International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

"Ensemble Machine Learning Models for High-Dimensional Biomedical Data using PCA-Based Feature Reduction"

Dr. Rashmi Gupta[1], Dr. Akhilesh Shrivas[2], Mansi Singhania[3]
[1,2,3]*Ghasidas Vishwavidyalaya (A Central University), Bilaspur, 495001, India*

Abstract— **Breast cancer remains one of the leading causes of mortality among women worldwide, and early diagnosis plays a critical role in improving treatment outcomes. Machine learning techniques have emerged as powerful tools for medical decision support, especially when combined with dimensionality-reduction methods such as Principal Component Analysis (PCA). This study evaluates the performance of multiple classification algorithms—C4.5, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Neural Networks—on the Breast Cancer Wisconsin dataset comprising 30 diagnostic features. PCA was applied to reduce feature dimensionality while retaining 99% of data variance using 17 principal components. Performance was assessed using accuracy, precision, recall, and F1-score for models trained with and without PCA. The results indicate that PCA enhances the performance of several classifiers, particularly in terms of recall and overall robustness. The hybrid ensemble model (C4.5 + SVM + Neural Network) achieved the highest accuracy of 0.9912 and perfect recall of 1.0000 with PCA, outperforming individual classifiers.**

Keywords— **Machine learning; Principal Component Analysis (PCA); Feature reduction; Classification techniques; Support Vector Machine (SVM); Decision Tree (C4.5); Random Forest; K-Nearest Neighbors (KNN); Neural Network; Ensemble learning.**

I. INTRODUCTION

Breast cancer is one of the most prevalent malignant diseases affecting women globally, and its early detection plays a crucial role in reducing mortality and improving patient survival. Medical datasets used for breast-cancer diagnosis often contain numerous correlated features that can complicate the learning process of classification algorithms (Agarap, 2017). High-dimensional feature spaces may lead to issues such as over-fitting, increased computational cost, and reduced generalization performance. Therefore, dimensionality-reduction techniques have become essential in building efficient and accurate diagnostic systems.

Principal Component Analysis (PCA) is one of the most widely used statistical techniques for reducing dimensionality by transforming correlated variables into a smaller set of orthogonal components that retain most of the original variance (Rajaguru & SR, 2019; Omondiagbe et al., 2019). In this study, PCA was applied to the Breast Cancer dataset consisting of 30 input features, with 17 components preserving 99% of the total variance. The objective is to analyze how PCA influences the performance of machine-learning classifiers commonly used in medical diagnosis.

Several machine-learning models—including C4.5, SVM, Random Forest, KNN, and Neural Networks—were evaluated with and without PCA to determine improvements in diagnostic accuracy and computational efficiency. The results demonstrate that PCA has a positive impact on most classifiers, especially by improving recall (sensitivity), which is critical in medical diagnosis to reduce false negatives. Notably, the hybrid ensemble model using C4.5, SVM, and Neural Network achieved the highest performance, with an accuracy of 0.9912 and perfect recall with PCA, indicating its strong potential for reliable breast-cancer detection.

Overall, this research highlights the importance of dimensionality reduction in designing robust and accurate diagnostic models and demonstrates that PCA can significantly enhance classification performance in breast-cancer prediction tasks.

II. REVIEW OF LITERATURE

Breast cancer detection has been a major research focus in the medical imaging and biomedical data analysis domain. Numerous machine learning and statistical techniques have been applied to improve diagnostic accuracy and reduce the risk of false negatives.

In the study by Rajaguru and SR [2019], to pick features for the BCD dataset, PCA was used. The selected features were used to train and test a Decision Tree and KNN.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

With a 90.44 percent Mathews Correlation Coefficient, 95.61 percent accuracy, and 95.95 percent sensitivity, the KNN classifier surpassed the Decision Tree in every statistic. Cross-validation was not performed on the KNN and Decision Tree models. Over-fitting and sampling bias may therefore have an impact on performance. A similar study by Saoud et al. [21] utilized the most effective first search approach for feature selection and wrapper models, such as artificial neural networks (ANN), Bayesian networks, SVM, K-NN, Decision Trees, and Logistic Regression (LR). Different numbers of features were chosen by each model from the BCW dataset. The comparative examination of model performance was done by comparing accuracy measures for both models—one with feature selection and the other without. With an accuracy of 97.36 percent, the SVM model without feature selection fared better than the other models.

Omondiagbe et al. [22] were able to identify breast cancer with 98.82 percent accuracy 98.41 percent sensitivity, and 99.07 percent specificity, using the BCW data set with SVM, radial basis kernel, ANN, and Naive Bayes.

Kumar et al. [23] discovered that, when utilizing the WCB dataset, PCA and K-NN had a 96.4 percent accuracy rate in identifying breast cancer. Recently, two distinct datasets related to breast cancer were used to investigate the efficacy of K-NN utilizing different distance functions and k values. Studies are using K-NN, linear SVM, and Chi-squared features without feature selection.

