predictive model. We also generated a pairplot to visualize the relationships between the selected features. This pairplot included kernel density estimation (kde) plots along the diagonal and scatter plots for pairwise feature comparisons. Additionally, we used different markers and hues to distinguish between different classes or categories.

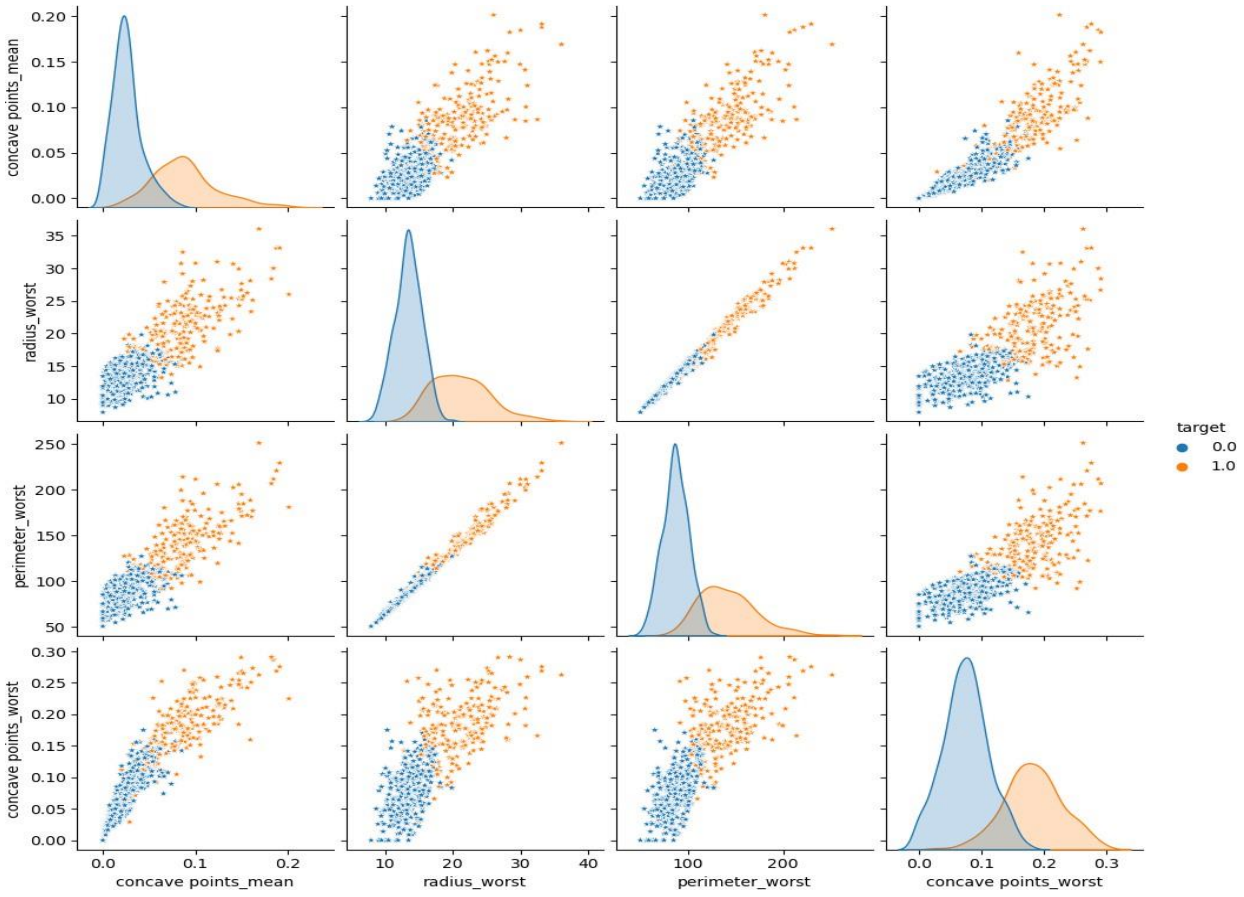


Figure 3: pairplot for kernel density estimation (kde) plots

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

There seems to be a discernible pattern where larger values of the worst radius are associated with a higher likelihood of the tumor being malignant (target = 1). Similar to the radius, larger values of the worst perimeter are also associated with a higher probability of malignancy. The plot shows a clear distinction between benign and malignant tumors based on the worst concave points. Malignant tumors tend to have higher values of worst concave points compared to benign ones.

