IJEMD-CSAI, 4 (1) (2025) https://doi.org/ 10.54938/ijemdcsai.2025.04.1.475

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

An Empirical Comparison of Machine Learning Algorithms for Breast Cancer Detection

Hamza Sabo Maccido [1*]

1. Department of Electrical and Computer Engineering, Faculty of Engineering, Baze University, Abuja, Nigeria.

Abstract

Breast cancer continues to be a major global health challenge, necessitating the development of accurate and reliable diagnostic systems. This study presents a comparative evaluation of multiple machine learning classification models aimed at enhancing breast cancer detection. Three feature selection techniques which are Principal Component Analysis (PCA), Pearson Correlation Coefficient (PCC), and Backpropagation Neural Networks (BNN) were employed to reduce dimensionality and extract relevant features. The performance of six classifiers which are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), and Artificial Neural Network (ANN) was analyzed based on accuracy, precision, recall, specificity and f1-Score. Results show that among the evaluated classifiers, Random Forest and Support Vector Machine (SVM) consistently delivered the highest performance, with Random Forest achieving up to 98.8% accuracy and SVM up to 98.0%, particularly when trained on features selected through Backpropagation Neural Networks (BNN). K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN) also demonstrated strong results, outperforming traditional models like Logistic Regression and Decision Tree in most scenarios. These outcomes underscore the superior classification capabilities of non-linear and ensemble-based models in handling complex feature interactions, affirming their suitability for accurate and robust breast cancer detection.

Keywords: Breast Cancer Classification; Machine Learning Models; Predictive Analysis; Feature Selection Techniques; Diagnostic Accuracy.

Email Address : hamza.maccido@bazeuniversity.edu.ng (Hamza Sabo Maccido *)

1. INTRODUCTION

Breast cancer remains one of the most frequently diagnosed cancers and is the leading cause of cancer-related deaths among women globally [1]. It is particularly prevalent in regions such as Eastern Europe and Africa, where mortality rates are disproportionately high [2]. According to the World Health Organization (WHO), more than 2.3 million women were diagnosed with breast cancer in 2020, resulting in approximately 685,000 deaths. Furthermore, by the end of that year, about 7.8 million women had been diagnosed within the preceding five years, making it the most commonly diagnosed cancer worldwide. Alarmingly, the number of cancer diagnoses is projected to rise to 19.3 million by 2025 [3].

Early and accurate diagnosis is essential to improving survival rates. While five-year survival in early-stage breast cancer can be as high as 81%, this drops drastically to 35% in late-stage cases [4]. Conventional diagnostic methods, including mammography, MRI, and ultrasound, although widely used, are limited by factors such as high costs, image quality issues, human error, and reduced sensitivity in dense breast tissues [5],[6],[7]. These limitations underscore the need for automated, accurate, and cost-effective diagnostic systems.

With the rapid growth of medical data and the increasing complexity of diagnostic processes, Machine Learning (ML) has emerged as a promising tool for disease diagnosis, particularly in breast cancer research [8]. ML techniques can uncover hidden patterns in large datasets and assist clinicians in making more informed and consistent decisions [9]. Various ML classifiers—such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Naive Bayes (NB) have demonstrated significant potential in medical diagnosis tasks, including cancer detection [10].

However, using raw, high-dimensional breast cancer data introduces noise and redundancy, which can reduce classification accuracy. Feature selection (FS) plays a critical role in identifying the most informative features for classification and improving the predictive performance of models [11]. In this context, both linear (e.g., PCA) and non-linear (e.g., Backpropagation Neural Network (BNN)) feature selection techniques are being explored.

Furthermore, breast cancer classification is often too complex to be effectively addressed by a single classifier, necessitating the use of ensemble learning methods. Ensemble classification combines multiple classifiers to enhance robustness and accuracy by leveraging the strengths of individual models while mitigating their weaknesses [12], [13].

This study focuses on the performance evaluation of individual classification models: SVM, KNN, ANN, NB, Logistic Regression (LR), and Decision Tree (DT) for breast cancer detection using the Wisconsin Breast Cancer Diagnostic Dataset (WBCD). FS is performed using a Principal Component Analysis (PCA), Pearson Correlation Coefficient (PCC) and BNN to reduce dimensionality and enhance model performance. The primary objective is to assess and compare the effectiveness of these models in accurately classifying breast cancer cases, thereby supporting early diagnosis, minimizing false positives, and aiding clinicians in informed decision-making.

2. MATERIALS & METHODS

2.1. Dataset and Feature Selection

This study utilizes the WBCD, a widely recognized dataset for evaluating the performance of classification models in breast cancer detection. The dataset contains 569 instances, each characterized by 30 numerical features derived from digitized images of fine needle aspirate (FNA) of breast masses. These features describe various properties of the cell nuclei, including radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension, among others. The outcome variable is a binary classification label, indicating whether a tumor is benign or malignant [14].

Before model training and evaluation, the dataset was preprocessed to ensure quality and consistency. This involved normalizing the feature values to a standard scale, checking for and handling missing data, and dividing the dataset into training (80%) and testing (20%) subsets to facilitate model evaluation on unseen data.

To enhance the effectiveness and efficiency of classification, FS was employed as a critical step in the data preparation process. Selecting the most relevant features helps reduce overfitting, improve accuracy, and decrease computational complexity. For this purpose, both linear and nonlinear feature selection methods were applied. PCA and PCC were used to identify linearly significant and uncorrelated features, while a BNN was utilized to identify non-linear dependencies and interactions among the features [15]. The reduced feature sets obtained from these techniques were then used as input for evaluating the performance of various classification models.

In this study, six classification models were employed: SVM, KNN, ANN, NB, LR, and DT.

3. METHODS & METHODOLOGY

This study adopts a structured methodology (as shown in Figure 1) aimed at evaluating the impact of different feature selection techniques on the performance of classification models for breast cancer detection. The WBCD comprising of 683 complete instances while 16 instances with missing values were handled through mean average making a total of 699 instances. There are 241 (34.5%) malignant records and 458 (65.5%) benign records. The records were defined using ten features. There are 241 (34.5%) malignant records and 458 (65.5%) benign records. The records were defined using ten features. Data preprocessing included normalization to a standard scale and partitioning into training and testing sets in an 80:20 ratio.

Three feature selection methods were applied to reduce dimensionality and enhance classifier performance: PCA and PCC served as linear approaches, while BNN was implemented for non-linear selection. PCA extracted orthogonal components that captured the highest data variance, whereas PCC filtered out redundant features with high correlation. BNN leveraged connection weights to identify complex, non-linear relationships among features [15].