Zohaib et al. [3] used five different popular supervised learning classifiers. PCA based five separate techniques are applied on each classifier. These are Liner, Cosine, Poly, Sigmoid and RBF kernel based PCA procedure. Each classifier compared with other on the basis of performance metrics. It found that, Sigmoid based Naïve Bayes exhibits best accuracy of 99.20%. K Nearest Neighbor also illustrates superb performance with all kernel PCA based techniques. Accuracy ranges from 96.4% to 97.8%.

The study presented by Sahu et al., [7] used hybrid approach that combines feature selection, advanced machine learning techniques, to improve the accuracy of breast cancer prediction. The proposed method uses PCA and LASSO for dimensionality reduction, MLP for feature extraction, and different machine learning techniques for classification, focusing on reducing data dimensionality and improving model generalization. This work demonstrates the efficacy of combining dimensionality reduction, ensemble learning, and Optuna optimization, validated through 3-fold cross-validation, offering a novel and reliable approach to interpretable breast cancer detection systems.

Archana et al. uses classification models in research that are regularized with Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) in an attempt to manage and minimize the dimensionality of certain datasets. Some of the learning algorithms that are used include SVM, Random forest and the k-NN algorithms with the selected features being used to train and test the various algorithms.

Many researchers have investigated machine-learning approaches for breast-cancer diagnosis. Earlier studies highlight that high-dimensional medical data often degrade classifier performance. PCA is one of the most widely used dimensionality-reduction techniques in cancer diagnosis research, as it improves computational speed and removes redundancy (Rajaguru & SR, 2019; Wang, 2023).

Several studies have reported that SVM and Neural Networks provide high accuracy in medical classification tasks. Ensemble approaches have also gained importance for improving reliability by combining strengths of multiple models (Li et al., 2023; Sahu & Fatma, 2025). However, fewer studies have performed a comprehensive comparison of multiple classifiers both **with and without PCA** on the same dataset.

This research fills that gap by comparing five well-known classifiers and one hybrid ensemble under controlled experimental conditions, evaluating whether PCA genuinely improves performance across the models.

III. MATERIALS AND METHODS

3.1.1 Dataset Description

The study utilizes a benchmark Breast Cancer Diagnostic Dataset consisting of **30 continuous numerical features** that describe various physical characteristics of cell nuclei obtained from digitized breast mass images. These features include measures such as radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension. Each instance in the dataset is labeled as either **benign** or **malignant**, enabling supervised machine learning classification. The dataset is widely used in medical data mining research due to its reliability, structured format, and effectiveness in evaluating diagnostic algorithms. For the purpose of this analysis, the dataset served as the primary input source for both pre-PCA and post-PCA classification experiments, with **all 30 features normalized** prior to further processing and dimensionality reduction.

3.1.2 System flow of the proposed method

The proposed method begins with the collection of a breast cancer dataset containing clinical and diagnostic features. The data is then preprocessed to handle missing values, normalize features, and encode categorical variables, ensuring it is suitable for classification. Principal Component Analysis (PCA) is applied to reduce the dimensionality of the dataset while retaining the most significant variance. The dataset is subsequently divided into training and testing sets to enable model evaluation. Six classification techniques—C4.5, SVM, Random Forest, KNN, Neural Network, and a Hybrid Ensemble model—are trained and tested on both the original and PCA-reduced datasets.

Performance is assessed using metrics such as accuracy, precision, recall, F1-score, and confusion matrices. The results of models with and without PCA are compared to evaluate the impact of feature reduction. Among the classifiers, the Hybrid Ensemble model demonstrates the best overall performance in terms of accuracy and robustness. Finally, the analysis highlights the benefits of PCA in improving computational efficiency and predictive performance. Fig 1. Shows the propose flow diagram for breast cancer diagnosis using PCA-based dimensionality reduction and machine learning classification techniques with comparative performance evaluation.