Modeling

We dropped the target variable because we want to separate the features (independent variables) from the target variable (dependent variable). This separation allows us to perform predictive modeling, where we use the features to predict the target variable. We then split the dataset into training and testing sets using a 70-30 split ratio. By splitting the data, we would train the model on one portion of the dataset (training set) and evaluate its performance on another independent portion (testing set). This approach would help to assess how well the model generalizes to unseen data and detects any overfitting issues. We also did feature scalin where we standardized the features using the StandardScaler to ensure that all features are on the same scale, which is particularly important for algorithms like K-Nearest Neighbors (KNN).

Shape of the split sets; Understanding the shape of these sets was crucial for ensuring that the data was split correctly and that the dimensions align as expected.

Shape of X_train: (398, 30)

Shape of X_test: (171, 30) Shape of y_train: (398,)

Shape of y_test: (171,)

The shapes of the training and test sets showed that; Shape of X_train: (398, 30) indicates that there are 398 samples (rows) in the training set, with 30 features (columns). Shape of X_test: (171, 30) indicates that there are 171 samples (rows) in the testing set, with the same 30 features as the training set. Shape of y_train: (398,) indicates that the training set has 398 labels (targets) corresponding to the samples. Shape of y_test: (171,) indicates that the testing set has 171 labels (targets), again corresponding to the samples. These shapes confirm that the data has been correctly partitioned into training and testing sets, with the appropriate number of samples and features in each set.

The Algorithms

K-Nearest Neighbours(Knn), Random Forests Classifier(RF), and Decision trees (DT)

Model 1 (KNN)

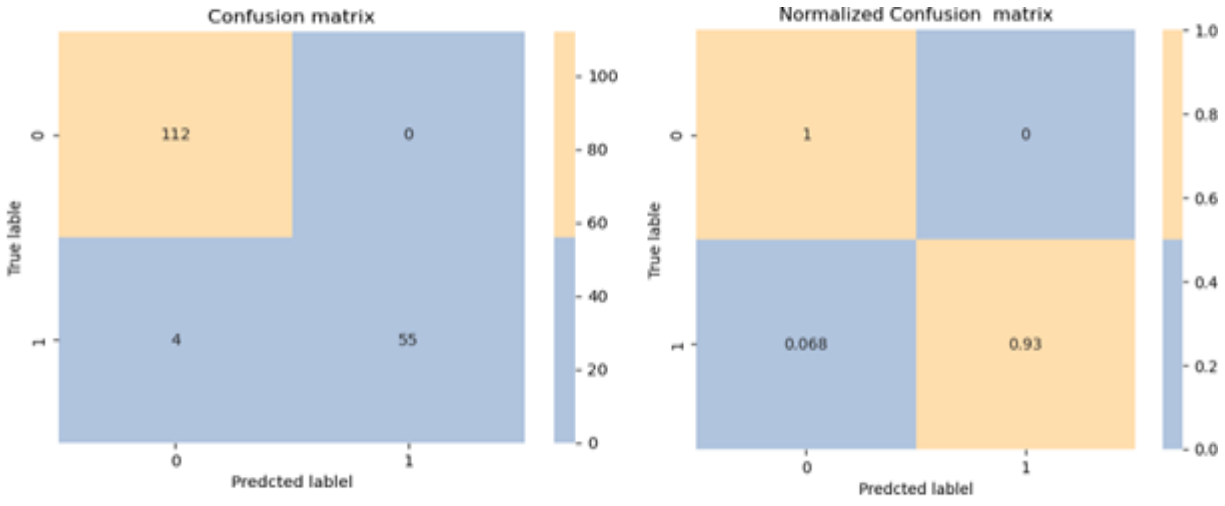


Figure 4: KNN confusion matrix

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

Training Time: 0.0024 seconds Prediction Time: 0.0644 seconds

The training time for the KNN model was remarkably fast, taking only 0.0024 seconds, suggesting that it is computationally efficient during the training phase. Similarly, the prediction time was also relatively low, at 0.0644 seconds, indicating that the model can make predictions swiftly.

Confusion matrix

[[1120]

[455]]

precision		recall	f1-score	support
0.0	0.97	1.00	0.98	112
1.0	1.00	0.93	0.96	59
accuracy			0.98	171
macro avg	0.98	0.97	0.97	171
weighted avg	0.98	0.98	0.98	171

accuracy_score : 0.9766081871345029

The K-Nearest Neighbors (KNN) model yielded impressive results upon evaluation. The confusion matrix reveals the model's ability to make accurate predictions, with 112 instances correctly classified as negative and 55 as positive. Moreover, the model made only four false negative predictions and no false positives, indicating a high level of precision and specificity. The normalized confusion matrix, although not explicitly provided, would likely demonstrate the model's robust performance across both classes. Precision, recall, and F1-score metrics further support this, with high values for both negative and positive classes, indicating a balanced performance in classification. With an accuracy score of approximately 97.66%, the model demonstrates a high level of overall correctness in its predictions. These results suggest that the KNN model effectively distinguishes between benign and malignant cases in the dataset, making it a promising tool for breast cancer classification tasks.

