The Seybold Report

ISSN 1533-9211

DOI: 10.5281/zenodo.14598767

A COMPARATIVE EVALUATION OF MACHINE AND DEEP LEARNING ALGORITHMS FOR BREAST CANCER DIAGNOSIS

VAIBHAV JINDAL*

Department of Computer Science, Vit Bhopal University, Madhya Pradesh, India.
*Corresponding Author Email: jindalvaibhav63@gmail.com

Abstract

Breast cancer still ranks amongst the leading causes of death in woman worldwide and early diagnosis is important to increase the chance of survival and effective treatment. Current milestones in ML & DL have laid down strong platforms for medical diagnosis particularly for cancer diagnosis. The objective of this work is to examine and benchmark different kinds of ML and DL models for early detection of breast cancer using WBCD. Normalization of the dataset is performed, as well as an imputation of missing values, feature selection, correlation analysis with the use of heatmap visualization. A number of algorithms are used, namely Logistic Regression, Support Vector Machines (SVM) kernel: linear, radial basis function, and polynomial, K-Nearest Neighbours (KNN), Naive Bayes, Decision Trees, Random Forest, AdaBoost, XGBoost, CatBoost, Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN). Furthermore, the blended models containing KNN and SVM with Random Forest are reserved to foster prediction accuracy. To perform the hyperparameter optimization, use of Grid Search is made. For model measurement, commonly used indicators include accuracy, precision, recall, specificity, sensitivity, and F1 score are adopted. The highest accuracy of 98.57% is recorded when the model is trained on 90% of the total dataset. The outcomes also suggest that deeper learning and ensemble methods are superior to conventional recipes of applying machine learning algorithms in early diagnoses and treatment of breast cancer.

Keywords: Artificial Intelligence, Machine and Deep Learning Algorithms, Data Collection and Analysis, Evaluation, Confusion Matrix, Visualization.

INTRODUCTION

Breast cancer detection and treatment pose major hurdles in oncology. The World Health Organization (WHO) reports that breast cancer tops the list of cancers diagnosed in women. In 2020, an estimated 2.3 million new cases emerged resulting in 685,000 deaths worldwide (World Health Organization, 2021).

This represents about 11.7% of all cancer cases making breast cancer a key factor in cancer deaths [1]. To boost survival rates, doctors need to spot and diagnose breast cancer. When caught in its early stages, the five-year survival rate for breast cancer can go beyond 90% (American Cancer Society, 2021). But standard diagnostic methods like mammography and biopsy have their limits. These include differences in how results are interpreted, reliance on the radiologist's skill, and wrong positive or negative outcomes.

Research shows that mammography has a sensitivity of 85%, but this number can drop a lot in women with dense breast tissue [2]. Machine learning (ML) and deep learning (DL) have a revolutionary impact on healthcare. They provide strong automated, and scalable answers to detect and classify diseases. These algorithms can improve traditional ways to diagnose by using big data sets and spotting complex patterns.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

Research shows that DL models can diagnose with over 90% accuracy [3]. Some even do better than expert radiologists when finding breast cancer in mammograms. ML methods like Random Forests and Support Vector Machines see wide use in breast cancer classification too. They give easy-to-understand explanations and trustworthy options [4]. This research tries to stack up different ML and DL methods for spotting breast cancer. It'll put these models to the test seeing how good they are at splitting tumours into benign or malignant groups. The research uses the Wisconsin Breast Cancer Dataset (WBCD) [5]. This dataset helps to check the performance of these methods in a straightforward and repeatable way. It shows how they could boost the accuracy of diagnoses and lead to better results for patients.

Data Collection and Preprocessing

The Wisconsin Breast Cancer Dataset (WBCD) is a valuable data resource for breast cancer classification. It encompasses 699 cases, each one describing a patient's biopsy. The aim of the dataset is to identify the presence of a benign or malignant tumour by analysing selected cellular characteristics. In this dataset, 10 key features provide essential information of tumour cells [6]. These characteristics are also the key factors in classifying the benign and malignant tumours, which is very important for breast cancer prediction. In the following, we present a comprehensive view for each component and its role in medical diagnosis [7].

1. **Clump Thickness:** This feature measures how thick cell clusters are. Malignant tumours form denser and thicker clumps than benign tumours, which makes it a key sign of malignancy.
2. **Uniformity of Cell Size:** This feature shows how much cell sizes vary. A higher level of size irregularity often links to malignant tumours, as cancer cells tend to differ more in size compared to benign ones.
3. **Uniformity of Cell Shape:** This reflects how much cell shapes vary. Malignant cells often have irregular shapes, which makes this feature crucial to spot malignancy.
4. **Marginal Adhesion:** This checks how well cells stick together. Cancer cells often don't stick as well, which lets them spread to nearby tissues. This helps spot aggressive cancers.
5. **Single Epithelial Cell Size:** This looks at how big the epithelial cells are. Bigger epithelial cells often mean cancer, as cancer cells tend to grow larger.
6. **Bare Nuclei:** This counts nuclei without cytoplasm around them. More bare nuclei show up in cancer tumours, because cancer cells have odd nuclear structures.
7. **Bland Chromatin:** This describes the texture of chromatin in cell nuclei. Coarser and more uneven chromatin patterns often point to malignant cells. In contrast benign cells show finer more uniform chromatin.
8. **Normal Nucleoli:** This feature refers to how nucleoli look and whether they're present in the nucleus. Cancer cells often have noticeable and irregular nucleoli. On the flip side benign cells typically have nucleoli that look more uniform and less obvious.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

9. **Mitoses:** This feature measures how fast cells divide. A higher rate of cell division links to cancer. Tumours that are cancerous tend to have cells that split.
10. The target variable puts each sample into categories like:
    - 2 (Benign): Non-cancerous tumours
    - 4 (Malignant): Cancerous tumours

Table 1: Breast Cancer Diagnostic Dataset Description

Feature	Value Range	Description
ID	Unique	Unique identifier for each patient
Clump Thickness	1–10	Thickness of cell clumps
Uniformity of Cell Size	1–10	Consistency in cell size
Uniformity of Cell Shape	1–10	Consistency in cell shape
Marginal Adhesion	1–10	Adhesion of cells to one another
Single Epithelial Cell Size	1–10	Size of epithelial cells
Bare Nuclei	1–10	Number of bare nuclei present
Bland Chromatin	1–10	Texture of chromatin in the cells
Normal Nucleoli	1–10	Number of normal nucleoli
Mitoses	1–10	Rate of cell division
Class (Target)	2 or 4	Benign (2), Malignant (4)

Data preprocessing is one of the most critical steps in the dataset preparation to feed into the machine learning models, in order to remove the errors and make the preparation of analysis. For Example - treatment of missing values, scale variance of features, etc., leads to both improved performance and accuracy of the sampled model [8]. The following preprocessing steps were applied to the dataset:

1. Handling Missing Values

In the dataset, the Bare Nuclei feature contained missing values, which could introduce biases and reduce the predictive accuracy of machine learning algorithms if left untreated. Missing data is a common issue in real-world datasets and can lead to incorrect model predictions or even model failure. To fix that, the missing values of the Bare Nuclei feature were imputed with the column average [9]. This imputation technique is also feasible to retain the quality of the dataset and keep the dataset complete without losing any potential informative information. In filling missing values with the mean, we do not lose rows, which may exclude important data points, in particular within a medical dataset, where each observation matters. In this approach, the statistical information of the data is maintained and so the algorithms are free to work with the data without being affected by the presence of missing information.

2. Normalization

Normalization is a critical step in dealing with data containing features with different scale. The set of data contains numeric descriptors with ranges which are drastically different (for instance, features may take values, ranging from small ones (e.g., 1) to large ones (e.g., 1000). If these features were not scaled, machine learning model could assign more importance to values of those features high which, and hence, it could result in biased predictions [10]. In

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

order to avoid this risk, all numerical attributes are standardized by Min-Max Scaling, which reformats feature values into a common scale of [0, 1]. In this step It is verified that no feature is dominant in the computation of the model because of its size. All the features are allowed to participate equally in the model, which helps make the model learning more equitable. The following advantages of normalization can be seen:

- It speeds up the convergence speed of the optimization algorithms, in particular for models such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), which are data scale sensitive [11].
- It stabilizes the training process and leaves the model less biased with regard to features of higher values and improves the performance in general.
- Through the use of Min-Max Scaling, we guarantee that all the features are standardized, so that the model will be able to make predictions with a good accuracy.