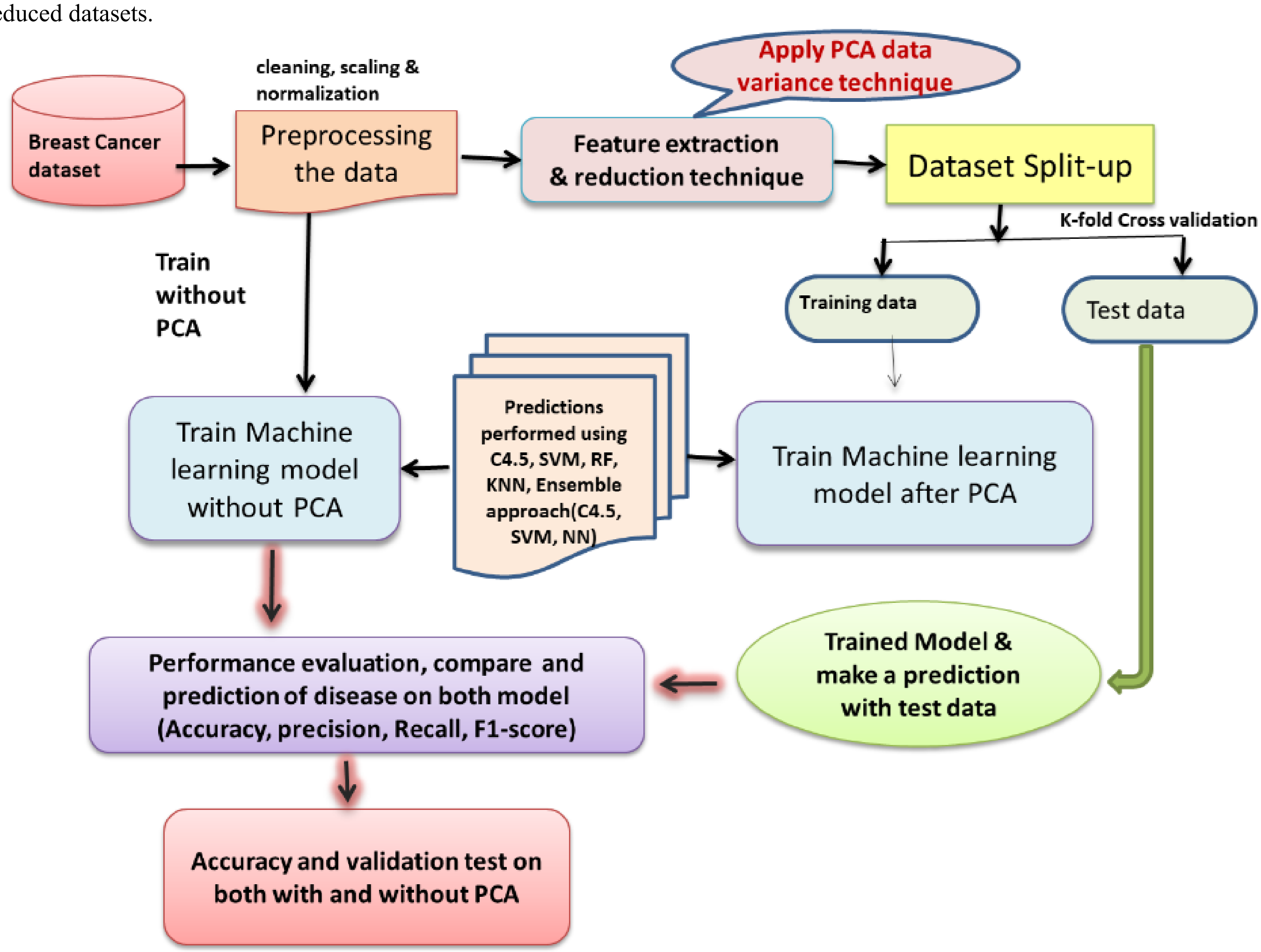

Fig 1 Learning pipeline model using classification algorithm

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

3.1.3 Software and Tools Used

The experimental analysis was carried out using the Python programming language, which provided an efficient and flexible environment for implementing machine learning algorithms and dimensionality reduction techniques. The study utilized essential Python libraries including NumPy for numerical computations, Pandas for dataset handling and preprocessing, Scikit-learn for applying PCA, implementing classification models such as SVM, Decision Trees, Random Forest, KNN, and Neural Networks, and for calculating performance metrics including accuracy, precision, recall, and F1-score. Additionally, Matplotlib were used for plotting and visualizing comparative results. All experiments were executed on a standard computing system using the Python ecosystem, ensuring reproducibility and ease of model evaluation (Rajaguru & SR, 2019).

3.2 Methodology

3.2.1 Data Preprocessing

Before applying the machine learning models, the dataset underwent essential preprocessing procedures. The dataset was first examined for missing values and inconsistencies, and it was confirmed that all records were complete with no missing data. To ensure uniform contribution of all variables and to improve model convergence, the 30 numerical features were standardized using Z-score normalization, which is commonly applied in machine learning workflows involving distance-based and variance-sensitive algorithms (Agarap, 2017; Mushtaq et al., 2019). This preprocessing step is particularly important for improving the effectiveness of PCA and classifiers such as KNN and SVM. After normalization, the dataset was divided into training and testing subsets using an 80:20 split ratio, a widely adopted practice to ensure unbiased and reliable performance evaluation (Omondiagbe et al., 2019).

3.2.2 Principal Component Analysis (PCA)

3.2.3 Classification Models

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset while retaining the most significant information. PCA works by transforming the original set of thirty correlated features into a smaller set of uncorrelated components, known as principal components (Mushtaq et al., 2019, Nirmala et al., 2020). In this study, the covariance matrix of the standardized data was computed, followed by Eigen value–eigenvector decomposition to identify the directions of maximum variance.

The components were then ranked according to their Eigen -values, and only those contributing to 99% of the total variance were selected. This procedure reduced the dataset to 17 principal components, effectively minimizing redundancy and noise while preserving the essential structure needed for accurate classification. The transformed PCA features were subsequently used as input for all machine learning models to evaluate improvements in performance and computational efficiency.