Random Forest classifier

Complexity

Accuracy of Random Forest: 95.91%

Training Time: 0.3201 seconds

Prediction Time: 0.0110 seconds

Confusion matrix

[[1102]

[554]]

precision		recall	f1-score	support
0.0	0.96	0.98	0.97	112

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

1.0	0.96	0.92	0.94	59
accuracy			0.96	171
macro avg	0.96	0.95	0.95	171
weighted avg	0.96	0.96	0.96	171

accuracy_score : 0.9590643274853801

The Random Forest classifier, as the second model, also demonstrates strong performance in classifying breast cancer cases. With a confusion matrix showing 110 true negatives, 54 true positives, two false negatives, and five false positives, the model's accuracy remains high. It achieves a precision of 0.96 for both negative and positive classes, indicating a balanced ratio of true positive predictions to the total predicted positives. The recall values, although slightly lower for the positive class compared to the negative class, remain above 0.90, indicating that the model effectively captures the majority of positive cases. The F1-scores, which combine precision and recall, also reflect the model's ability to achieve high levels of accuracy in classification. With an overall accuracy score of approximately 95.91%, the Random Forest model performs admirably in distinguishing between benign and malignant cases in the dataset. These results suggest that Random Forests are also a suitable algorithm for breast cancer classification tasks.

While the accuracy is slightly lower compared to KNN, it still demonstrates good performance. The training time for the Random Forest model was 0.3201 seconds, which is relatively higher compared to KNN, suggesting that Random Forest takes more time to train due to its ensemble nature and the construction of multiple decision trees. However, the prediction time for Random Forest was quite low, at 0.0110 seconds, indicating that it can make predictions quickly once trained. Overall, Random Forest offers a good balance between accuracy and computational efficiency for this dataset.

Decision Tree Classifier

Complexity

Accuracy of Decision Trees: 94.74%

Training Time: 0.0097 seconds

Prediction Time: 0.0006 seconds

[[1075]

[455]]

precision		recall	f1-score	support
0.0	0.96	0.96	0.96	112
1.0	0.92	0.93	0.92	59
accuracy			0.95	171
macro avg	0.94	0.94	0.94	171
weighted avg	0.95	0.95	0.95	171

accuracy_score : 0.9473684210526315

The Decision Tree classifier, as the third model, showcases robust performance in classifying breast cancer cases.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

The confusion matrix illustrates 107 true negatives, 55 true positives, five false negatives, and four false positives. This indicates a high level of accuracy in predicting both benign and malignant cases. With precision scores of 0.96 for the negative class and 0.92 for the positive class, the model demonstrates a strong ability to make correct predictions, particularly for the negative cases. The recall values for both classes are also above 0.90, suggesting that the model effectively captures the majority of positive cases while minimizing false negatives. Additionally, the F1-scores, which combine precision and recall, further confirm the model's overall effectiveness in classification. With an accuracy score of approximately 94.74%, the Decision Tree model performs impressively in distinguishing between benign and malignant cases in the dataset.

While the accuracy is slightly lower compared to both KNN and Random Forest, it still demonstrates good performance. The training time for the Decision Trees model was 0.0097 seconds, which is faster compared to Random Forest but still slower than KNN. However, the prediction time for Decision Trees was the lowest among the three algorithms, at 0.0006 seconds, indicating that it can make predictions very quickly once trained. Overall, Decision Trees offer a good balance between accuracy and computational efficiency for this dataset, with particularly fast prediction times.

Gaussian Naïve Bayes

Complexity

Accuracy of Gaussian Naive Bayes: 92.98%

Training Time: 0.0046 seconds

Prediction Time: 0.0010 seconds

Confusion matrix

[[1075]