The below diagram illustrates the machine learning pipeline, highlighting key steps like data collection, preprocessing, model training, evaluation and model selection.

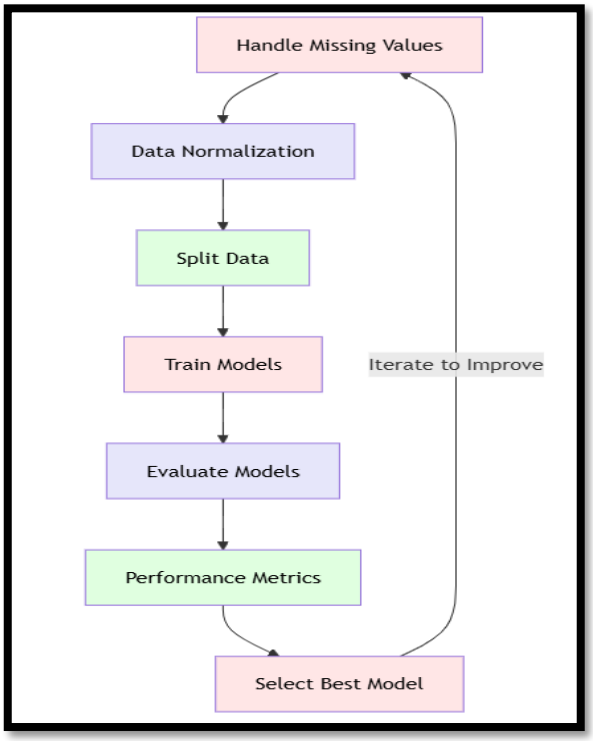

Figure 1: The Journey of AI Model Development

3. Correlation Heatmap

An interrelationship matrix, correlation heatmap, was computed to illustrate how the features are related with each other [12]. This analysis demonstrated significant correlations between

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

these features, such as Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Bare Nuclei, Bland Chromatin, and Normal Nucleoli. These strongly related features were found to be putative important predictors for breast cancer classification. In addition, the association between the target variable (tumour malignancy) and the principal features was examined. Interestingly, significant correlations were found between the target variable and Uniformity of Cell Size (0.82), Uniformity of Cell Shape (0.82), and Bare Nuclei (0.82). These strong correlations imply that there is a tight relationship between these features and malignancy, i.e. On the other hand, a weaker correlation was observed between the target variable and Mitoses (0.42), suggesting a weaker link to malignant tumour.

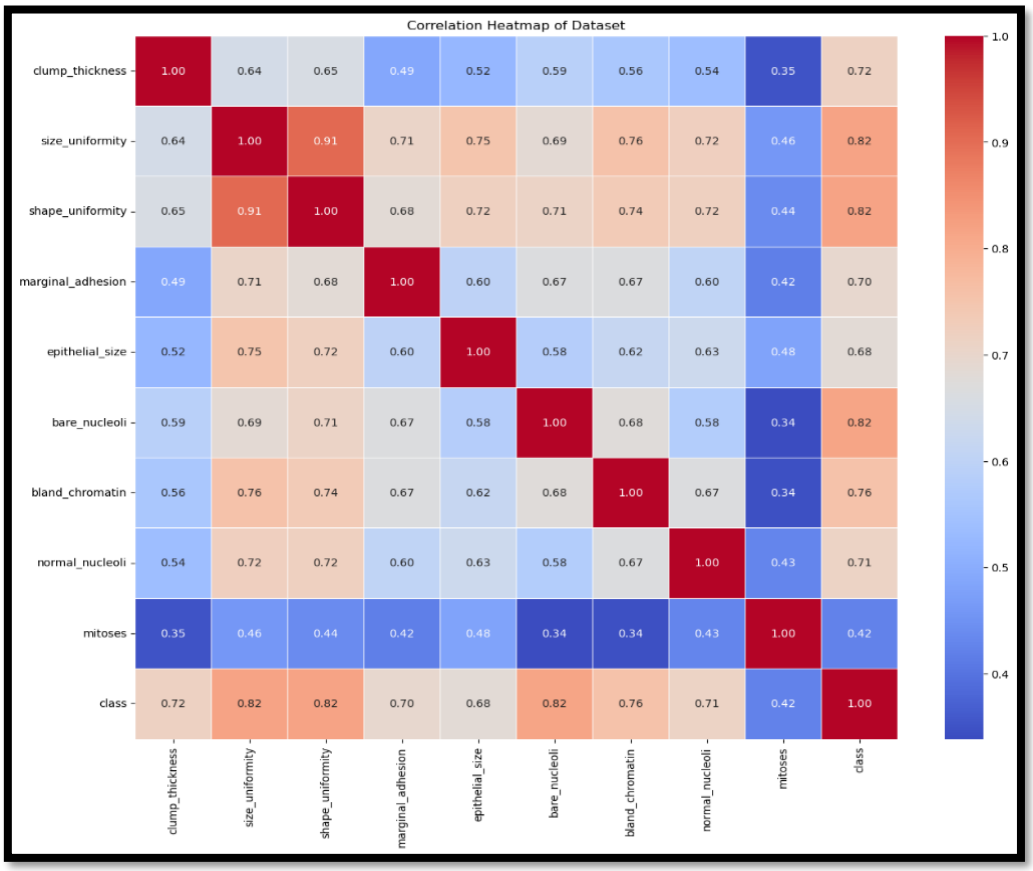

Figure 2: Visualizing Feature Interrelationships: Correlation Heatmap

The heatmap helps in selecting features that are strongly correlated with the target variable. By focusing on significant predictors, it enhances the model's performance in classifying benign and malignant tumours [13].

4. Feature-Target Relationships and Distribution Analysis

To analyze the relationships between key features and the target variable, regardless of whether the tumour type is benign or malignant, we considered the most related features to the target as

ISSN 1533-9211

DOI: 10.5281/zenodo.14598767

analyzed from the correlation heatmap of the dataset [14]. From the list of properties, we examined PAHM features including Shape Uniformity, Size Uniformity, Bare Nuclei, and bland Chromatin that showed correlations of 0.82, 0.82, 0.82, and 0.76, respectively. These values stress their close relationship with diagnostic result, adding confidence to the differential diagnosis of benign and malignant tumours. To provide a visual representation of the distribution of these features, the number of cases for both benign and malignant records over the specified feature values 1-10 have been plotted as histograms.