Six supervised learning algorithms (SVM, KNN, ANN, NB, LR, DT) were each trained using the feature subsets produced by these selection techniques. The models' classification performance was evaluated using standard metrics like accuracy, precision, recall, specificity and f1-Score. Confusion matrices were also used to assess each model's prediction reliability in differentiating malignant from benign cases. MATLAB was used for model training and analysis, while Microsoft Excel was used to validate computations related to PCA and PCC. This comprehensive classification-focused framework enabled a comparative analysis in improving diagnostic accuracy for breast cancer.

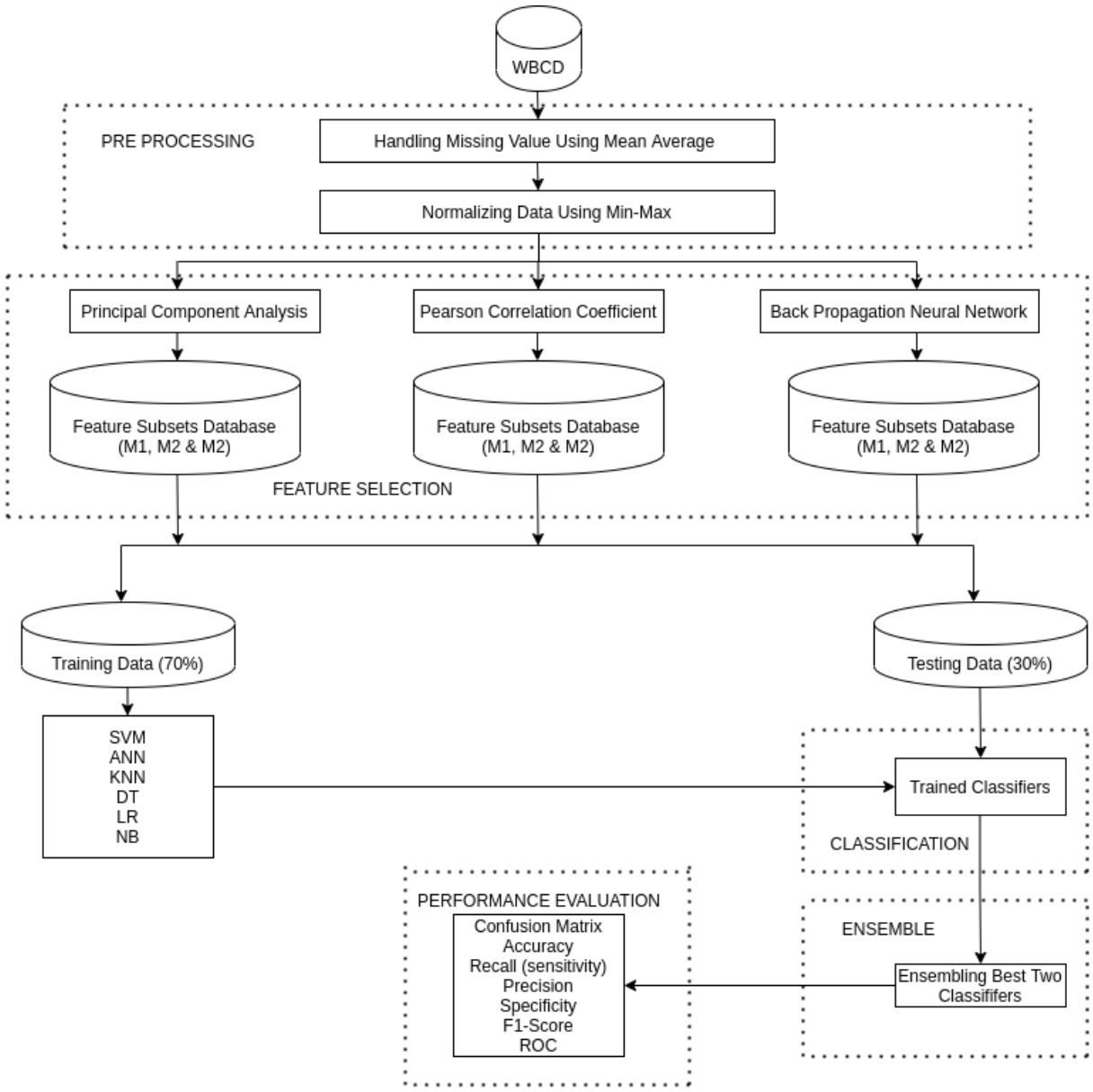


Figure 1: Methodology Framework

3.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful classification algorithm that seeks to find the optimal hyperplane to separate classes in a multidimensional space [16]. Both linear and nonlinear kernel functions were tested to assess the impact of feature selection on SVM performance. Hyperparameters, including the penalty term and kernel-specific parameters, were tuned to achieve optimal classification

outcomes. SVM is well-suited for biomedical applications due to its robustness and strong generalization, particularly in breast cancer detection [17].

3.2. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are capable of learning intricate patterns through interconnected layers of neurons. The model used consisted of an input layer, one or more hidden layers, and an output layer [18]. Training was conducted using the backpropagation algorithm to minimize the prediction error. ANN's flexibility allows it to model both linear and non-linear data relationships, making it a valuable tool for classification in complex datasets like WBCD [19].

3.3. K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) algorithm is a simple, yet effective classification method based on instance-based learning [20]. It classifies test samples by assigning them the majority class among their k closest neighbors, calculated using the Euclidean distance. Cross-validation was used to determine the optimal value of k that maximized classification performance [21]. KNN's performance was analyzed across different feature subsets to assess its sensitivity to feature dimensionality.

3.4. Naive Bayes (NB)

Naive Bayes (NB) is a probabilistic classifier grounded in Bayes' theorem, assuming conditional independence among features [22]. Despite its simplicity, it has shown competitive performance in biomedical classification tasks. In this study, NB was trained to estimate posterior probabilities of class labels, and model parameters were learned using maximum likelihood estimation [21]. Its classification output was evaluated in the context of each feature selection method.

3.5. Logistic Regression (LR)

Logistic Regression (LR) is a widely adopted model for binary classification problems. It maps input features to the probability of belonging to a particular class using the sigmoid function [21]. The model was optimized via gradient descent to determine the most suitable weights. Its role in this study was to serve as a baseline linear classifier for evaluating the influence of feature selection on classification accuracy in breast cancer detection [23].

3.6. Decision Tree (DT)

Decision Tree (DT) classifiers work by recursively splitting the dataset based on feature values to create a hierarchical tree structure [24]. Each node represents a feature test, and each leaf node signifies a predicted class. The Gini index was used as the splitting criterion, and pruning techniques were applied to avoid overfitting [25]. DT's interpretability made it valuable for assessing how different features—selected by each method—impacted classification decisions.