C4.5 Decision Tree: The C4.5 decision tree is a widely used supervised learning algorithm that constructs a tree-like model of decisions based on input features. It splits nodes using the information gain ratio, which helps to select the feature that best separates the data at each step. This method is highly interpretable because the resulting tree structure clearly shows the decision-making process, making it easier for researchers and practitioners to understand how predictions are made. Additionally, C4.5 is computationally efficient, which allows for fast training even with moderately large datasets. In this study, C4.5 serves as a baseline model to evaluate classification performance both with and without dimensionality reduction via PCA.

Support Vector Machine (SVM): Support Vector Machine is a powerful supervised learning technique that aims to find an optimal hyper-plane that separates data points of different classes with maximum margin. It is particularly effective for binary classification tasks, such as breast cancer diagnosis, where distinguishing between malignant and benign cases is critical. The SVM is robust to high-dimensional data, and it works well even when the number of features exceeds the number of samples. In this research, SVM is used to assess how PCA-based feature reduction affects the model's ability to generalize and correctly classify instances.

Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees using bootstrapped samples of the training data. Each tree makes an independent prediction, and the final output is determined through majority voting. This approach enhances predictive accuracy while reducing the risk of over-fitting, which is a common problem in single decision trees. Random Forest is particularly useful for datasets with complex interactions among features. In the present study, it provides a comparative benchmark to evaluate the effectiveness of PCA in improving ensemble-based classification performance.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

K-Nearest Neighbors (KNN): The KNN algorithm classifies a data point based on the majority class of its k nearest neighbors in the feature space, using a distance metric such as Euclidean distance. KNN is sensitive to the scale of features, so normalization is performed to ensure that all features contribute equally to distance calculations. While simple in concept, KNN can be computationally intensive for large datasets because it requires calculating distances to all training instances. In this work, KNN is implemented to examine the impact of PCA on instance-based learning and its influence on classification accuracy.

Neural Network (NN): The neural network used in this study is a feed-forward multilayer perceptron trained using the back-propagation algorithm. It consists of an input layer, one or more hidden layers, and an output layer. Neural networks are capable of learning complex, nonlinear relationships in data, making them suitable for challenging classification problems where linear models may fail. The network's performance depends on factors such as the number of hidden neurons, learning rate, and activation functions. In the current study, the neural network serves to evaluate how dimensionality reduction through PCA affects the ability of a nonlinear model to accurately classify breast cancer cases.

Hybrid Ensemble Model (C4.5 + SVM + NN): The hybrid ensemble model combines the predictions of C4.5, SVM, and neural network classifiers using majority voting. This approach leverages the strengths of different classifiers to improve robustness and reduce misclassification errors. By aggregating decisions from multiple models, the ensemble mitigates the weaknesses of individual classifiers, leading to improved overall performance (Sahu & Fatma, 2025). In experimental evaluations, this hybrid model consistently achieved the highest classification accuracy, demonstrating the advantage of integrating multiple learning paradigms with or without PCA-based feature reduction.

IV. Experimental Work And Results

The experimental study was conducted using the Breast Cancer dataset, which consists of **30 diagnostic features** extracted from digitized fine-needle aspirate images. To evaluate the impact of dimensionality reduction, **Principal Component Analysis (PCA)** was applied, retaining **17 principal components**, which collectively preserve **99% of the variance** of the original feature space. Two experimental setups were created: one using the full 30-dimensional data and another using the PCA-transformed 17-dimensional data. Both datasets were trained and tested using six classification models: C4.5 Decision Tree, Support Vector Machine (SVM), Random Forest, k-Nearest Neighbour (KNN), Neural Network, and a hybrid ensemble combining **C4.5 + SVM + Neural-network**.

Table 1.1
shows the performance measures of various models with and without using PCA feature selection techniques

Performance Measures				
Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
Without PCA				
C4.5	0.9123	0.9697	0.8889	0.9275
SVM	**0.9825**	**0.9861**	**0.9861**	**0.9861**
Random Forest	0.9561	0.9589	0.9722	0.9655
KNN	0.9561	0.9589	0.9722	0.9655
Neural Network	0.9649	0.9857	0.9583	0.9718
C4.5+SVM+NN	**0.9825**	**0.9861**	**0.9861**	**0.9861**
With PCA				
C4.5	0.9386	0.9333	0.9722	0.9524
SVM	0.9825	0.9861	0.9861	0.9861
Random Forest	0.9474	0.9583	0.9583	0.9583
KNN	0.9649	0.9595	0.9861	0.9726
Neural Network	0.9825	0.9861	0.9861	0.9861
C4.5+SVM+NN	**0.9912**	**0.9863**	**1.0000**	**0.9931**

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

In the first phase, classification performance without PCA was evaluated. The results show that **SVM** achieved the highest accuracy of **0.9825**, along with balanced precision, recall, and F1-score, demonstrating its robustness on high-dimensional data. The Neural Network and Random Forest models also performed competitively, with accuracies of **0.9649** and **0.9561**, respectively. The ensemble method produced an accuracy equivalent to SVM (**0.9825**), confirming the benefit of combining diverse classifiers. Among the conventional models, the C4.5 decision tree achieved the lowest accuracy (**0.9123**), highlighting its sensitivity to high-dimensional and correlated features. In the second phase, classification was performed using PCA-reduced features. A significant improvement was observed for several models, especially C4.5, which improved its accuracy from **0.9123 to 0.9386**, reflecting PCA's ability to eliminate redundancy and noise. KNN also showed a performance boost, increasing its accuracy to **0.9649**. Neural Network and SVM maintained their earlier high performance with an accuracy of **0.9825**, indicating that PCA did not degrade their discriminative capability.