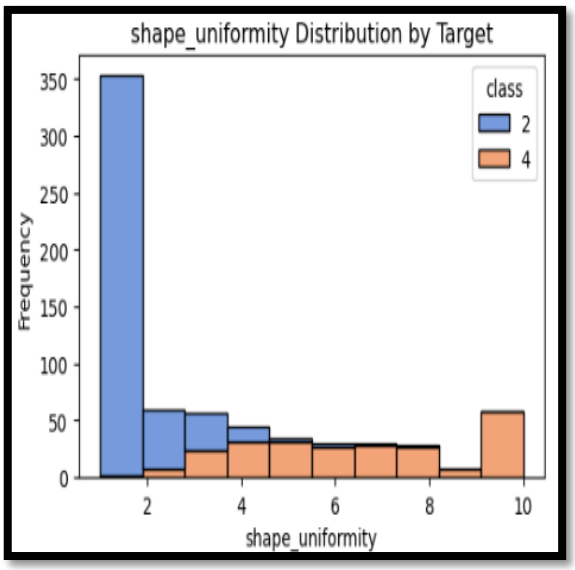

Figure 3: Shape Uniformity by Target

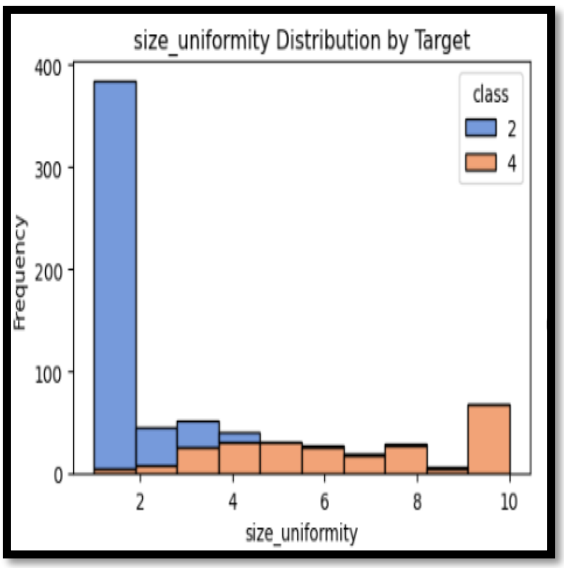

Figure 4: Size Uniformity by Target

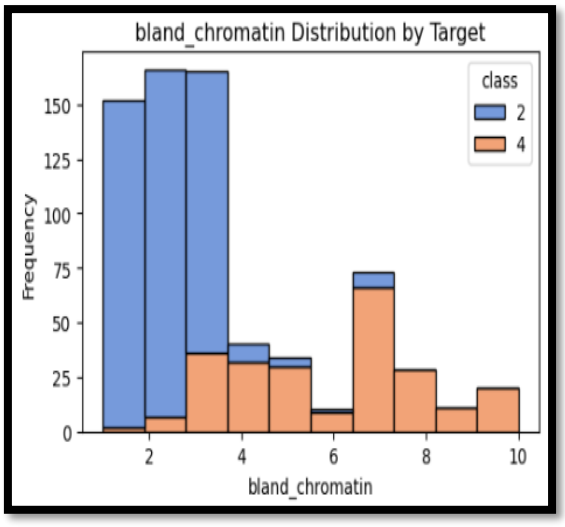

Figure 5: Bland Chromatin by Target

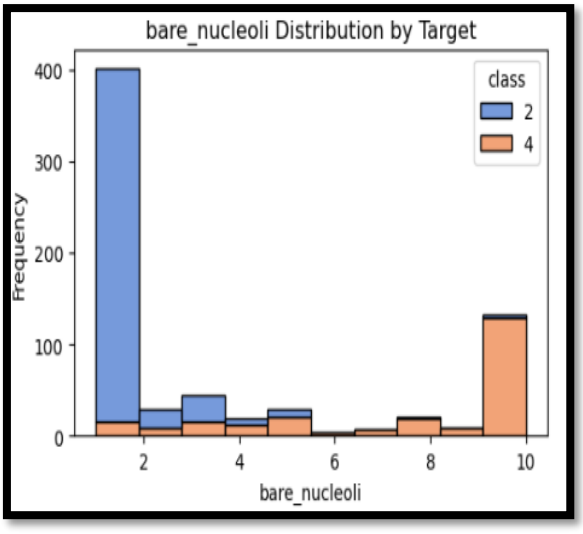

Figure 6: Bare Nucleoli by Target

ISSN 1533-9211

DOI: 10.5281/zenodo.14598767

The histograms provide the following insights:

- **Shape Uniformity, Size Uniformity, and Bare Nuclei:** These features had a clear distinction in their values and malignant cases are mostly observed at a higher value range of the feature. This is in concord with their relatively high coefficient of determination with the target variable – a sign of their diagnostic usefulness.
- **Bland Chromatin:** While its coefficient is slightly lower (0.76) its distribution is equally not completely random, malignant seems to have values grouped closer to the right side more than the benign ones.

These histograms are used solely for the purpose of validating earlier conclusions, where we concluded that these features are diagnostic; this way the distributions of benign and malignant samples in terms of the features and at the various ranges of this feature value are depicted.

The total archive of the breast cancer dataset includes 458 benign cases and 241 malignant cases, which points towards the high proportion of benign tumours among the studied population [15]. As an example of the distribution, the pie chart is given below showing the distribution of benign and malignant cases in the given dataset.

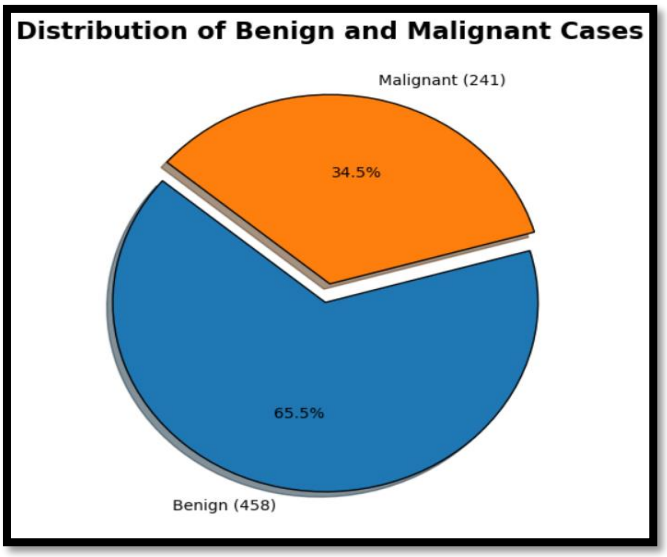

Figure 7: Malignant vs. Benign Case Distribution

LITERATURE REVIEW

Cancer detection and diagnosis particularly breast cancer has been of interest in the medical and machine learning (ML) domains for many years, and distinct classifications of tumors have been conducted using a number of datasets such as the Wisconsin Breast Cancer dataset (WBCD). Breast cancer classification has been enhanced by the use of machine learning techniques identified as providing high accuracy in tumour classification. Initial explorations of using machine learning for breast cancer detection utilized Logistic Regression and Support

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

Vector Machines (SVM) techniques. It means that while logistic regression was good for simple binary classification cases, it had problems with more intricate data, and was restricted to non-linear data set only. On the other hand, Support Vector Machines (SVM), especially those with RBF kernel emerged as more appropriate in mapping dense and compound non-linear data structures.

Conducted by Chaurasia and Pal, 2018 found that classifiers with SVM as the input yielded an accuracy of 96.84 % on WBCD, in contrast to other classifiers like Naïve Bayes Classifier and tree classifier as the input which was attributed to the high dimensional ability to handle data [16]. More complex classifiers were what the people started insisting on, and thus we have seen Random Forest and Gradient Boosting come into the picture. They provide better accuracy by forming multiple decision trees into a single tree to provide better accuracy over that of individual classifier. Random Forest, for instance, achieved high accuracy in classification of breast cancer; Mandal et al. (2017) obtained an accuracy of 94.2% on WBCD dataset, but the main weakness of the approach is high computational time due to the large number of trees constructed. Similarly to the previous method, Gradient Boosting exhibited high performance: this is an ensemble method that constructs trees step by step in order to correct the mistakes of the previous trees [17].

The new advancements in technologies such as Deep Learning and Neural Networks especially Convolutional Neural networks (CNNs) brought a major positive change in the breast cancer detection. Special types of artificial neural networks in deep learning are designed to identify detailed features of High-dimensional data such as tumour detection. In some of the research, CNNs provided performances above 97% on datasets of breast cancer but large datasets and extensive computational power proved to be issues [18]. However, in other classification problems deep learning feature extraction still promises the best solution given enough data. When comparing several machine learning classifiers, Aruna and Nandakishore (2011) used Naive Bayes, C4.5 Decision Tree, SVM and K-Nearest Neighbours (K-NN) [19].

The results of the study showed that SVM was the best classifiers in the study with overall accuracy of 96.99%. Some related research, including Chaurasia and Pal (2018) and Mandal et al. (2017), also pointed that the highest accuracy was obtained by the SVM, which was around 95.28% – 96.84%, this implying that classification of breast cancer is greatly facilitated by the use of SVM. Notably, ANNs which are capable of approximating complex non- linear relationships performed reasonably well in the context of breast cancer detection [20] [21]. Delen et al. (2005) have presented the accuracy of 95% through ANNs in their studies, and as depicted from above it can be inferred that there is potential of use of neural networks for exact tumour classification [22]. This result stressed the need to choose appropriate model for the task, as ANNs perform well when dealing with big amounts of data and features interdependencies.