3.7. Hyperparameter Settings

In this study, all classifiers were implemented using their default hyperparameter settings as provided by the MATLAB Classification Learner Toolbox. This approach was adopted to ensure consistency across models and to reflect a baseline performance that can be reasonably expected without extensive tuning. The default configurations have been widely validated in previous literature and offer a practical benchmark for evaluating the impact of different feature selection methods on model performance.

3.8. Evaluation Metrics

To assess the predictive effectiveness of the classification models used for breast cancer diagnosis, a range of standard performance evaluation criteria were employed. These metrics were derived from the confusion matrix, a vital tool in quantifying classification accuracy and errors. The confusion matrix captures the distribution of predicted versus actual labels, offering a granular view of each model's performance [2].

In this study, the confusion matrix was used to record four critical outcomes:

- **True Positive (TP):** Instances correctly identified as malignant (cancerous).
- **False Positive (FP):** Benign (non-cancerous) cases incorrectly predicted as malignant.
- **True Negative (TN):** Benign cases correctly classified as non-cancerous.
- **False Negative (FN):** Malignant cases mistakenly classified as benign.

From these outcomes, five key evaluation metrics were computed to measure classifier performance:

• **Accuracy**

Represents the proportion of total correct predictions out of all predictions made.

$$accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} * 100$$

• **Precision**

Indicates the proportion of correctly predicted malignant cases out of all cases predicted as malignant.

$$precision = \frac{TP}{(TP + FP)} * 100$$

• **Recall (Sensitivity)**

Measures the model's ability to correctly identify malignant cases.

$$recall = \frac{TP}{(TP + FN)} * 100$$

• **Specificity**

Captures the model's ability to correctly classify benign cases.

$$specificity = \frac{(TN)}{(TN + FP)} * 100$$

• **F1-Score (F-Measure)**

Provides a harmonic mean of Precision and Recall, offering a balanced metric especially valuable when classes are imbalanced.

$$f1 - score = 2 * \frac{(\text{Recall} \quad \text{Precision})}{(\text{Recall} + \text{Precision})} *$$

These metrics collectively provided a robust framework to evaluate and compare the 233 effectiveness of different classifiers trained on features selected by both linear and non-linear 234 techniques. They

ensured not only that the models could detect cancer accurately but also that 235 false positives and negatives were minimized, thus enhancing the diagnostic reliability of the 236 machine learning approaches employed [2].

4. RESULTS & DISCUSSION

The descriptive statistical analysis presented in Table 1 offers a detailed summary of the distributional characteristics of ten diagnostic features used in breast cancer evaluation.

Among these variables, CT records the highest mean value (4.42), highlighting its significance in the dataset, while Mitoses (M) has the lowest mean (1.59), reflecting its typically low occurrence in benign cases. BN exhibits the greatest variability, with a standard deviation of 3.62 and a variance of 13.13, suggesting it may play a key role in distinguishing between classes. In contrast, the feature C shows the least variability, supporting its likely categorical nature.

The kurtosis values indicate that M has a sharp peak distribution with notable outliers (12.66), whereas C shows a flatter distribution (-1.58); the remaining features largely approximate normality. Skewness results reveal pronounced right skew in M (3.56) and moderate skew in SECS (1.71) and MA (1.52), suggesting the presence of higher-end outliers.

Most feature values range between 1 and 10, indicating a normalized or scaled dataset, while the class label ranges from 2 to 4, presumably distinguishing benign from malignant cases. These findings point to the need for potential transformations or normalization, particularly for features with high skewness and kurtosis, to improve the performance of classification algorithms that are sensitive to input distributions. This underscores the critical role of preprocessing and thoughtful feature selection in enhancing model effectiveness.

Table 1: Descriptive Statistical Analysis of the parameters

Parameter	Mean	SD	SV	Kurtosis	Skewness	Min	Max
Clump Thickness (CT)	4.42	2.82	7.93	-0.62	0.59	1.00	10.00
Uniformity of Cell Size (UCSZ)	3.13	3.05	9.31	0.10	1.23	1.00	10.00
Uniformity of Cell Shape (UCSH)	3.21	2.97	8.83	0.01	1.16	1.00	10.00
Marginal Adhesion (MA)	2.81	2.86	8.15	0.99	1.52	1.00	10.00
Single Epithelial Cell Size (SECS)	3.22	2.21	4.90	2.17	1.71	1.00	10.00
Bare Nuclei (BN)	3.56	3.62	13.13	-0.79	0.98	1.00	10.00
Bland Chromatin (BC)	3.44	2.44	5.95	0.18	1.10	1.00	10.00
Normal Nucleoli (NN)	2.87	3.05	9.32	0.47	1.42	1.00	10.00
Mitoses (M)	1.59	1.72	2.94	12.66	3.56	1.00	10.00
Class Label (C)	2.69	0.95	0.90	-1.58	0.65	2.00	4.00

4.1. Feature Selection Analysis

FS in this study was conducted using three techniques: PCA, PCC, and BNN. The primary goal of each method was to reduce the dimensionality of WBCD while retaining the most relevant features for classification. PCA transformed the original features into a new set of uncorrelated principal components that captured the highest variance in the data.

PCA effectively reduced the number of features while preserving approximately 95% of the total variance (Table 2). The majority of this variance was captured by the first few components, indicating that essential information was retained within a lower-dimensional representation. This makes PCA particularly well-suited for linear classifiers (Table 3). Specifically, the first seven principal components were selected based on the cumulative explained variance, which accounted for about 95.14% of the total variance. Table 3 shows the resulting feature subsets obtained through PCA.

Table 2: Eigenvalue and percentage of data explained by each factor.