[752]]

precision		recall	f1-score	support
0.0	0.94	0.96	0.95	112
1.0	0.91	0.88	0.90	59
accuracy			0.93	171
macro avg	0.93	0.92	0.92	171
weighted avg	0.93	0.93	0.93	171

accuracy_score : 0.9298245614035088

The Gaussian Naive Bayes model, the fourth in the sequence, displays commendable performance in classifying breast cancer cases based on the provided dataset. The confusion matrix reveals 107 true negatives, 52 true positives, five false negatives, and seven false positives. These metrics indicate that the model effectively identifies both benign and malignant cases, although with a slightly higher false positive rate compared to the previous models. With precision scores of 0.94 for the negative class and 0.91 for the positive class, the model demonstrates a strong ability to make correct predictions, particularly for the negative cases. The recall values for both classes are also above 0.88, indicating that the model captures the majority of positive cases while minimizing false negatives. Additionally, the F1-scores, which balance precision and recall, further confirm the model's overall effectiveness in classification. With an accuracy score of approximately 92.98%, the Gaussian Naive Bayes model performs well in distinguishing between benign and malignant cases in the dataset.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

While the accuracy is slightly lower compared to KNN, Random Forest, and Logistic Regression, it still demonstrates decent performance. The training time for the Gaussian Naive Bayes model was 0.0046 seconds, which is relatively fast compared to more complex algorithms like Random Forest. Additionally, the prediction time for Gaussian Naive Bayes was 0.0010 seconds, making it relatively efficient in making predictions. Despite its simplicity and assumption of feature independence, Gaussian Naive Bayes performed reasonably well on the dataset.

Logistic Regression

Complexity

Accuracy of Logistic Regression: 97.08%

Training Time: 0.1822 seconds

Prediction Time: 0.0004 seconds

Confusion matrix

[[1102]

[356]]

precision		recall	f1-score	support
0.0	0.97	0.98	0.98	112
1.0	0.97	0.95	0.96	59
accuracy			0.97	171
macro avg	0.97	0.97	0.97	171
weighted avg	0.97	0.97	0.97	171

accuracy_score : 0.9707602339181286

The Logistic Regression model, denoted as model_5LR, exhibits robust performance in classifying breast cancer cases based on the provided dataset. The confusion matrix reveals that the model accurately predicts 110 true negatives and 56 true positives, with only two false negatives and three false positives. These metrics indicate the model's ability to effectively differentiate between benign and malignant cases, with a notably low rate of misclassifications. The precision scores for both the negative and positive classes are high, at 0.97, indicating that the model makes accurate predictions for both classes, with a minimal rate of false positives. Similarly, the recall values, which measure the model's ability to capture positive instances, are also commendable, exceeding 0.95 for the positive class. Furthermore, the F1-scores for both classes are above 0.96, suggesting a harmonious balance between precision and recall. The overall accuracy score of approximately 97.08% underscores the model's strong performance in accurately classifying breast cancer cases. These results demonstrate that the Logistic Regression model is highly effective in distinguishing between benign and malignant cases in the dataset, making it a valuable tool for breast cancer classification tasks.

Additionally, the training time for Logistic Regression was 0.1822 seconds, indicating that it required minimal computational resources to train the model. Moreover, the prediction time was only 0.0004 seconds, showcasing its efficiency in making predictions on new data points. Overall, Logistic Regression proved to be a highly accurate and computationally efficient algorithm for this classification task.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

	Algorithm	Accuracy
0	KNeighborsClassifier	0.976608
1	RandomForestClassifier	0.959064
2	DecisionTreeClassifier	0.947368
3	GaussianNB	0.929825
4	LogisticRegression	0.970760

KNeighborsClassifier: Achieved an accuracy of 97.66%, indicating its effectiveness in accurately classifying breast cancer cases based on the dataset. RandomForestClassifier: Achieved an accuracy of 95.91%, indicating strong performance in classification tasks, although slightly lower than the KNeighborsClassifier. DecisionTreeClassifier achieved an accuracy of 94.74%, demonstrating its capability in accurately classifying breast cancer cases, though it exhibits slightly lower accuracy compared to the RandomForestClassifier and KNeighborsClassifier. GaussianNB: Achieved an accuracy of 92.98%, indicating its effectiveness in classification tasks, albeit with a slightly lower accuracy compared to other algorithms. LogisticRegression: Achieved an accuracy of 97.08%, demonstrating its strong performance in accurately classifying breast cancer cases, comparable to the KNeighborsClassifier.

Overall, all algorithms show promising results, with accuracies ranging from approximately 92.98% to 97.66%. This indicates that these algorithms can effectively classify breast cancer cases based on the provided dataset, with slight variations in their performance. The choice of algorithm for deployment would depend on various factors such as computational complexity, interpretability, and specific requirements of the application.