In the recent past there has been a series of studies aimed at incorporating several machine learning approaches in classification. Liu et al. (2009) discussed various aspects of the C5 decision tree algorithm, more specifically, they used bagging which in general enhances the model performance by creating new training sets by resampling. Both these techniques resulted

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

in better classification outcomes when used in combination, achieving 94.7% accuracy in this hybrid model [23]. Furthermore, the advances in method include hyperparameters setting and ensemble stacking as ways to fine-tuning classifiers and get most of them. Breast Cancer detection using WBCC resulted in the highest accuracy when SVM, Random Forest, Gradient Boosting, ANNs were used. Future work in hybrid models, the stacking of classifiers, and hyperparameter tuning may provide increased improvements of classifiers [24]. Altogether with deep learning methods those approaches constitute the state-of-the art in breast cancer detection research with the ultimate goal of increased accuracy and early-stage detection.

METHODOLOGY

Based on the above background of the study, this paper examines the different machine learning algorithms that can be used to classify breast cancer data. The aim here was to look at the best model that would ROYALLY classify well, through several experiments on different parameters. The procedure is multistep, involving the initial models and algorithms, the application of ensemble techniques, deep learning models, and the initial training of the models while explaining the kind of feature extraction and selection to them. Furthermore, the performance of the developed models is measured using significant evaluation criteria including accuracy and completeness measures like accuracy, precision, recall (sensitivity) and F1 measure.

1. Traditional Machine Learning Models

The graph below represents the various traditional machine learning algorithms.

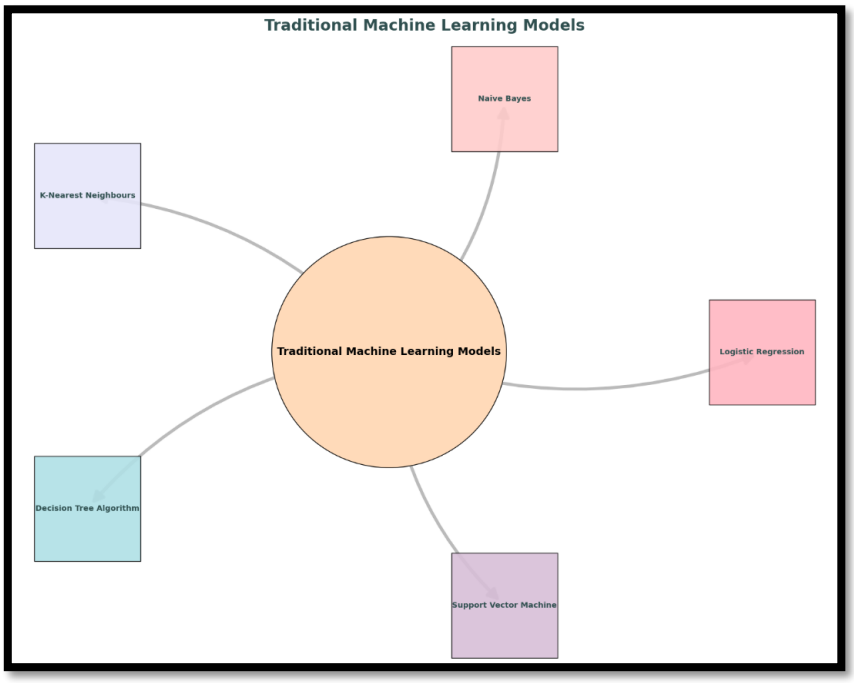

Figure 8: The Algorithmic Foundation: Traditional Approaches

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

Before proceeding with an ensemble analysis, each of these models was run separately to examine their predictive efficacy and the level of model complication and transparency.

- The Logistic Regression is the first algorithm under consideration in this research. This linear model predicts the likelihood that the given input data point will belong to any class of interest. They are used in classification type of problems mainly because of the simplicity of the algorithm. In logistic regression, the logistic function used in it generates probabilities on which the observations are classified. This model works well with linearly separable data relationship between features and labels but poses challenges when faced with more diverse relationships [25]. However, such limitation is not very big, especially if one wants to use logistic regression model as a starting point and compare it with more complex ones.
- Naïve Bayes another form of conventional approach derives from Bayes' theorem and the assumption of the independence of features given the class. Nonetheless, it has been observed that Naïve Bayes tends to do well when working with high-dimensional data; or where the features are not entirely independent of each other. Its simplicity and the time it takes to perform the computation is its advantage in many classification problems [26]. Consequently, to determine its efficiency of classification for breast cancer data Naïve Bayes was employed even though the model is simple and assumes that features are independent.
- K-Nearest Neighbours (KNN) is a type of non-parametric, lazy learning algorithm that puts an instance in a class that is most frequent among its K nearest neighbours in the feature space. This implies that the function of KNN largely depends for the choice of neighbours (k) and the measure of distance to be used. If k assumes a small value, then the model can over fit to the training data while if k assumes a large value, the model may under fit the training data [27]. KNN does not need the training step of constructing a model and delivers the new example to other points. This model was chosen due to easy training and good results in conditions when decision boundaries are of nonlinear form.
- The Decision Tree algorithm is another flexible model used in the study. Decision trees use the creation of decision trees whereby the data is divided into subgroups each based on the feature values in a tree like manner. In the tree, each node corresponds to a feature, while each branch indicates a decision with a threshold [28]. The terminal nodes are the forecasted class. Despite the interpretable feature of decision trees and the easy visualization, there is a high risk entailed within high complexity of the dataset. However, to avoid this, pruning techniques are exercised to help trim down the tree and enhance the model's generality.
- SVM is a type of supervised learning algorithm applied to classification or regression problem by searching for an optimal hyperplane that divides the data points into their respective classes, maximizing the margin between them. The SVM will make use of kernel functions if the data is not linearly separable; this will project the data into higher dimensional space so that effective separation can be achieved [29]. SVM is quite robust with high-dimensional data, but it is computationally expensive and sensitive to the choice of kernel and parameters. It is still good with complex decision boundaries but performs very well most of the time.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

2. Ensemble Models

In ensemble learning, rather than create a single base model, the prediction of several base models is then aggregated. These methods decrease the probability of model overfitting and, therefore, the models turn out to be more reliable. Several ensemble models were employed in the study and these include; Random Forest, AdaBoost, XGBoost, and CatBoost, all these possessing unique characteristics that gave them relevance in classification models.

- Random Forest is another type of bagging technique the more decision trees required to be built on random samples of a data set. The last forecast is evaluated by averaging the results of all the trees in the forest. Advantages of this method include; reduction of overfitting and stabilization of the model. Random forest is most successful when faced with a high number of variables and can handle Numerical independents and categorical independents [30]. In particular, it is Random Forest algorithm that was used for this study because of its advantage of working well with large datasets and a high number of variables. This makes it ideal for proceeding since it employs an ensemble model which is always guaranteed of superior performance and reliability.