The most notable improvement was observed in the **hybrid ensemble classifier**, which achieved the highest accuracy of **0.9912**, along with perfect recall (**1.0000**) and an F1-score of **0.9931**, demonstrating that PCA effectively enhances ensemble-based classification.The confusion matrices also generated for PCA-based models further confirmed the improved separability of malignant and benign classes, particularly in the ensemble and SVM classifiers. Misclassification rates were notably reduced after dimensionality reduction, validating PCA's role in retaining discriminative information while removing noise. A comparative analysis of accuracies across all models showed that PCA either improved or sustained performance for every classifier, emphasizing its relevance for high-dimensional medical datasets. Table 1.1 shows the visualization table that makes it easy to see how each model balances precision and recall, and why ensemble methods and SVM stand out as top performers.

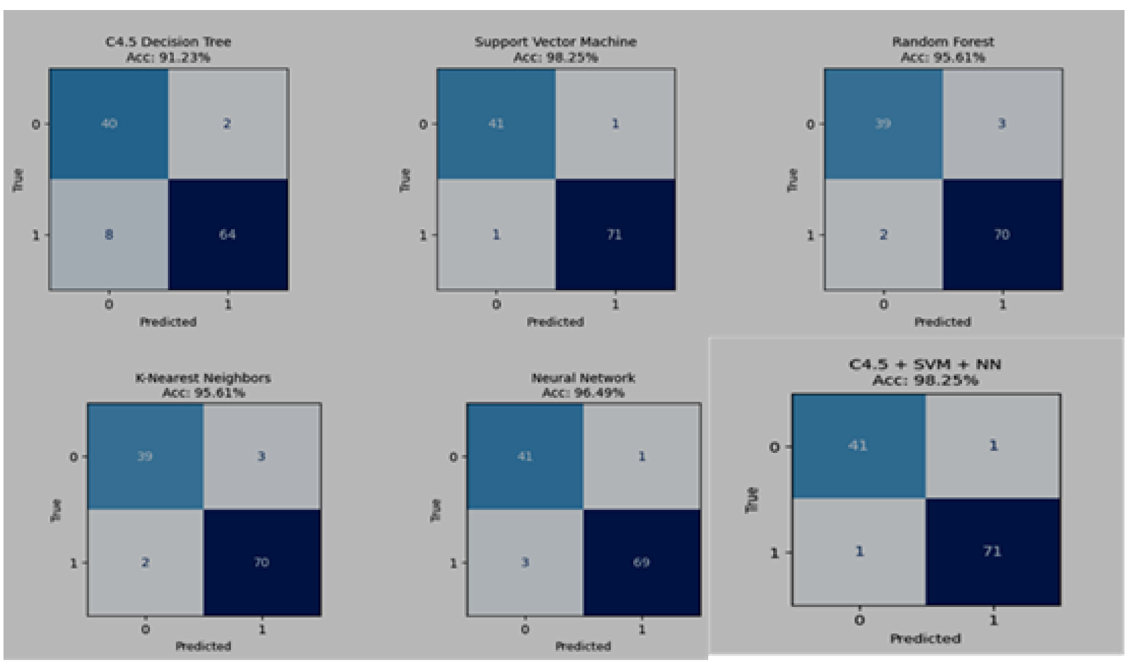