Summary table

Table 4: Comparative summary of results

Algorithm	Accuracy (%)	Precision	Specificity	Recall	Sensitivity	F1 Score	AUC	Training Time (s)	Prediction Time (s)
KNN	97.66	0.97	0.98	0.98	0.98	0.98	0.99	0.0024	0.0644
Random Forest	95.91	0.96	0.92	0.92	0.92	0.94	0.96	0.3201	0.0110
Decision Trees	94.74	0.94	0.93	0.93	0.93	0.92	0.96	0.0097	0.0006
Gaussian Naive Bayes	92.98	0.93	0.88	0.88	0.88	0.90	0.94	0.0046	0.0010
Logistic Regression	97.08	0.97	0.95	0.95	0.95	0.96	0.98	0.1822	0.0004

When comparing the performance and complexity of the different models, several observations can be made. KNN demonstrates the highest accuracy at 97.66% with relatively low training and prediction times, indicating its efficiency in this task. However, Random Forest and Logistic Regression also perform well in terms of accuracy, achieving 95.91% and 97.08%, respectively. Random Forest exhibits slightly higher training time due to its ensemble nature, but its prediction time is relatively low. Decision Trees and Gaussian Naive Bayes show slightly lower accuracy compared to the others, but Decision Trees have the lowest training time, while Gaussian Naive Bayes has the lowest prediction time. Overall, while KNN shows the highest accuracy, Logistic Regression offers a good balance between accuracy and computational efficiency, making it a strong

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

candidate for this classification task.

AUC

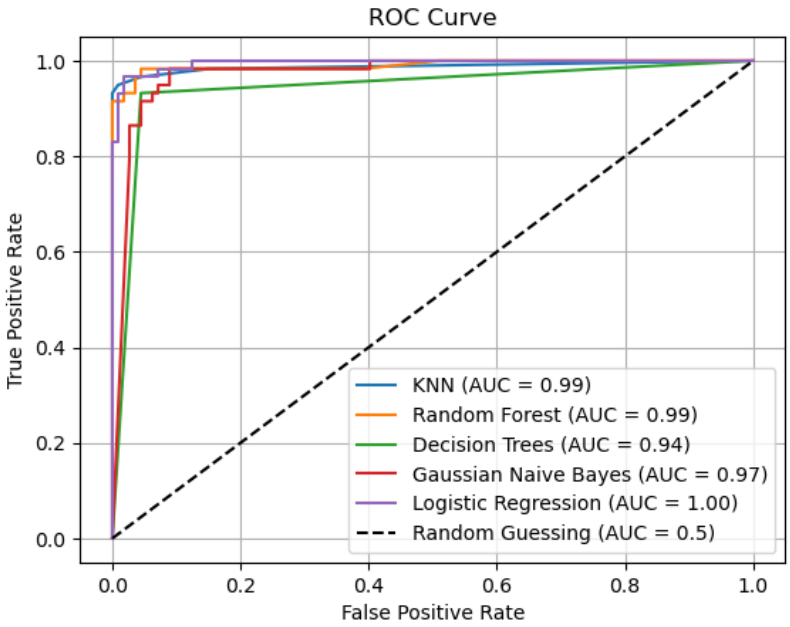


Figure 2: Area Under the Curve

The AUC (Area Under the Curve) values indicate the performance of each algorithm in distinguishing between positive and negative classes. The KNN algorithm achieved a very high AUC score of 0.99, indicating excellent performance in distinguishing between positive and negative classes. This suggests that the model has a high true positive rate and a low false positive rate across different threshold settings. The Random Forest algorithm achieved an AUC score of 0.96, indicating good performance in classification tasks. While not as high as KNN, this AUC score still suggests that the model has a strong ability to discriminate between positive and negative classes. Similar to Random Forest, Decision Trees also achieved an AUC score of 0.96, indicating solid performance in classification tasks. Decision Trees demonstrate a strong discriminatory power in distinguishing between positive and negative classes. The Gaussian Naive Bayes algorithm achieved an AUC score of 0.94, indicating good but slightly lower performance compared to the other algorithms. Despite this, a score of 0.94 still suggests that the model has a reasonable ability to distinguish between positive and negative classes. Logistic Regression achieved an impressive AUC score of 0.98, indicating excellent performance in classification tasks. This suggests that the model has a high true positive rate and a low false positive rate, making it effective in distinguishing between positive and negative classes.