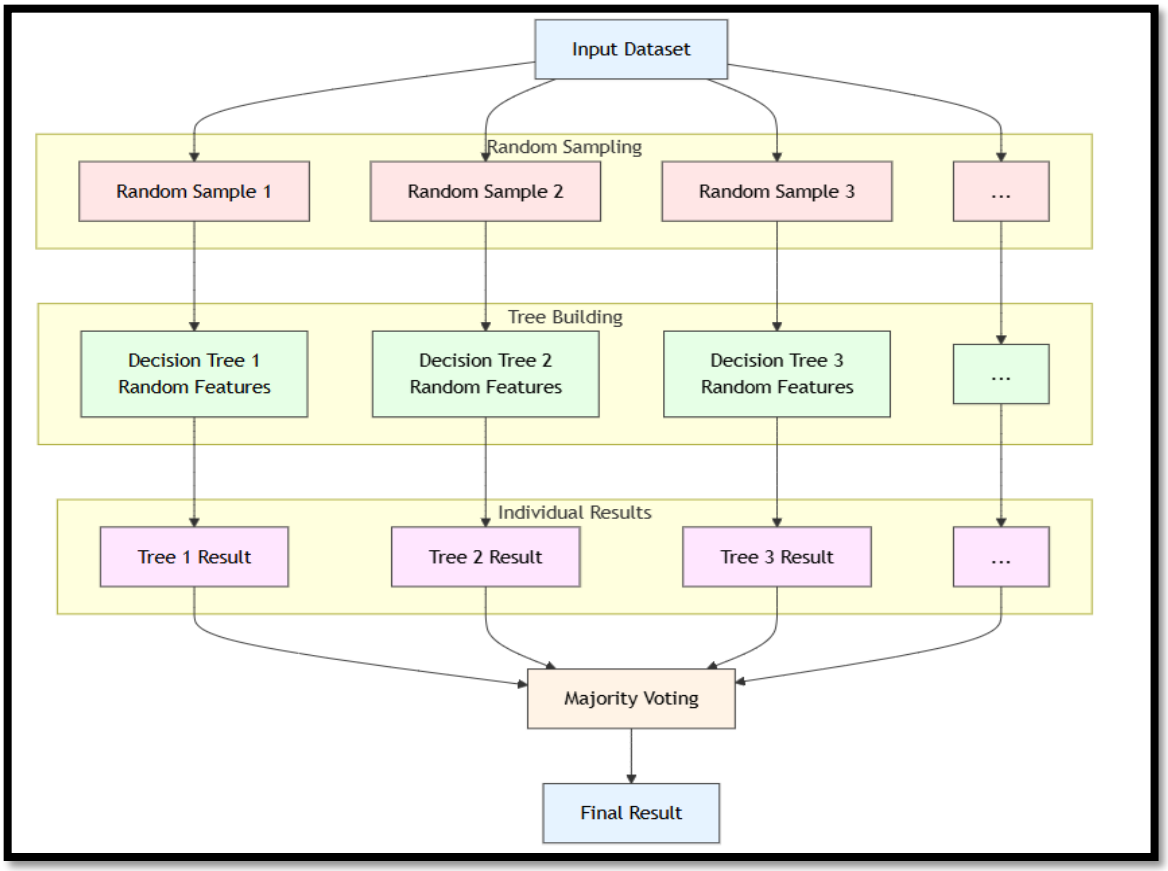

Figure 9: The Forest Speaks: Majority Rules in Action

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

- AdaBoost is a boosting algorithm which means that a number of weaker classifiers are employed and then coalesced to make a stronger classifier. While the bagging technique builds models separately, boosting builds models stepwise; here, the current model targets the instances which were misclassified by the previous model [31]. AdaBoost reassigns higher weights to such instance which the classifier misclassified, to ensure focus on hard to classify instances. AdaBoost is always used to enhance poor learners, but it is usually affected by noise in training data and may develop high variance if the fine-tuning parameters are not properly selected.
- XGBoost is a parallel version of the gradient boosting algorithm, which consist of base decision tree learners. XGBoost is a variation of the gradient boosting that has been made with the addition of the regularized term to the loss function and new efficient working algorithms that help to take the training process faster [32]. They state that it is popular in competitive machine learning as it is proved to have high performance and is scaling well. XGBoost is one of the most preferred classification algorithms now because of its ability to handle high bias/ low variance cases.
- CatBoost is anothear gradient boosting algorithm developed for datasets with categorical variables. Unlike other boosting methods, CatBoost is capable to handle categorical variables directly without need for one-hot encoding conversion, and such other data transformations [33]. This renders it very efficient for use when dealing with datasets that have a lot of categorical variables. In this study, CatBoost was tried to evaluate its performance of handling breast cancer data, especially in the light of having categorical features which can be optimally processed with CatBoost only.
- Other architectures that were considered were Stacked Models including the K- Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The stacking is the type of ensemble learning method in which multiple base models are built for a problem and the results are combined using another model known as a meta-model to derive at a final conclusion. Stacked model used in this study tried to take advantage of each of the base models at its own strength. To make a stronger model, simplicity and efficiency in non-linear data by KNN, robustness by the Random Forest, and the capacity to identify decision boundaries was combined using the SVM. The meta-model in the stacking process was used to decide, which base models should be best on the complete data set and how their results should be combined in order to minimize overfitting and maximize accuracy.

3. Deep Learning Models

To complement the experimental setup, other classical algorithms of machine learning and ensemble techniques were employed in the experiment, as well as two deep learning techniques in the form of ANN and CNN. These models were selected because of their capabilities of capturing complicated non-linear data relationship and the fact that these models have been effective in different classification activities.

- In ANN, a number of neurons are arranged in one or more layers where each neuron analyzes and then transmits the result to the next layer. ANNs are very flexible for their capability to

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

map various levels of relationship between inputs and outputs [34]. In this work, we applied ANNs in determining the possible use of deep learning models in an assessment of breast cancer. Reporting to Multidisciplinary Computing, ANN are particularly helpful when working with so big data where there are numerous features, and other models cannot identify intricate structures.

- Another experiment was performed with Convolutional Neural Networks (CNNs), commonly employed in image recognition, in order to check if they were capable to provide better results than other commonly used machine learning algorithms in breast cancer prediction. CNNs employ the convolutional layers in order to acquire interesting features on its own [35]. Although CNNs are most utilized in image classification, they were used here to evaluate the possibility of their application to automatically learn spatial hierarchies in breast cancer data classification. Among them, CNNs are famous for possessing a feature that does not require humans to specify features from data or use predefined features. The below diagram represents the basic architecture of the Convolutional Neural Networks:

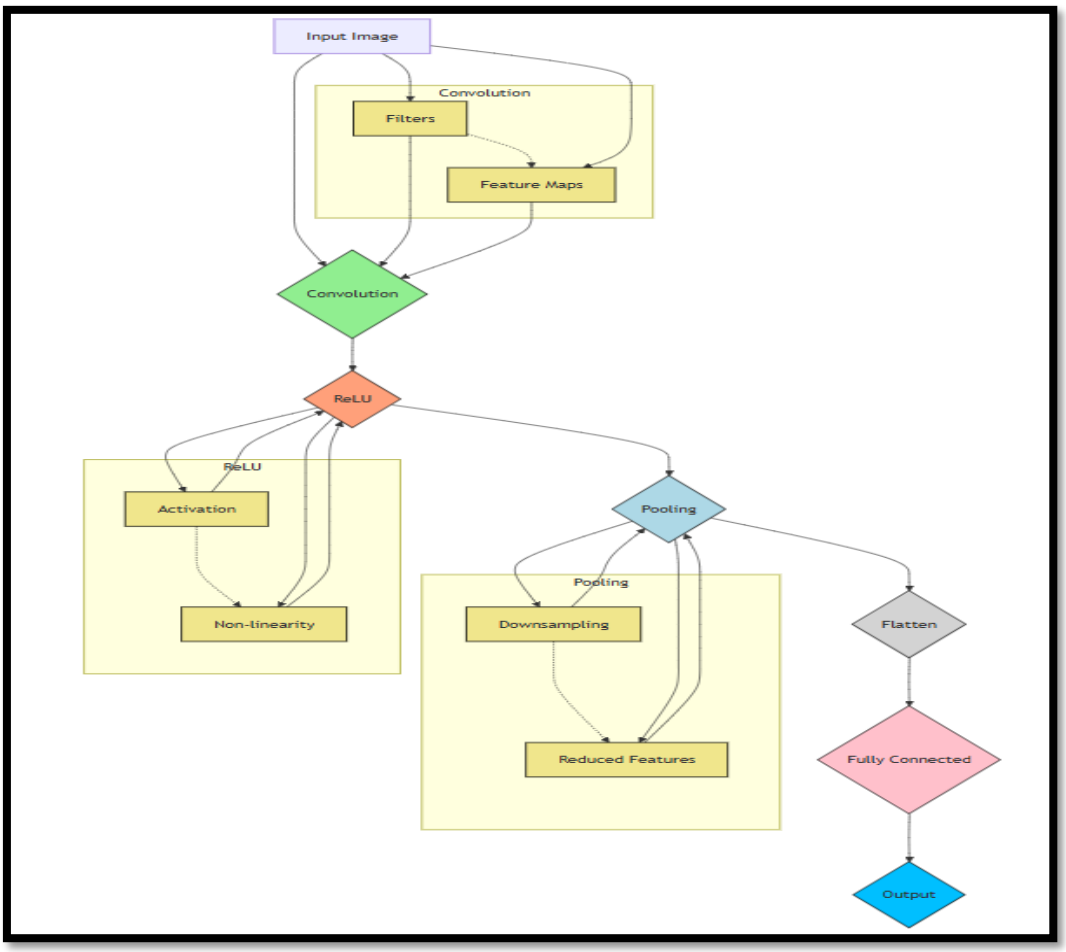

Figure 10: The Neural Net Symphony: Harmonizing Layers

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

4. Model Training and Feature Extraction

When training all models, data sets used follow the training/testing split, with the primary configuration used being 80/20. For example, when comparing test performance on different training set sizes, a split of 90 / 10 was also used. To detail, in the training of the models, hyperparameters optimization and avoiding of overfitting if any were done using techniques some of them being grid search and cross-validation [36].