Number	Value	Difference	Proportion	CV	CP
1	**6.70864**	5.91513	0.6709	6.70864	0.6709
2	**0.79352**	0.24635	0.0794	7.50216	0.7502
3	**0.54716**	0.07938	0.0547	8.04933	**0.8049**
4	0.46778	0.08777	0.0468	8.51711	0.8517
5	0.38001	0.06038	0.038	8.89711	0.8897
6	0.31963	0.02199	0.032	9.21675	**0.9217**
7	0.29764	0.03498	0.0298	9.51439	0.9514
8	0.26266	0.128	0.0263	9.77705	0.9777
9	0.13466	0.04636	0.0135	9.91171	0.9912
10	0.0883	---	0.0088	10	1

Table 3: Principal Component Analysis Feature Subsets

SUBSETS (M)	ATTRIBUTES
M1	UCSH, UCSZ, BN
M2	UCSH, UCSZ, BN, BC, CT, NN
M3	UCSH, UCSZ, BN, BC, CT, NN, MA, SECS, M

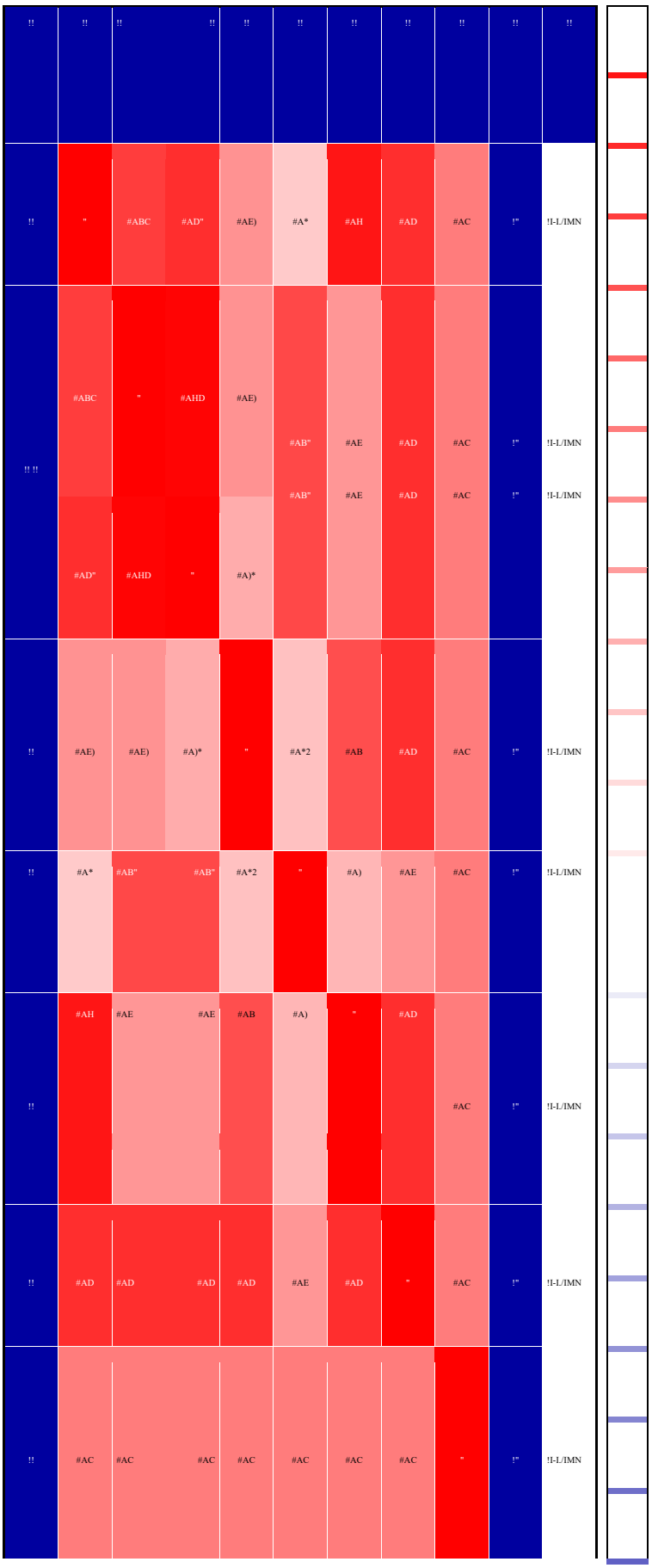
The PCC matrix presented in Table 4 highlights key linear relationships among the diagnostic features and their association with the tumor classification label (C). Among all features, Uniformity of Cell Shape (UCSH) (0.819), Uniformity of Cell Size (UCSZ) (0.818), and Bare Nuclei (BN) (0.813) exhibit the highest correlation with the class label, suggesting they are highly informative for distinguishing between malignant and benign tumors and should be prioritized during feature selection. Other notable features with strong correlations to C include Clump Thickness (CT) at 0.717, Bland Chromatin (BC) at 0.757, and Normal Nucleoli (NN) at 0.7121, indicating their strong diagnostic relevance.

Conversely, Mitoses (M) shows the weakest correlation with C (0.423), implying limited standalone predictive power. The matrix also reveals high inter-feature correlations, particularly between UCSZ and UCSH (0.907), and between UCSZ and Single Epithelial Cell Size (SECS) (0.753), suggesting the presence of multicollinearity. Additional redundancy is observed among SECS, BC, BN, and NN, which are all moderately to strongly correlated with one another and with the target class. This redundancy underscores the risk of overfitting when using highly correlated features in classification models.

Overall, UCSH, UCSZ, and BN stand out as the most predictive features for breast cancer classification. However, the significant multicollinearity among several features emphasizes the need for dimensionality reduction or regularization methods—such as PCA or Lasso regression—to improve model efficiency and interpretability. Figure 2 visualizes the correlation matrix, and Table 5 lists the feature subsets selected based on PCC analysis. Table 4: Correlation analysis between the input and output variables

Parameters	CT	UCSZ	UCSH	MA	SECS	BN	BC	NN	M	C
CT	1									
UCSZ	0.645	1								
UCSH	0.655	0.907	1							
MA	0.487	0.705	0.683	1						
SECS	0.523	0.753	0.720	0.599	1					
BN	0.583	0.685	0.708	0.662	0.579	1				
BC	0.559	0.756	0.736	0.667	0.618	0.674	1			
NN	0.536	0.723	0.719	0.604	0.631	0.580	0.666	1		
M	0.350	0.459	0.439	0.418	0.481	0.337	0.344	0.4281	1	
C	**0.717**	**0.818**	**0.819**	**0.698**	**0.686**	**0.813**	**0.757**	**0.7121**	**0.423**	**1**

"


!!
!!
!!
!!
!!
!!
!!
!!
!!
!!
!!
!!
"
#ABC
#AD"
#AE)
#A*
#AH
#AD
#AC
!"
!!-L/IMN
!! !!
#ABC
"
#AHD
#AE)
#AB"
#AE
#AD
#AC
!"
!!-L/IMN
#AD"
#AHD
"
#A)*
#AB"
#AE
#AD
#AC
!"
!!-L/IMN
!!
#AE)
#AE)
#A)*
"
#A*2
#AB
#AD
#AC
!"
!!-L/IMN
!!
#A*
#AB"
#AB"
#A*2
"
#A)
#AE
#AC
!"
!!-L/IMN
!!
#AH
#AE
#AE
#AB
#A)
"
#AD
#AC
!"
!!-L/IMN
!!
#AD
#AD
#AD
#AD
#AE
#AD
"
#AC
!"
!!-L/IMN
!!
#AC
#AC
#AC
#AC
#AC
#AC
#AC
"
!"
!!-L/IMN