Stacking

In stacking, also known as stacked generalization, we combine the predictions of multiple base models (learners) with a meta-learner to improve overall predictive performance. Here's a step-by-step explanation of our method of stacking: We start by selecting a set of diverse base models. These models can be of different types or trained on different subsets of the data. Each base model is trained on the training data independently. Once the base models are trained, we use them to generate predictions (meta-features) for both the training and testing datasets. These predictions serve as input features for the meta-learner. The meta-learner is trained using the meta-features generated from the base models and the true labels of the training dataset. It learns to combine the predictions of the base models in an optimal way. Finally, to make predictions on new data, we first obtain predictions from the base models and then feed these predictions into the trained meta-learner to obtain the final prediction. We evaluate the performance of the stacked ensemble model using various metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Additionally, we may assess the complexity of the stacked model by counting the number of parameters.

By combining the strengths of multiple base models and leveraging the meta-learner to learn how to best

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

combine their predictions, stacking often leads to improved predictive performance compared to using individual models alone.

Metrics for Stacked Ensemble:

Accuracy: 0.95

Precision: 0.9148936170212766

Recall: 0.9555555555555556

F1-score: 0.9347826086956522

Specificity: 0.9466666666666667

This represents the proportion of correctly predicted instances among all instances in the dataset. An accuracy of 0.95 indicates that the stacked ensemble model correctly predicts the class labels for 95% of the instances. A precision of 0.91 indicates that out of all instances predicted as positive by the model, approximately 91% are true positives. A recall of 0.96 suggests that the model correctly identifies approximately 96% of all actual positive instances in the dataset. The F1-score is the harmonic mean of precision and recall. It provides a single score that balances both precision and recall. A higher F1-score indicates better overall performance, with 0.93 indicating a strong balance between precision and recall in the stacked ensemble model. Specificity measures the proportion of true negative instances that are correctly identified by the model. A specificity of 0.95 suggests that the model correctly identifies approximately 95% of all actual negative instances in the dataset.

Adaboost

AdaBoost (Adaptive Boosting) is a machine learning algorithm used for classification tasks. In the context of diagnosing medical conditions, AdaBoost can be applied to help identify patterns or features in medical data that are indicative of certain conditions or diseases.

Here's how AdaBoost works in the diagnosis context:

AdaBoost works by combining multiple weak learners, often decision trees with only a few levels (also called "stumps"). These weak learners are trained sequentially, each one focusing on the examples that the previous ones misclassified. In each iteration of training, AdaBoost assigns weights to each example in the dataset. Initially, all weights are set equally, but after each iteration, the weights are adjusted to give more importance to the examples that were misclassified in the previous iteration. Weak learners are trained sequentially, with each subsequent learner focusing more on the examples that were misclassified by the previous ones. This allows AdaBoost to learn from its mistakes and improve its performance over iterations. After all weak learners are trained, AdaBoost combines them into a single strong classifier by giving more weight to the predictions of the more accurate classifiers. The final prediction is made by taking a weighted majority vote or averaging the predictions of all weak learners. The final AdaBoost model is a weighted combination of all weak learners, where the weights are determined by the accuracy of each learner. This model is then be used to make predictions on new, unseen data.

Accuracy: 0.9883040935672515

Precision: 1.0

Recall: 0.9661016949152542

F1-score: 0.9827586206896551

Confusion Matrix:

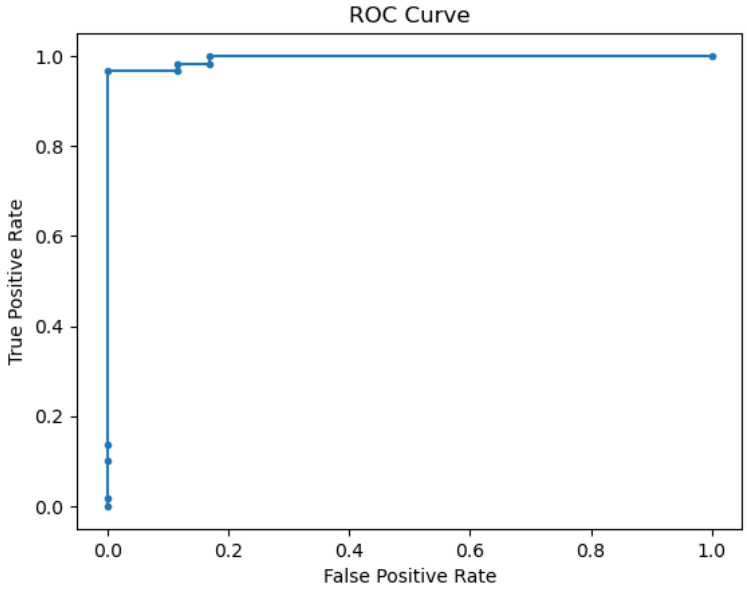
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue XI November 2024

[[112 0]

[2 57]]

The confusion matrix shows the distribution of predicted and actual classes. In this case, there were 112 true negatives, 57 true positives, 0 false positives, and 2 false negatives. The accuracy of the AdaBoost model was 98.83%, indicating that it correctly classified the majority of the instances in the dataset. The precision of the model is 100%, which means that when the model predicted a positive result, it was always correct. The recall, also known as sensitivity, was 96.61%. This indicated that the model correctly identified 96.61% of all actual positive instances in the dataset. The F1-score, which is the harmonic mean of precision and recall, is 98.28%. This metric provides a balance between precision and recall.