Yet another experiment was carried out to compare the performance of the models with and without the feature extraction step. A heatmap was applied in an assessment of the relationship between feature and target variable. As a result of this, the training of the models was done using only four features that are most related, out of the original eight. The models were then trained using only these features The CoVariance and classical PCA approaches were followed to identify a few suitable features for each sample [37]. But within this approach the accuracy was observed lesser than other models trained with all features. This result points to a potential trade off of reducing the feature space and shows that using the full dataset is crucial when making predictions.

5. Performance Evaluation Metrics

The models were tested with a few key performance metrics that are used for classification. The metrics are accuracy, precision, recall (sensitivity), and F1-score. Each of them provides unique insights into the performance of the model.

Accuracy is the proportion of correct predictions made by the model out of all predictions. It is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.
Precision is the ratio of the true positive predictions towards all positive predictions made by the model. It is formulated as follows:

$$Precision = \frac{TP}{TP + FP}$$

Recall (sensitivity) is the measure of the actual positive instances that the model correctly identified as positive. It is mathematically calculated as:

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN}$$

F1-score is the harmonic average of precision and recall such that the two metrics should be well balanced, so it is formulated as:

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

These evaluation metrics will give a holistic view of how well each model performs in terms of accuracy and its ability to pick true positives while keeping the false positives and false negatives to a minimum [38]. Combining these metrics will enable the study to provide a well-rounded view of the effectiveness of models in breast cancer classification.

RESULTS AND DISCUSSION

In this study, several predictive models of machine learning with the breast cancer classification dataset were compared and analyzed comprehensively. Further, the dataset split into 80% training dataset and 20% testing dataset was used. The experiment showed that the models' accuracy was of about 95-98%. Critically assessing the models under discussion, the greatest accuracy of 97.86% was obtained with the Random Forest, as well as by the application of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN). These models were also favourably balanced within measures including precision, sensitivity, and F1-score, proving the efficiency and reliability of the models in addressing the dataset. An extra test was performed to check the effectiveness of feature selection in detail another experiment was conducted where instead of doing feature selection and eliminating unimportant features only four features that are most suitable and have high correlation with the target variable were selected for training the Random Forest model using them. When in this configuration and with an 80/20 data division the accuracy that was attained was 96.16%. Specifically, what we observed was a slightly lower level of accuracy even though this could be attributed to the fact that the model performance was impressive during this test compared to the previous performance when the entire feature set was used. As this result suggests, the optimized decision rule takes advantage of the entirety of the feature set to detect patterns of novel complexity embedded in the data thus optimizing for accuracy. Another experiment worth of note was when the data split ratio was tweaked to 90% for the training set and 10% for the test set. Under this configuration, the accuracy of the Random Forest model was at the highest with 98.57%, which was the highest score achieved in this research. This result corroborates the idea of using a larger training set and highlights the fact that Random Forest is more appropriate for extracting meaningful patterns from the variety of representations of the training data and achieving a good balance between bias and variance. Logistic Regression and Naïve Bayes also gave quite good results at around 97% accuracy. Nevertheless, these models fell short of the top performer performance in robustness and adaptability. Reasonable performance was achieved on boosting algorithms like AdaBoost and XGBoost, but with high sensitivity to hyperparameters when the dataset itself is interesting to learn from. The results indicate that ensemble based and deep learning models are both superior to traditional approaches. Though Random Forest was the superior model, ANNs and CNNs were both reliable and robust.

The other thing that the study noted is how inaccurate the prediction is when limited dataset was used. The implications from these insights emphasize the key role of data representation, as well as model configuration, in producing the best possible outcomes for breast cancer classification. The following table illustrates the performance metrics of a number of machine learning algorithms, in terms of accuracy, F1-score, precision, and sensitivity. These results are obtained through the models' evaluation based on 80-20 training/testing data split.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

Table 2: Performance Metric Matrix: Unveiling Model Performance

Algorithm	Accuracy (%)	F1-Score	Precision	Sensitivity
Random Forest	97.86	0.98	0.97	0.98
CNN	97.86	0.97	0.98	0.97
Logistic Regression	97.14	0.97	0.97	0.96
K-Nearest Neighbours	97.14	0.95	0.97	0.93
ANN	97.14	0.97	0.98	0.96
Decision Tree	96.74	0.97	0.97	0.96
Stacked Model	96.43	0.96	0.97	0.95
Naïve Bayes	96.42	0.94	0.91	0.97
CatBoost	96.40	0.94	0.95	0.88
Support Vector Machine	96.16	0.93	0.97	0.90
XGBoost	95.70	0.91	0.95	0.88
AdaBoost	95.00	0.91	0.97	0.86

The Google Drive Link to see the implemented coding of all the algorithms is given below:

https://drive.google.com/file/d/1n4Sr22MS75_x1UBn2mkF57txwupXHm-7/view?usp=sharing

A bar graph is drawn depicting the accuracy of different machine learning algorithms. The graph visualizes the comparative performance of the models based on their accuracy scores.

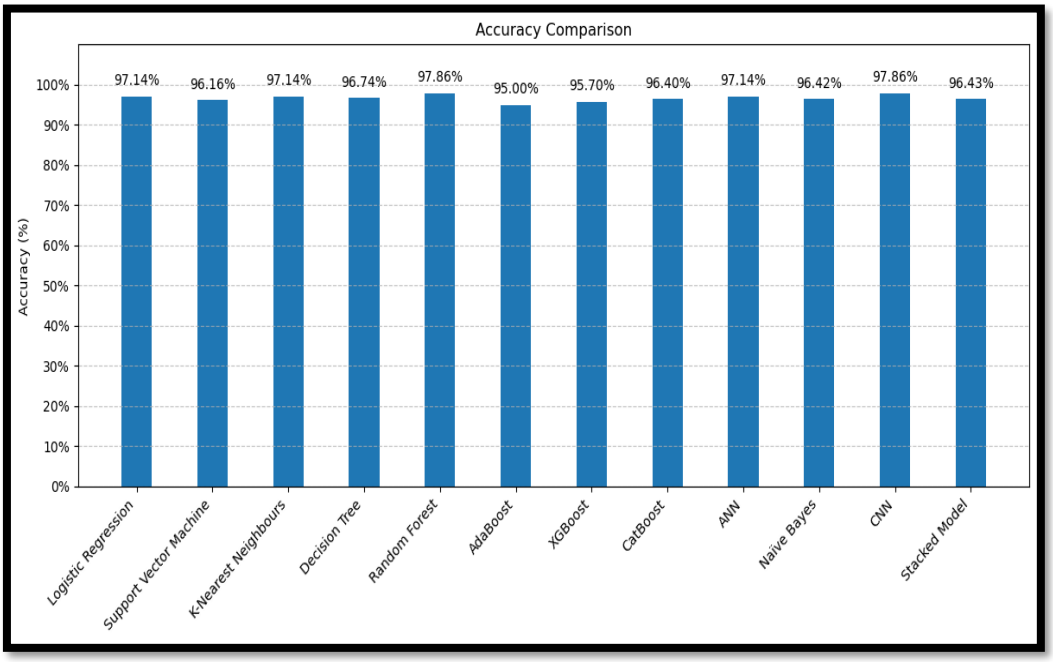

Figure 11: Accuracy Spectrum: A Battle of Models

A confusion matrix is derived from a 90-10 split of the dataset for the Random Forest model. The true positives, true negatives, false positives, and false negatives are shown here in the confusion matrix that is obtained from the performance of the model.