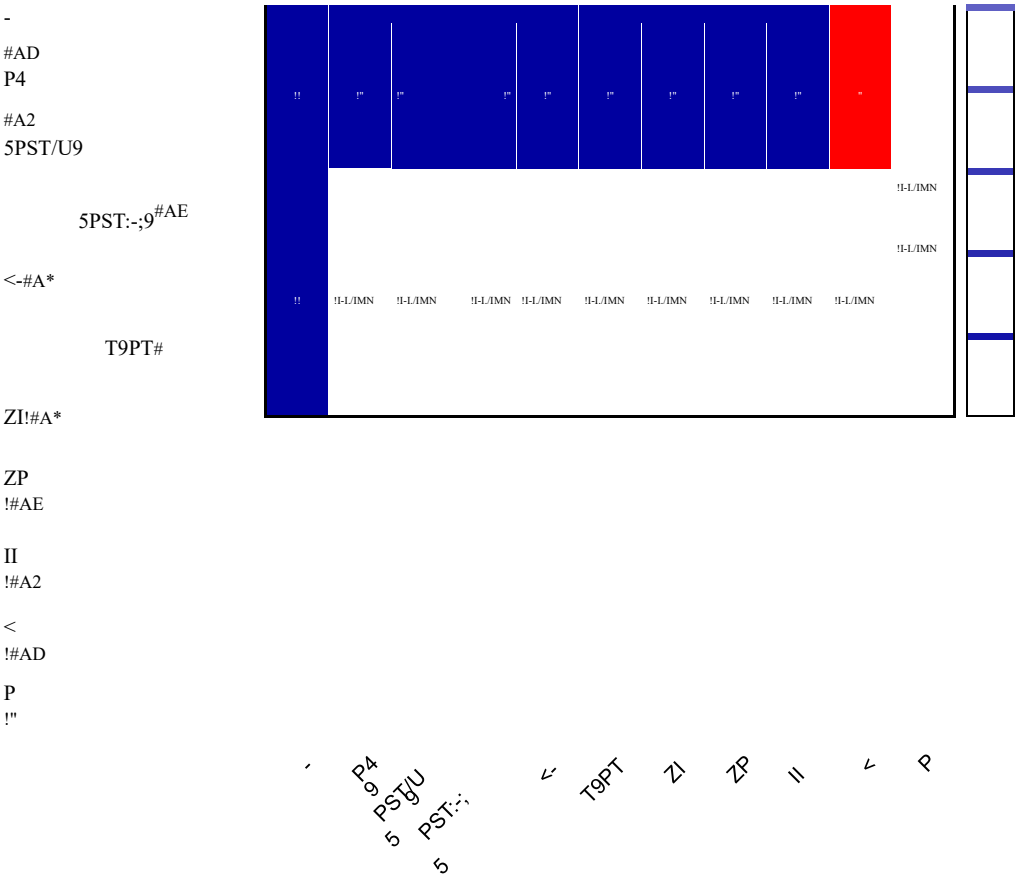


Figure 2: Pearson Correlation Coefficient analysis between the input and output variables

Table 5: Pearson Correlation Coefficient Feature Subsets

SUBSETS (M)	ATTRIBUTES
M1	UCSH, UCSZ, BN
M2	UCSH, UCSZ, BN, BC, CT, NN
M3	UCSH, UCSZ, BN, BC, CT, NN, MA, SECS, M

The BNN utilized for feature selection was constructed using a fully connected feedforward architecture. This network includes an input layer, two hidden layers, and an output layer. The first and second hidden layers contain 32 and 16 neurons respectively, both employing the ReLU (Rectified Linear Unit) activation function. The output layer uses a sigmoid activation function to support binary classification. The model was trained using the binary cross-entropy loss function, optimized with the Adam optimizer, and configured with a learning rate of 0.001. Training was conducted over 100 epochs with a batch size of 32. To prevent overfitting, early stopping was applied based on validation loss. This configuration was selected after preliminary tuning to balance training efficiency and predictive accuracy.

The feature importance analysis, based on Root Mean Square Error (RMSE), ranks the diagnostic attributes according to their contribution to the model's predictive performance (Table 6). Uniformity of Cell Size (UCSZ) achieved the lowest RMSE (0.2276), indicating its strong influence on classification accuracy. It was followed closely by Uniformity of Cell Shape (UCSH) with an RMSE of 0.238 and Bland Chromatin (BC) at 0.279, all of which align with previously identified strong linear correlations and suggest their central role in distinguishing between benign and malignant tumors.

In contrast, features such as Mitoses (M), Bare Nuclei (BN), and Clump Thickness (CT) yielded higher RMSE values of 0.404, 0.336, and 0.3231 respectively, indicating weaker contributions to the BNN's predictive power. The relatively low importance of M further corroborates earlier findings from the descriptive statistics and correlation analysis, suggesting it is less informative for classification tasks.

To ensure statistical reliability and robustness of the RMSE-based rankings, a 10-fold cross-validation strategy was implemented during model training and evaluation. The reported RMSE values in Table 6 represent the mean ± standard deviation across all folds. For instance, the lowest RMSE was recorded as 0.238 ± 0.021, while the highest reached 0.402 ± 0.026. Despite the seemingly narrow range, these results consistently appeared across different data partitions, confirming the BNN's stability and the statistical significance of feature importance rankings. In summary, the BNN effectively captured complex, non-linear relationships within the dataset and prioritized features like UCSZ, UCSH, and BC as the most influential for breast cancer classification. These results validate the efficacy of BNN-based feature selection for enhancing model performance and highlight its potential in medical diagnostic applications.

Table 6: Back Propagation Neural Network Feature Subsets

Attributes	RMSE	Ranking
UCSZ	0.2276	1
UCSH	0.238	2
BC	0.279	3
SECS	0.2891	4
NN	0.3053	5
MA	0.3216	6
CT	0.3231	7
BN	0.336	8
M	0.404	9

Table 7: Back Propagation Neural Network Feature Subsets

SUBSETS (M)	ATTRIBUTES
M1	UCSZ, UCSH, BC
M2	UCSZ, UCSH, BC, SECS, NN, MA
M3	UCSZ, UCSH, BC, SECS, NN, MA, CT, BN, M

4.2. Comparative Analysis of all Classifier Performance

After evaluating six classification models (SVM, KNN, LR, DT, NB, and ANN), SVM models (Table 8) demonstrated robust performance across all configurations. The highest classification accuracy of 97.3% was achieved using a Quadratic kernel on the non-linear M3 subset, while the best linear performance was 96.5% with the Medium Gaussian kernel on M2. These results emphasize SVM's ability to model complex boundaries, particularly when supported by non-linear feature transformations. Overall, SVM consistently ranked among the top-performing classifiers, underscoring its effectiveness in high-dimensional medical datasets.