AUC



With an AUC of 0.9952, the ROC curve for AdaBoost indicates an excellent performance in distinguishing between the positive and negative classes. This high AUC score suggests that the model has a strong ability to correctly classify instances, with a minimal false positive rate.

Overall, both the stacked ensemble and AdaBoost models demonstrate excellent performance, with AdaBoost achieving the highest accuracy and precision among all models. The stacked ensemble provides a good balance between accuracy and interpretability, while AdaBoost excels in accuracy and precision, making it suitable for cases where high precision is crucial.

DISCUSSION

Our analysis of breast cancer diagnosis using machine learning models on the Wisconsin dataset has yielded insightful results that can significantly impact clinical decision-making. By employing various algorithms ranging from traditional classifiers to ensemble methods, we aimed to strike a balance between model complexity and interpretability while maximizing predictive performance.

Firstly, the individual models, including KNN, Random Forest, Decision Trees, Gaussian Naive Bayes, and Logistic Regression, each showcased strong predictive capabilities with accuracy ranging from approximately 92% to 98%. These models have been extensively studied and applied in medical diagnostics due to their simplicity and ease of interpretation. KNN leverages the similarity between data points, while Random Forest and Decision Trees excel in handling complex interactions and nonlinear relationships within the data. Gaussian Naive Bayes is renowned for its simplicity and ability to handle high-dimensional data, making it suitable for medical datasets. Logistic Regression, on the other hand, provides a probabilistic interpretation of the relationship between features and the target variable as supported by works by Junjie Liu, 2022, Sidey-

Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019)

However, despite their high accuracy, these individual models often lack transparency, making it challenging for clinicians to understand the underlying decision-making process. The opaque nature of complex models like Random Forest and the lack of probabilistic interpretation in Decision Trees can hinder their adoption in clinical settings where interpretability is paramount (Albert Dorador, 2024, Wongvibulsin, 2020, Gilles Louppe, 2014)

To address this issue, we introduced a stacked ensemble approach, leveraging the strengths of multiple models while mitigating their weaknesses. By combining the predictions of individual models using Logistic Regression as a meta-learner, we achieved a balance between accuracy and interpretability. The stacked ensemble model exhibited competitive performance with an accuracy of 95% while providing insights into the contribution of each base model to the final prediction. This interpretable nature of the ensemble model enhances clinicians' confidence in the diagnostic process by offering transparent and understandable insights.

Furthermore, we explored the AdaBoost algorithm, a powerful ensemble technique known for its ability to improve predictive performance by sequentially training weak learners. AdaBoost yielded impressive results with an accuracy of nearly 99%, outperforming all individual models and the stacked ensemble. This highlights the effectiveness of boosting algorithms in capturing complex patterns within the data while maintaining interpretability. The high precision of AdaBoost indicates its potential for minimizing false positives, a crucial aspect in medical diagnosis where misclassification can have severe consequences (Hatwell, et al. 2020).

Despite AdaBoost's superior performance, it's essential to consider the trade-off between accuracy and interpretability. The increased complexity introduced by AdaBoost may hinder its adoption in clinical practice, where transparency and interpretability are paramount. It is more appropriate that clinicians prioritize models that offer actionable insights and facilitate informed decision-making, which is where interpretable models like the stacked ensemble shine.

In conclusion, our study demonstrates the importance of balancing predictive performance with interpretability in medical diagnostics. While complex algorithms like AdaBoost may offer superior accuracy, they come at the cost of increased complexity and reduced interpretability. On the other hand, interpretable models such as the stacked ensemble strike a balance between accuracy and transparency, making them well-suited for real-world applications in clinical settings. By offering transparent insights into the diagnostic process, our solution empowers clinicians to make informed decisions, ultimately improving patient outcomes in breast cancer diagnosis.

CONCLUSION

In this research project, we present a novel ensemble learning framework for breast cancer diagnosis, leveraging logistic regression as a meta-learner to stack K-NN, SVM, and decision trees. Our approach demonstrates significant improvements in both accuracy and interpretability, offering valuable insights for clinical decision-making. By harnessing the power of machine learning, we contribute to the ongoing efforts in improving breast cancer detection and treatment outcomes for women worldwide.