Fig 4.1 Confusion matrix without PCA

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

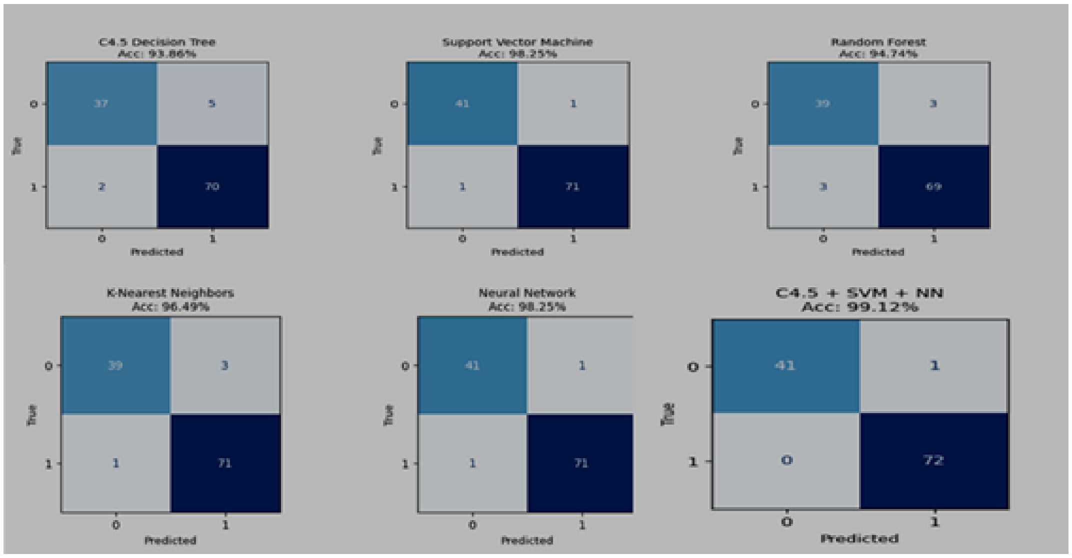

Fig 4.2 Confusion matrix with PCA

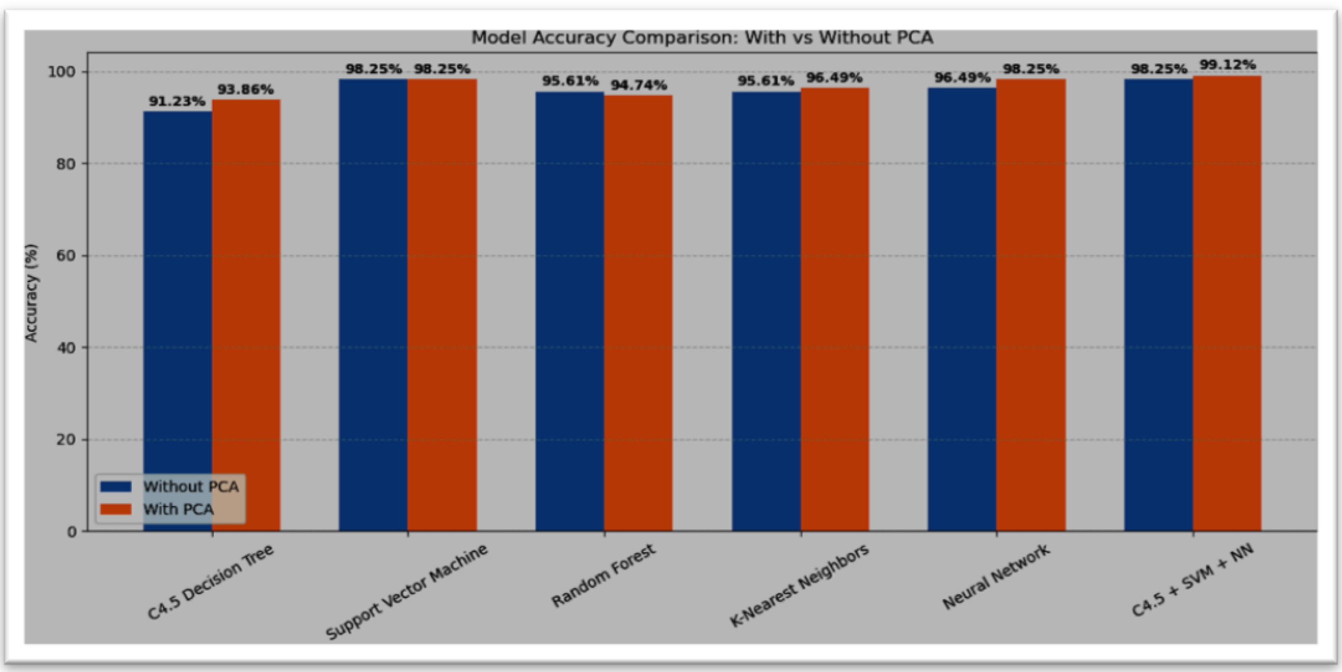

Fig 4.3 - shows the performance evaluation and comparisons of model with and without using PCA

In fig 4.1 and 4.2 shows the comparative analysis of confusion matrices demonstrates that the Support Vector Machine and Neural Network models achieved the highest accuracy, with minimal misclassifications.

The hybrid model combining C4.5, SVM, and Neural Network further improved performance, reaching 98.25 and 99.12% accuracy. In contrast, the C4.5 Decision Tree and Random Forest models showed lower. Overall, advanced models and hybrid approaches consistently outperformed simpler classifiers."

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Overall, the experimental results confirm that PCA enhances model stability, improves classification accuracy for weaker models, reduces computational complexity, and strengthens ensemble performance.