KNN classifiers also performed notably well as seen on table 9. The Cosine and Weighted variants reached 97.1% and 96.9% accuracy, respectively, on the non-linear M3 subset—surpassing the best linear result of 96.3% on M2/M3. These outcomes affirm the model's sensitivity to both the choice of distance metric and the feature representation. KNN's performance further highlights its capacity to adapt effectively to local patterns in transformed feature spaces.

As a linear classifier, LR (Table 10) achieved a solid accuracy of 95.7% on the linear M2 subset, with a modest increase to 96.4% using the non-linear M3 subset. While LR did not outperform more complex models like SVM or KNN, its high interpretability and relatively competitive performance make it an attractive option for clinical settings where model transparency is essential.

DT models showed stable but modest performance (Table 11), with accuracy peaking at 94.8% using the Medium variant on the non-linear M1 subset, and 94.7% with the Fine and Medium variants on linear M1. This relatively consistent performance across feature types suggests limited sensitivity to feature transformations but also highlights a potential for overfitting. The results indicate that DTs may benefit more from ensemble methods such as Random Forest or boosting to enhance generalizability.

NB classifiers exhibited clear improvement when applied to non-linear feature subsets (Table 12). The Kernel NB model achieved 96.6% on non-linear M3, surpassing the 95.7% achieved with the Gaussian variant on linear M3. These results confirm that NB can benefit from enriched feature representations, particularly when independence assumptions are relaxed by more expressive feature selection strategies.

ANN classifiers maintained consistently high performance across both feature categories as seen on table 13. The best result was 96.3% on the linear M3 subset, followed closely by 96.1% on the non-linear M3. This close parity indicates ANN's flexibility in learning both linear and non-linear

relationships, supporting its application in complex medical classification problems where pattern recognition across diverse input spaces is required.

Across all classifiers, the non-linear M3 subset consistently produced the highest accuracies, particularly for SVM (97.3%), KNN (97.1%), and NB (96.6%). This trend underscores the importance of aligning the classifier architecture with the nature of the feature transformations. Models capable of exploiting complex, high-order relationships such as SVM and KNN—demonstrated improved generalization when coupled with non-linear features. SVM and KNN emerged as the top performers across both feature categories, suggesting their suitability for breast cancer detection tasks involving heterogeneous or high-dimensional data. ANN offered a strong balance between accuracy and adaptability, while LR and DT, though relatively simpler, provided competitive performance and retained their appeal for interpretable or resource-constrained applications.

Table 8: Accuracy results of SVM classifiers using linear and non-linear FS

	LINEAR FS (%)			**NON-LINEAR FS** (%)		
SVM	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**
Linear	94.6	96	95.9	94.4	95.6	96.6
Quadratic	95.3	96	96.4	95.6	95.7	97.3
Cubic	95.7	95.9	95.4	95.3	95.7	96.3
Fine Gaussian	94.7	95	94.7	95	95.7	94.1
Medium Gaussian	95.3	95.9	95.9	95	95.9	96.7
Coarse Gaussian	94.8	95.7	95.9	94.6	95.3	96.9

Table 9: Accuracy results of KNN classifiers using linear and non-linear FS

KNN	LINEAR FS (%)			NON-LINEAR FS (%)		
	M1	M2	M3	M1	M2	M3
FINE	94.4	94.1	93.4	93	94	95.1
MEDIUM	94.8	95.7	95.7	95	95.3	96.7
COARSE	93.7	94	93.7	94.6	93.3	95.3
COSINE	94.8	96.3	96.3	95.4	96	97.1
CUBIC	94.7	96.1	95.3	95	95.4	96.7
WEIGHTED	95.3	95.9	95.9	94.3	96	96.9

Table 10: Accuracy results of LR classifiers using linear and non-linear FS

LR	LINEAR FS (%)			NON-LINEAR FS (%)		
	M1	M2	M3	M1	M2	M3
LOGISTIC REGRESSION	94.3	95.7	95.6	94.4	95.1	96.4

Table 11: Accuracy results of DT classifiers using linear and non-linear FS

DT	LINEAR FS (%)			NON-LINEAR FS (%)		
	M1	M2	M3	M1	M2	M3
FINE	94.7	94.1	94.4	94.7	93.8	93.4
MEDIUM	94.7	94.3	94.4	94.8	94	93.6
COARSE	94.6	93.6	93.6	94.3	93.3	93.3

Table 12: Accuracy results of NB classifiers using linear and non-linear FS

NB	LINEAR FS (%)			NON-LINEAR FS (%)		
	M1	M2	M3	M1	M2	M3
GUASSIAN	94.6	95.7	95.4	94.8	95.6	95.7
KERNEL	92.4	95.3	95.7	94.3	94.6	96.6

Table 13: Accuracy results of ANN classifiers using linear and non-linear FS

ANN	LINEAR FS (%)			NON-LINEAR FS (%)		
	M1	M2	M3	M1	M2	M3
ARTIFICIAL NEURAL NETWORK	95.6	96	96.3	95.7	95.3	96.1

4.3 Performance Evaluation Analysis of Best Classification Models

A detailed comparative analysis of the best performing under linear and non-linear FS shows clear performance differences once Accuracy, Precision, Recall, Specificity, and F1Score are considered. SVM emerges as the leading model: with a Quadratic kernel trained on the non-linear M3 feature subset, it attains the highest accuracy of 97.3 %, coupled with 94.8 % precision, 97.5 % recall, 97.2 % specificity, and a 96.1 % F1-Score. Even its best linear configuration, a Medium Gaussian kernel on M2, maintains a strong 96.5 % accuracy, underlining SVM's robustness on both feature types.

KNN follows closely. Using a Cosine distance metric on the non-linear M3 subset, KNN achieves 97.1 % accuracy, with recall and F1-Score matching SVM's recall (97.5 %) and reaching 95.9 %, respectively; its linear counterpart records a still-impressive 96.3 % accuracy. LR, while inherently linear, rises to 96.4 % accuracy and a 94.9 % F1-Score on the non-linear M3 subset, offering a transparent alternative for clinical contexts where interpretability is paramount. DT records the lowest accuracies of 94.8 % for non-linear features and 94.7 % for linear ones with a comparatively modest F1-Score of 92.7 %, reflecting higher false-positive rates and potential overfitting. NB benefits markedly from the non-linear Kernel variant, reaching 96.6 % accuracy, 93.2 % precision, and 97.1 % recall, outperforming its Gaussian counterpart and demonstrating that richer feature representations mitigate the model's strong independence assumptions. ANN remains consistently competitive: it peaks at 96.3 % accuracy on linear M3 and maintains 96.1 % on the corresponding non-linear subset, confirming its adaptability to varying feature structures.