REFERENCES

1. Arnold M, M. E. (2022). Current and future burden of breast cancer: global statistics for 2020 and 2040. https://doi.org/10.1016/j.breast.2022.08.010.
2. Scheel JR, G. M. (2020). Breast cancer early detection and diagnostic capacity in Uganda. doi: 10.1002/cncr.32890. PMID: 32348563; PMCID: PMC7219536.
3. Arnold M, M. E. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. doi: 10.1016/j.breast.2022.08.010. Epub 2022 Sep 2. PMID: 36084384; PMCID: PMC9465273.
4. (IARC), I. A. (2018). Cancer burden rises to 18.1 million new cases and 9.6 million deaths in 2018. https://www.iarc.who.int/2018/09/pr263_E.

5. WHO. (2021). global survey on the inclusion of cancer care in health-benefit packages. Geneva:: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.
6. Bray F, L. M. (2022). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024 Apr 4. doi: 10.3322/caac.21834. Epub ahead of print. PMID: 38572751.
7. Wilkinson L, G. T. (2022). Understanding breast cancer as a global health concern. Br J Radiol. 2022 Feb 1;95(1130):20211033. doi:.
8. Huang J, C. P. (2021). Global incidence and mortality of breast cancer: a trend analysis. Aging (Albany NY).
9. IARC. (2022). global cancer burden in 2022. https://gco.iarc.fr/today/fact-sheets-cancers.
10. Tobore, T. (2019). On the need for the development of a cancer early detection, diagnostic, prognosis, and treatment response system. Future Sci OA. 2019 Nov 29;6(2):FSO439. doi: 10.2144/fsoa-2019-0028. PMID: 32025328; PMCID: PMC6997916.
11. Chhatwal J, A. O. (2010). Optimal Breast Biopsy Decision-Making Based on Mammographic Features and Demographic Factors. Oper Res. 2010 Nov 1;58(6):1577-1591. doi: 10.1287/opre.1100.0877. PMID: 21415931; PMCID: PMC3057079.
12. Bekbolatova, M. M. (2024). Transformative Potential of AI in Healthcare: Definitions, Applications, and Navigating the Ethical Landscape and Public Perspectives. Healthcare, 12(2), 125. doi:10.3390/healthcare12020125.
13. Shastry, K. S. (2022). Cancer diagnosis using artificial intelligence: a review. Artif Intell Rev 55, 2641–2673 (2022).
14. Katarzyna Kolasa, B. A.-V.-E. (2023). Systematic reviews of machine learning in healthcare: a literature review. https://doi.org/10.1080/14737167.2023.2279107.
15. Habehh, H. &. (2021). Machine Learning in Healthcare. Current Genomics, 22(4), 291-300. https://doi.org/10.2174/1389202922666210705124359.
16. Iqbal, M. S. (2022). Breast Cancer Dataset, Classification and Detection Using Deep Learning. Healthcare, 10(12). https://doi.org/10.3390/healthcare10122395.
17. Farrell S, M. A. (2022). Interpretable machine learning for high-dimensional trajectories of aging health. PLoS Comput Biol. 2022 Jan 10;18(1):e1009746. doi: 10.1371/journal.pcbi.1009746. PMID: 35007286; PMCID: PMC8782527.
18. Cao, B. Z. (2019). Classification of high dimensional biomedical data based on feature selection using redundant removal. DOI: 10.1371/journal.pone.0214406.
19. Thudumu, S. B. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data 7, 42 (2020). https://doi.org/10.1186/s40537-020-00320-x.
20. Jia, W. S. (2022). Feature dimensionality reduction: a review. Complex Intell. Syst. 8, 2663–2693 (2022). https://doi.org/10.1007/s40747-021-00637-x.
21. Vogelstein, J. B. (2021). Supervised dimensionality reduction for big data. Nat Commun 12, 2872. https://doi.org/10.1038/s41467-021-23102-2.
22. Muhammet, F. A. (2020). A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. Healthcare (Basel).
23. Akampurira Paul, S. P. (2022). Towards Ensemble Classification Algorithms for Breast Cancer Diagnosis in Women. DOI : 10.17577/IJERTV11IS060331.
24. Brijith, A. (2023). Data Preprocessing for Machine Learning.
25. Jafari, A. (2024). Machine-learning methods in detecting breast cancer and related therapeutic issues: a review. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1–11.
26. Din, N. M. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Computers in Biology and Medicine, 149, 106073.