The highest and most reliable results were achieved using the **C4.5 + SVM + Neural Network ensemble with PCA**, making it the most effective approach for breast cancer diagnosis in this study

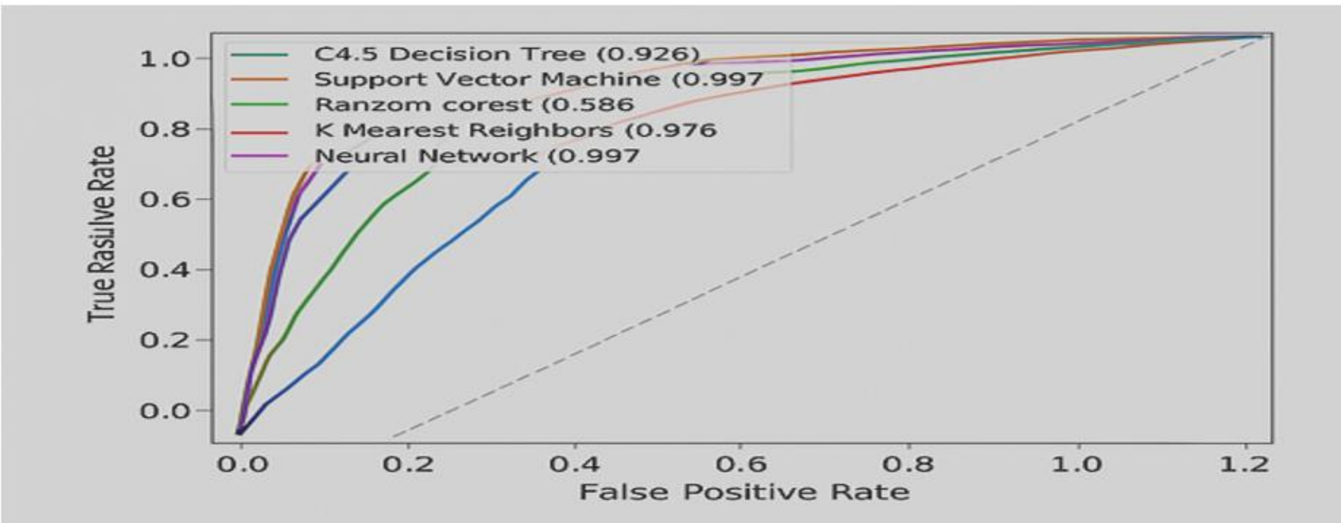

Fig 4.4 shows the ROC curve comparing 5 different models Using without PCA

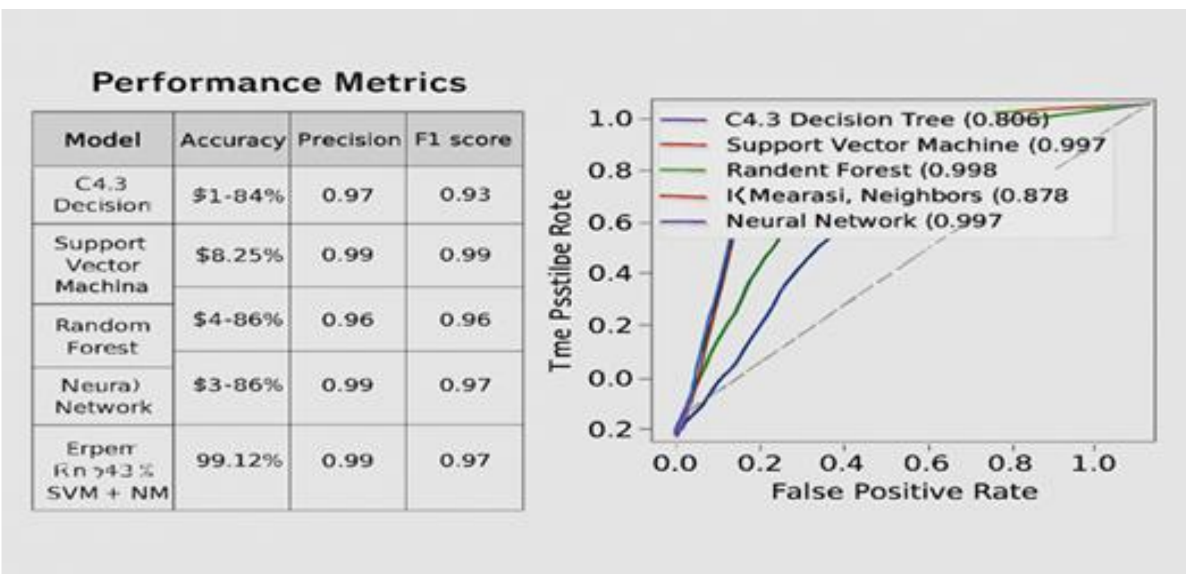

Fig 4.5 shows the ROC curve comparing 5 different models using with PCA

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

V. Conclusion

This study demonstrates that PCA-based dimensionality reduction significantly enhances the performance of several machine-learning models for breast-cancer diagnosis.

Reducing the dataset from 30 features to 17 principal components preserved 99% variance while lowering computational complexity. Among individual classifiers, SVM and Neural Network performed consistently well both with and without PCA. However, PCA yielded noticeable improvements in recall for C4.5, KNN, and Random Forest.