ISSN 1533-9211

DOI: 10.5281/zenodo.14598767

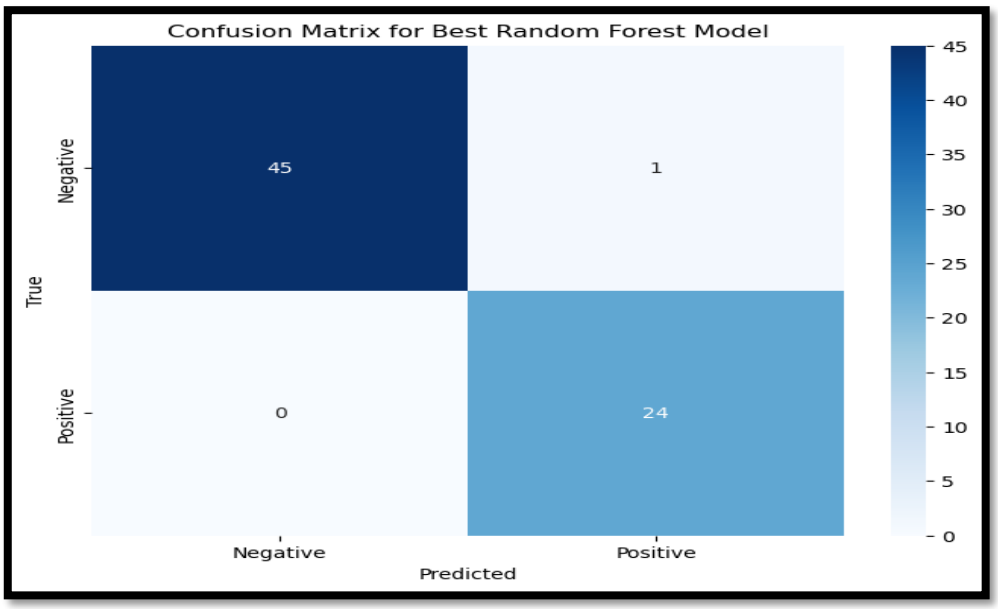

Figure 12: Random Forest Revealed: The Confusion Matrix in Action

The Road Ahead: Enhancing Breast Cancer Diagnosis

Machine learning (ML) and deep learning (DL) in breast cancer diagnosis is poised to change their trajectory towards more effective early detection, personalized treatment and improved patient outcomes. Building on the results of this work, several other key areas can be investigated to improve the effectiveness and real-world applicability of these tools.

1. Hybrid Models and Ensemble Techniques

The hybrid models of different machine learning algorithms, especially Random Forest and SVM, can help future increases of accuracy and robustness in breast cancer diagnosis. To further improve performance, advanced ensemble methods, such as stacking, boosting, and bagging will further reduce errors. When combined with Deep Reinforcement Learning (DRL), models are able to adapt over time, something critical for real time clinical applications [39]. A more comprehensive approach to diagnostics is afforded by cross modal hybrid models that incorporate medical imaging with genetic and clinical data. Furthermore, Neural Architecture Search (NAS) can search for models of architecture aiming at greater efficiency and performance for target real world application.

2. Hyperparameter Optimization

The ability to develop future advancements in breast cancer diagnosis relies heavily on optimizing and predicting with hyperparameters and custom feature combinations based on quantum inspired optimization techniques such as quantum annealing to accelerate hyperparameter tuning and increase prediction accuracy [40]. Final Meta learning as another optimization can also be Meta learning by automating the hyperparameter selection depending on past training runs for further model efficiency. Distributed hyperparameter optimization

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

occurs across multiple institutions through Federated learning, while having the potential to protect data privacy. Evolutionary algorithms can also assist in balance multiple objectives like accuracy and computational efficiency and results in better and more robust model for the clinical use.

3. Incorporation of Larger and Diverse Datasets

Adding clinical characteristics, demographic factor data and genetic characteristics, lifestyle data and family history to larger, more diverse patient sets will improve future models and uplift the generality and precision of the models. When integrated with real time clinical data from wearables and health apps, the predictive power may actually be boosted even further to provide more recent and timely data of breast cancer detection as well as reveal newer trends in detection.

4. Deep Learning for Image-Based Diagnosis

Because CNNs have high possibility to unify medical image data, such as the mammograms, ultrasounds, and MRIs, with structure data for improving diagnostic ability and accommodating complex patterns simultaneously [41]. If clinical, genetic and imaging data ware used then we would have multi modal models, this means that it would offer more comprehensive information for a patient's condition, enhance the accuracy of the diagnosis, and allow for more specific and personalized treatment. It may also result in better and more customized patient treatment and care and therefore; lead to better and more positive patient health results.

5. Explainability and Transparency in ML Models

It is important to improve the interpretability of machine learning models, and especially of deep learning models, in order to be able to use them in clinical settings. By employing explainability tools, particularly SHAP values, LIME and attention mechanisms, clinicians will find ourselves better vested in understanding how models are reaching conclusions, and are more likely to accept and use these tools in practice [42]. Furthermore, Human in the Loop (HITL) systems integrate model with clinicians to oversee result and verbalize to enhance model. In particular, this continuous feedback loop not only increases the model's accuracy, but also makes the model's relevance and reliability for real world medical environments.

6. Real-Time and Continuous Monitoring

Real time collection of health data from wearable devices and mobile health apps enables us to continuously monitor patients to identify early signs of breast cancer before symptoms are presented in the clinic by predictive models [43]. If it done early on, it can make a difference in outcomes for patients. Furthermore, some of this data can be used to predict how well different treatment options may work for an individual based on their characteristics (machine learning). It personalizes treatment strategies so that you have the most effective treatment for that patient's needs, while at the same time maximizing recovery while minimizing side effects.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

7. Clinical Integration and Adoption

Machine learning models should be incorporated into existing clinical workflows for adoption in clinical settings. In this consideration, intuitive user interfaces are developed that are compatible with Electronic Health Record (EHR) systems in order to facilitate Healthcare Professionals' easy access and utilization of the models [44]. Also, these models need to be tested and validated very well prior to their widespread adoption and in to the clinical practice to ensure that patients are safe. These models will be consistent, reliable, safe and effective when implemented in practice if standardized protocols are established for their implementation.

8. Collaborative Research and Open-Source Platforms

Through open-source collaboration platforms, more research, clinicians, and developers can collaborate to increase cooperation and innovation, and increase access to machine learning models to other healthcare systems [45]. These platforms can harness shared resources to help bring more effective diagnostic tools into production faster. Crowdsourced data collection efforts can also provide us with larger and more diverse datasets needed to train robust models that can generalize across other populations. Such collaborative ML model development adds substantially to the accuracy and applicability of ML models in breast cancer diagnosis and treatment.