Overall, SVM with the Quadratic kernel on M3 stands out as the best performer across all metrics, while KNN provides the most balanced alternative, virtually matching SVM's recall and F1 but at a

slight cost in precision. LR offers a light yet respectable option, NB shows the greatest benefit from non-linear features, and ANN delivers stable high performance regardless of feature type. DT, though acceptable, trail the other models. These results underscore the need to align classifier architecture with the complexity of feature transformations: non-linear subsets, particularly M3, consistently boost performance and should be favored when high diagnostic accuracy in breast cancer detection is required.

Table 14: Performance Evaluation Metrics for Best Classification Results of Linear and Nonlinear Feature Sets (M1–M3)

Model	FS (Subset)	Accuracy (%)	TP	TN	FP	FN	Precision	Recall	Specificity	F1-Score
SVM - Medium Gaussian	Linear (M2)	96.5	234	441	17	7	93.2	97.1	96.3	95.1
SVM - Quadratic	Non-Linear (M3)	97.3	235	445	13	6	94.8	97.5	97.2	96.1
KNN - Cosine	Linear (M2/M3)	96.3	233	440	18	8	92.8	96.7	96.1	94.7
KNN - Cosine	Non-Linear (M3)	97.1	235	444	14	6	94.4	97.5	96.9	95.9
LR	Linear (M2)	95.7	232	437	21	9	91.7	96.3	95.4	93.9
LR	Non-Linear (M3)	96.4	233	441	17	8	93.2	96.7	96.3	94.9
DT - Fine & Medium	Linear (M1)	94.7	230	432	26	11	89.8	95.4	94.3	92.6
DT – Medium	Non-Linear (M1)	94.8	230	433	25	11	90.2	95.4	94.5	92.7
NB - Gaussian	Linear (M2)	95.7	232	437	21	9	91.7	96.3	95.4	93.9
NB - Kernel	Non-Linear (M3)	96.6	234	441	17	7	93.2	97.1	96.3	95.1
ANN	Linear (M3)	96.3	233	440	18	8	92.8	96.7	96.1	94.7
ANN	Non-Linear (M3)	96.1	233	439	19	8	92.5	96.7	95.9	94.5

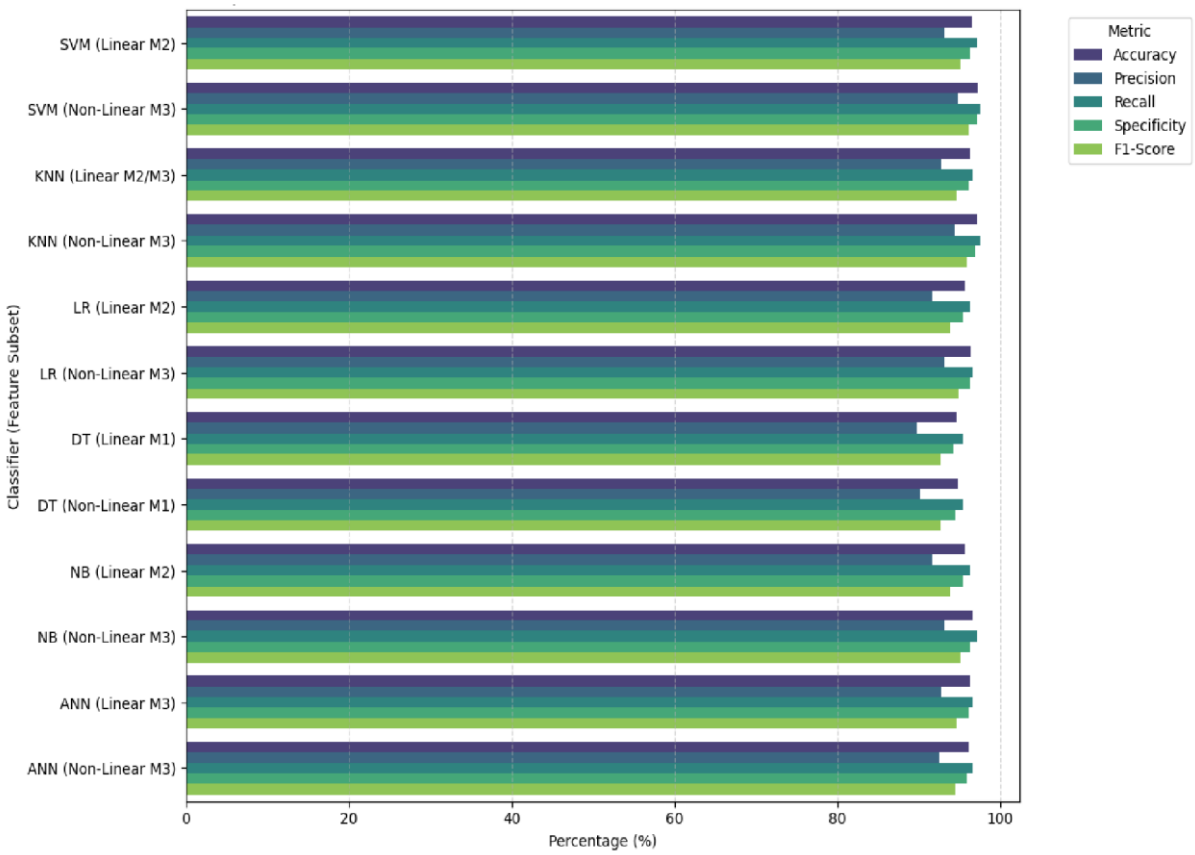


Figure 3: Comparative Performance of Best Classifiers with Linear and Non-Linear Feature Subsets

5. CONCLUSION

In summary, this study compared six widely used machine-learning classifiers (SVM, KNN, LR, DT, NB, and ANN) on the WBCD after both linear (M1–M3) and non-linear (BNNderived) feature transformations. Results show that non-linear features substantially improve diagnostic performance, with SVM employing a Quadratic kernel on the M3 subset achieving the highest accuracy, precision, recall, specificity, and F1-score. KNN (Cosine, M3) closely followed, while ANN delivered consistently strong results across feature types. LR offered a highly interpretable yet competitive alternative, NB benefited most from non-linear kernels, and DT lagged behind, suggesting susceptibility to overfitting. These findings confirm that pairing appropriate classifiers with richer, non-linear feature representations is essential for maximizing breast-cancer detection accuracy.