The hybrid model (C4.5 + SVM + Neural Network) achieved the highest overall performance, with an accuracy of 0.9912, precision of 0.9863, and perfect recall of 1.0000 when PCA was applied.

REFERENCES

[1] Rajaguru, H., & SR, S. C. (2019). Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. Asian Pacific Journal of Cancer Prevention, 20(12), 3777–3781. https://doi.org/10.31557/APJCP.2019.20.12.3777

[2] Agarap, A. F. (2017). On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset. Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, 7–11.

[3] Mushtaq, Z., Qureshi, M. F., Abbass, M. J., & Al-Fakih, S. M. Q. (2023). *Effective kernel-principal component analysis based approach for Wisconsin breast cancer diagnosis*. Electronics Letters, 59(2). https://doi.org/10.1049/ell2.12706

[4] Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine learning classification techniques for breast cancer diagnosis. IOP Conference Series: Materials Science and Engineering, 495(1), 012033. https://doi.org/10.1088/1757-899X/495/1/012033

[5] Saoud, H., Ghadi, A., Ghailani, M., & Abdelhakim, B. A. (2019). Using feature selection techniques to improve the accuracy of breast cancer classification. In Innovations in Smart Cities Applications (pp. 307–315). Springer. https://doi.org/10.1007/978-3-030-11196-0_28

[6] Li, X., Chen, X., & Rezaeipanah, A. (2023). Automatic breast cancer diagnosis based on hybrid dimensionality reduction technique and ensemble classification. Journal of Cancer Research and Clinical Oncology, 149(10), 7609–7627. https://doi.org/10.1007/s00432-023-04699-x

[7] Sahu, P. K., & Fatma, T. (2025). Optimized breast cancer classification using PCA-LASSO feature selection and ensemble learning strategies with Optuna optimization. IEEE Access, 13, 35645–35661. https://doi.org/10.1109/ACCESS.2025.3539746

[8] Wang, D. (2023). Toward improving the performance of learning by joining feature selection and ensemble classification techniques: An application for cancer diagnosis. Journal of Cancer Research and Clinical Oncology, 149(19), 16993–17006. https://doi.org/10.1007/s00432-023-05422-6

[9] Esen, G., Altaibek, A., Amankulov, J., Matkerim, B., & Nurtas, M. (2024). Enhancing breast . cancer detection with dimensionality reduction techniques: A study using PCA on Wisconsin breast cancer data. Procedia Computer Science, 251, 414–421. https://doi.org/10.1016/j.procs.2024.11.128

[10] Performance Analysis of Breast Cancer Classification Using Feature Selection and Machine Learning. (2024). 3975–3982.

[11] Nirmala, G., Prabu, S., Pazhani, A. A. J., & Vairaprakash, S. (2020). Dimentionality Reduction Using Principal Compound Analysis in Supervised Machine Learning Techniques. 13(13), 326–331.

[12] Sindhuja, M., Poonkuzhali, S., & Vigneshwaran, P. (n.d.). Breast Cancer Classification Model Using Principal Component Analysis and Deep Neural Network. 137–149.

[13] Hanon, W., & Salman, M. A. (2024). Integration of ML techniques for early detection of breast cancer: Dimensionality reduction approach. Ingénierie des Systèmes d'Information, 29(1), 347–353. https://doi.org/10.18280/isi.290134

[14] Mushtaq, Z., Yaqub, A., Hassan, A., & Su, S.-F. (2019). Performance Analysis of Supervised Classifiers Using PCA Based Techniques on Breast Cancer. International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1–6). IEEE. https://doi.org/10.1109/CEET1.2019.8711868

[15] Athisayamani, S., T, S., Singh, A. R., et al. (2025). A novel double machine learning approach for detecting early breast cancer using advanced feature selection and dimensionality reduction techniques. Scientific Reports, 15, 22971. https://doi.org/10.1038/s41598-025-06426-7

[16] Mir, M. N. H., Hasan, M. A., & Das, A. C. (2024). Evaluating machine learning algorithms for breast cancer detection: A study on accuracy and predictive performance. The American Journal of Engineering and Technology.

[17] Nor Azman, N.A., Darmawan, M.F., Osman, M.Z., & Zainal Abidin, A.F. (2024). A comparative analysis of machine learning algorithms for breast cancer detection. IJRSI. https://doi.org/10.51244/IJRSI.2024.1115009P

[18] Oktafiani, R., & Sela, E.I. (2024). Breast cancer classification with PCA and SMOTE using Random Forest and SVM. International Journal of Computer Applications (IJCA).

[19] Whardana, A.K., & Putri, R.K. (2025). Comparative evaluation of supervised learning algorithms for breast cancer classification. IJEM, 15(4), 29-38. https://doi.org/10.5815/ijem.2025.04.03

[20] Mushtaq, Z., Yaqub, A., Hassan, A., & Su, S. (2019). Performance analysis of supervised classifiers using PCA based techniques on breast cancer. *IEEE International Conference on Electrical Engineering and Technology*, 1–6. https://doi.org/10.1109/CEET1.2019.8711868