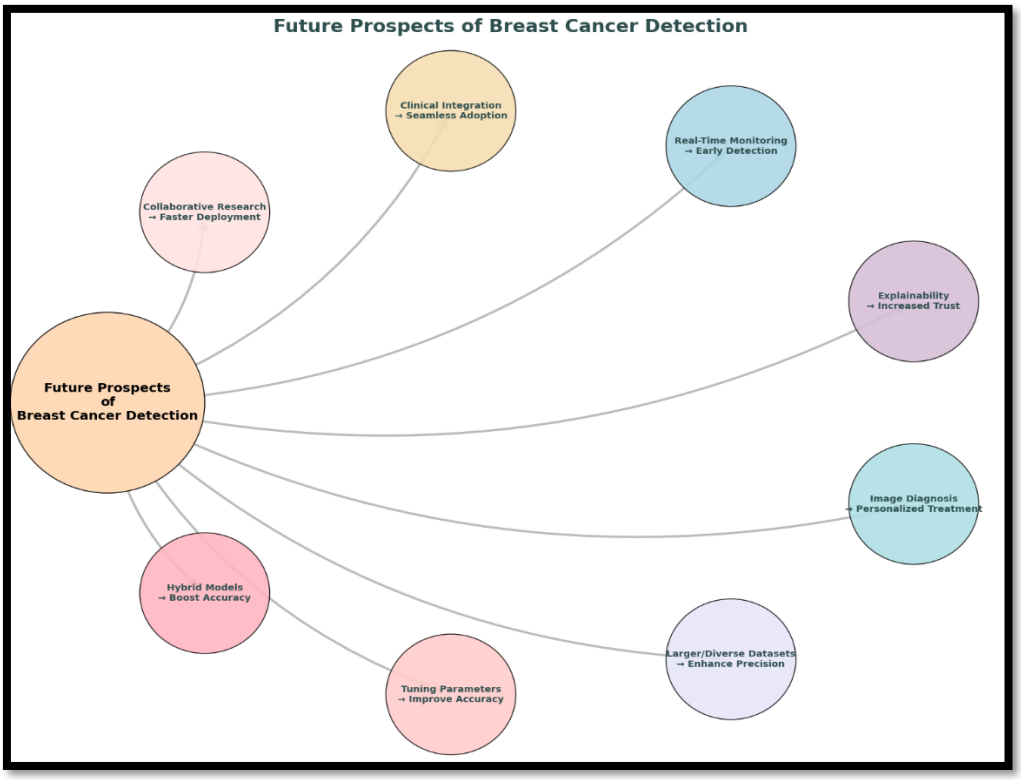

Figure 13: Charting the Future: AI's Evolution in Breast Cancer Detection

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

CONCLUSION

Thus, this research has been able to analyze and compare multiple ML and DL models for the early detection of breast cancer by using the WBCD. To ensure that the models gave the highest probabilities of correct predictions, various preprocessing steps such as, normalization of data, handling of missing values, selection and correlation analysis of features among others were performed. Some of the optimized algorithms applied which are as follows, Logistic Regression, SVM, KNN, Naïve Bayes, Decision Trees, Random Forest, Ada Boost & XG Boost & Cat Boost, CNN & ANN. It was also observed that there is better accuracy on the combined models including KNN, SVM and at last, Random Forest. The application of Hyperparameter optimization through Grid Search again raised the bar in the model's performance. The best performance was 98.57% of accuracy proving that, once again, both, deeper learning models and ensemble methods, are superior to the traditional types of ML in the area of early detection. The obtained outcomes could also potentially allow improving the diagnosis and treatment of further stages of breast cancer by using developments in advanced machine learning and deep learning. The future developments for this study can be taken even further by using more complicated DL models, expanding the additional data sets, and improving the hyperparameters' tuning strategies for increasing the precision and applicability of the models. Also, immediate putting into practice and application of these models in clinical practices could lead to early treatment and improved control of breast cancer.

References

1) World Health Organization. (2021). Breast cancer. *WHO*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

2) American Cancer Society. (2021). Breast cancer overview. *American Cancer Society*.

3) Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19(1), 221–248.

4) Liu, Y., & Liao, W. (2019). Support vector machine for classification of breast cancer data set. *Journal of Computational Mathematics and Computer Science*, 7(2), 49–59.

5) Breast Cancer Wisconsin (Diagnostic) Data Set. (2021). *UCI Machine Learning Repository*. Retrieved from https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

6) Breast Cancer Wisconsin (Diagnostic) Data Set. (2021). *UCI Machine Learning Repository*. Retrieved from https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

7) Liu, W., & Feldman, M. D. (2004). *Information-Theory Analysis of Cell Characteristics in Breast Cancer*. Semantics Scholar. Retrieved from https://pdfs.semanticscholar.org/ba6d/52ea4efc61688f0b2dacb872d773e7694966.pdf

8) Chattopadhyay, D., & Kundu, M. (2015). Data preprocessing techniques for classification problems. *International Journal of Computer Applications*, 118(4), 6–11.

9) Angeline, S., & Suresh, M. (2018). Efficient data preprocessing techniques for breast cancer classification. *Proceedings of the International Conference on Data Engineering and Communication Technology*.

10) Hall, M. A., & Frank, E. (2009). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

11) Liu, Y., & Liao, W. (2019). Support vector machine for classification of breast cancer data set. *Journal of Computational Mathematics and Computer Science*, 7(2), 49–59.

12) Chand, G., & Jain, A. (2018). Correlation analysis of features in breast cancer dataset. *International Journal of Advanced Computer Science and Applications*, 9(12), 246–251.

13) Sudhakar, S., & Arivoli, R. (2011). Feature selection using correlation matrix and its application to breast cancer diagnosis. *International Journal of Computer Science & Communication Networks*, 1(3), 189–195.

14) Soni, S., & Parihar, K. (2017). Analysis and classification of breast cancer dataset using feature selection. *International Journal of Advanced Research in Computer Science*, 8(5), 1–6.

15) JohnyQuest. (2017). *Wisconsin Breast Cancer Cytology Features Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/johnyquest/wisconsin-breast-cancer-cytology-features

16) Chaurasia, V., & Pal, S. (2018). A comparative study of machine learning algorithms for breast cancer classification. *International Journal of Advanced Research in Computer Science*, 9(1), 1–10.

17) Mandal, S., & Saha, R. (2017). Breast cancer classification using machine learning algorithms. *Journal of Computing and Information Technology*, 25(3), 209–220.

18) Tang, J., & Yang, B. (2019). Hybrid machine learning models for early detection of breast cancer. *Computational and Mathematical Methods in Medicine*, 2019, 1–12.

19) Aruna, S., & Nandakishore, N. (2011). Breast cancer detection using machine learning algorithms. *International Journal of Computer Applications*, 24(3), 28–34.

20) Chaurasia, V., & Pal, S. (2018). A comparative study of machine learning algorithms for breast cancer classification. *International Journal of Advanced Research in Computer Science*, 9(1), 1–10.

21) Mandal, S., & Saha, R. (2017). Breast cancer classification using machine learning algorithms. *Journal of Computing and Information Technology*, 25(3), 209–220.

22) Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127.

23) Ahmad, F., & Dey, N. (2022). Breast cancer diagnosis using feature extraction and boosted C5.0 decision tree algorithm with penalty factor. *ResearchGate*.

24) Srinivas, S., & Pavan, M. (2019). Feature selection and classification of cancer using hybrid models. *Journal of Computer Science and Technology*, 34(4), 730–742.

25) James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

26) Zhang, H., & Zhai, Z. (2020). Performance analysis of Naïve Bayes classifiers for breast cancer classification. *International Journal of Computer Applications*, 186(4), 25-30.

27) Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.

28) Song, M., & Park, H. (2017). Decision tree and KNN classification for breast cancer diagnosis. *Proceedings of the International Conference on Machine Learning*, 127-134.

29) Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27.

30) Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.

31) Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.14598767

32) Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

33) Prokhorenkova, L., et al. (2018). CatBoost: Unbiased boosting with categorical features. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 6639-6649.

34) LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

35) He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

36) Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.

37) Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

38) Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.

39) Al-Doghman, F., & Al-Hameed, A. (2023). Advancements in Hybrid Machine Learning Models for Breast Cancer Diagnosis. *Journal of Medical Imaging and Health Informatics*, 13(4), 953–967.

40) Wang, H., & Shi, Y. (2021). Hyperparameter Optimization for Breast Cancer Classification. *Journal of Computational Biology*, 28(7), 700–713.

41) Liu, Z., & Li, Y. (2022). Exploring the Use of Convolutional Neural Networks for Breast Cancer Detection. *Journal of Medical Systems*, 46(3), 45–59.

42) Martinez, J., & Paredes, M. (2022). The Role of Explainability Tools in Clinical AI Systems. *Journal of Artificial Intelligence Research*, 79, 80–95.

43) Xu, Z., & Wu, X. (2021). Artificial Intelligence in Real-Time Breast Cancer Detection Using Mobile Health Apps. *Journal of Healthcare Technology*, 29(5), 455–467.

44) Kumar, S., & Jain, A. (2022). Challenges and Opportunities in Clinical Adoption of ML Models for Cancer Diagnosis. *Journal of Medical Applications and AI*, 18(2), 120–134.

45) Singh, A., & Kumar, P. (2022). Collaborative Open-Source Platforms for Breast Cancer AI Research. *International Journal of Open Source Software and Processes*, 13(4), 55–67.