Despite these promising outcomes, several limitations should be acknowledged. First, the analysis relied on a single public dataset; therefore, class distribution, imaging protocols, and demographic diversity may not fully reflect real-world clinical variability. Second, the study did not explore ensemble methods, cost-sensitive learning, or cross-validated hyperparameter optimization, each of which could further enhance performance. Third, while confusion-matrix-derived metrics were examined, additional measures such as area under the ROC curve and calibration were beyond scope. Finally, computational cost and explainability were not systematically quantified, yet they are critical for clinical deployment.

Future work should validate the top-performing models on larger, multi-institutional datasets; incorporate ensemble or hybrid approaches to reduce variance and bias; and employ explainable-AI techniques so clinicians can better trust automated assessments. Investigating class-imbalance handling, real-time inference efficiency, and integration with complementary data sources will further strengthen the reliability and utility of machine-learning tools for early breast-cancer diagnosis.

Acknowledgments: We extend our heartfelt gratitude to Baze University Abuja, Nigeria for their funding and support, Humans and Machines Limited for technical assistance, and the creators of the Wisconsin Breast Cancer Dataset for their contributions. Special thanks to our academic advisors, and peers for fostering insightful discussions and providing a platform to share our findings.

Author Contributions: H.M. conceived the study, collected and preprocessed the data, performed the experiments, and wrote the manuscript.

Additional Information

Competing Interests Statement: The author declares no competing financial or nonfinancial interests.

6. REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[2] S. Laghumati, B. Cherradi, A. Tmiri, O. Danouni, and S. Hamida, "Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques," in *Proc. 3rd Int. Conf. Advanced Communication Technologies and Networking (CommNet)*, 2020, doi: 10.1109/CommNet49926.2020.9199633.

[3] M. Ragab, A. Sharkas, and O. Attallah, "Breast cancer diagnosis using an efficient CAD system based on multiple classifiers," *Diagnostics*, vol. 9, no. 4, p. 165, 2019, doi: 10.3390/diagnostics9040165.

[4] R. Turkki, et al., "Breast cancer outcome prediction with tumour tissue images and machine learning," *Breast Cancer Research and Treatment*, vol. 177, no. 1, pp. 41–52, 2019, doi: 10.1007/s10549-019-05281-1.

[5] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microwave Magazine*, vol. 12, no. 7, pp. 78–94, 2011, doi: 10.1109/MMM.2011.942702.

[6] C. Erickson, "Automated detection of breast cancer using SAXS data and wavelet features," Jul. 2005.

[7] M. Madadi, S. Zhang, and L. M. Henderson, "Evaluation of breast cancer mammography screening policies considering adherence behavior," *Eur. J. Oper. Res.*, vol. 247, no. 2, pp. 630–640, 2015, doi: 10.1016/j.ejor.2015.05.068.

[8] M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in *Proc. Int. Conf. Computer and Applications (ICCA)*, 2017, pp. 306–311, doi: 10.1109/COMAPP.2017.8079784.

[9] M. Abdar and V. Makarenkov, "CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer," *Measurement*, vol. 146, pp. 557–570, 2019, doi: 10.1016/j.measurement.2019.05.022.

[10] E. M. F. El Houby, "A survey on applying machine learning techniques for management of diseases," *J. Appl. Biomed.*, vol. 16, no. 3, pp. 165–174, 2018, doi: 10.1016/j.jab.2018.01.002.

[11] P. Zarbakhsh and A. Addeh, "Breast cancer tumor type recognition using graph feature selection technique and radial basis function neural network with optimal structure," *J. Cancer Res. Ther.*, vol. 14, no. 3, p. 625, 2018, doi: 10.4103/0973-1482.183561.

[12] R. Dhanya, I. R. Paul, S. S. Akula, M. Sivakumar, and J. J. Nair, "F-test feature selection in Stacking ensemble model for breast cancer prediction," *Procedia Comput. Sci.*, vol. 171, pp. 1561–1570, 2020, doi: 10.1016/j.procs.2020.04.167.

[13] M. Hosni, I. Abnane, A. Idri, J. M. Carrillo de Gea, and J. L. Fernández Alemán, "Reviewing ensemble classification methods in breast cancer," *Comput. Methods Programs Biomed.*, vol. 177, pp. 89–112, 2019, doi: 10.1016/j.cmpb.2019.05.019.

[14] W. Wolberg, "UCI Machine Learning Repository: Breast Cancer Coimbra Data Set," Univ. California, 1992. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29

[15] D. R. Wilson and T. R. Martinez, "Reduction techniques for instance-based learning algorithms," *Mach. Learn.*, vol. 38, no. 3, pp. 257–286, 2000, doi: 10.1023/A:1007626913721.

[16] S. R. Gunn, "Support Vector Machines for Classification and Regression," Univ. Southampton, 1998. [Online]. Available: https://www.isis.ecs.soton.ac.uk/resources/svm/

[17] I. Maglogiannis, E. Zafiropoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM-based classifiers," *Appl. Intell.*, vol. 30, no. 1, pp. 24–36, 2009, doi: 10.1007/s10489-007-0073-z.

[18] S. Walczak and N. Cerpa, "Artificial neural networks," in *Encyclopedia of Physical Science and Technology*, Elsevier, 2003, pp. 631–645, doi: 10.1016/B0-12-227410-5/008371.

[19] N. Fatima, L. I. Liu, S. H. A. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[20] A. Mert, N. Kiliç, E. Bilgili, and A. Akan, "Breast cancer detection with reduced feature set," *Comput. Math. Methods Med.*, vol. 2015, Article ID 265138, 2015, doi: 10.1155/2015/265138.

[21] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast tumor classification using an ensemble machine learning method," *J. Imaging*, vol. 6, no. 6, 2020, doi: 10.3390/jimaging6060039.

[22] S. A. S. Al-Sabbah, S. F. Mohammad, and M. M. Eanad, "Use of the Naïve Bayes function and the models of artificial neural networks to classify some cancer tumors," *Indian J. Public Health Res. Dev.*, vol. 10, no. 4, pp. 1563–1569, 2019, doi: 10.5958/09765506.2019.00938.0.

[23] M. Sumner, E. Frank, and M. Hall, "Speeding up logistic model tree induction," in *Lect. Notes Comput. Sci.*, vol. 3721, pp. 675–683, 2005, doi: 10.1007/11564126_72.

[24] L. Rokach, "Decision forest: Twenty years of research," *Inf. Fusion*, vol. 27, pp. 111– 125, 2016, doi: 10.1016/j.inffus.2015.06.005.

[25] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Elsevier, 535 2012, doi: 10.1016/C2009-0-61819-5.