3.2. Performance Proposed Method Comparison with ROC Curve

We have implemented a Receiver Operating Characteristic (ROC), which serves as the conventional tool for model selection and assessment in problems involving the classification of two classes [41]. The ROC curve can be calculated by utilizing the True Positive Rate (TPR) and False Positive Rate (FPR) results obtained from the calculation of the confusion matrix shown in (15) and (16). The TPR and FPR values for each model are presented in Table 3, and the visualization is depicted in Fig. 11, with the x-axis representing the TPR (True Positive Rate) and the y-axis representing the FPR (False Positive Rate) measurement for each algorithm. The formula used is as follows:

$$TPR = \frac{TP}{\text{Actual Positive}} = \frac{TP}{\text{TP} + \text{FN}} \tag{15}$$

$$FPR = \frac{FP}{\text{Actual Negative}} = \frac{FP}{\text{TN} + \text{FP}} \tag{16}$$

It is evident that there is convergence among all the machine learning classifier models. The highest accuracy is achieved by the Support Vector Machine (SVM) and Logistic Regression (LR), both with an accuracy of 97.3%, while the lowest accuracy is observed in the Decision Tree model, which achieves an accuracy of 91.8%. The SVM model exhibits a higher ROC curve and a better FPR value compared to Logistic Regression, despite both models having the same accuracy. This indicates that SVM outperforms Logistic Regression in terms of classification performance. The results of the FPR and TPR for all the methods used are presented.

Table 3. Result of FPR and TPR each algorithm

Method	FPR	TPR
ANN	0.0,0.01408541,1.0	0.0,0.93023256,1.0
SVM RBF	0.0,0.0,1.0	0.0,0.94023256,1.0
DT	0.0,0.05633803,1.0	0.0,0.88697674,1.0
RF	0.0,0.04225352,1.0	0.0,0.94348837,1.0
KNN	0.0,0.04225352,1.0	0.0,0.90697674,1.0
LR	0.0,0.01408451,1.0	0.0,0.95348837,1.0
NB	0.0,0.01419451,1.0	0.0,0.89372093,1.0

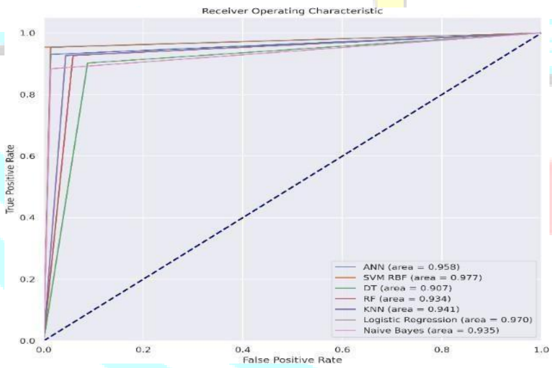

Fig. 11. ROC of all models

3.3. Performance Comparison Previous Study

Comparison of our work with the most related works show in Table 4.

Table 4. Performance comparison with previous study

Author	Dataset	Method	Accuracy
S. Ara *et al.*, 2021 [12]	UCI WBCD, 569 instances, 32 features	SVM	96.5%
Verghese *et al.*, 2021 [13]	UCI WBCD, 569 instances, 32 features	SVM:RBF	94.5%
H. Chiu *et al.*, 2020 [14]	UCI WBCD, 569 instances, 32 features	MLP + SVM	86.9%
Assegie *et al.*, 2020 [15]	Kaggle, 569 instances, 32 features	DT	92.5%
Proposed	**Kaggle, 569 instances, 32 features**	**SVM:RBF & LR**	**97.3%**

4. CONCLUSION

Based on research conducted using datasets obtained from the Kaggle site, we have explored breast cancer classification using feature selection with Principal Component Analysis (PCA) implemented into several supervised machine learning algorithms. The results obtained indicate that SVM and LR achieve the highest accuracy, reaching 97.3%. However, the ROC curve shows that the SVM graph is higher than the LR graph, which can be attributed to the results of the confusion matrix calculation, where the False Positive (FP) value is 0 and the False Positive Rate (FPR) is also 0. A FP and FPR value of 0 is considered favorable, as it signifies that the classification model accurately predicts instances as negative when they do not belong to the class in question. In cases such as breast cancer disease, minimizing FP is crucial. When the FP value is 0, it indicates that the model does not mistakenly classify something as positive when it is actually negative. Consequently, this is considered a positive outcome. Thus, the overall performance of SVM with RBF (Radial Basis Function) kernel and utilizing the c-value selection approach surpasses that of all the machine learning algorithms tested in this study. For future research, several avenues can be explored to further improve accuracy and enhance the classification diagnosis for breast cancer patients. These include applying alternative feature selection methods and optimizers, such as forward selection, to obtain the optimal set of attributes and selecting different features to increase the accuracy value.

REFERENCES

[1] X. Lin, L. Liu, and Z. Yu, "A Generic-Driven Wrapper Embedded With Feature-Type-Aware Hybrid Bayesian Classifier for Breast Cancer Classification," *IEEE Access*, vol. 7, pp. 119931–119942, 2019, https://doi.org/10.1109/ACCESS.2019.2932505.

[2] R. Roslidar and A. Rahman, "A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection," *IEEE Access*, vol. 8, 2020, https://doi.org/10.1109/ACCESS.2020.3004056.

[3] I. Hirra *et al.*, "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling," *IEEE Access*, pp. 24273–24287, 2021, https://doi.org/10.1109/ACCESS.2021.3056516.

[4] H. N. Khan, A. R. Shahid, A. H. Dar, and H. Alquhayz, "Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 165724–165733, 2019, https://doi.org/10.1109/ACCESS.2019.2953318.

[5] A. Hassan, "Performance Analysis of Supervised Classifiers using PCA based Techniques on Breast Cancer," *Int. Conf. Eng. Emerg. Technol.*, pp. 1–6, 2019, https://doi.org/10.1109/CEET1.2019.8711868.

[6] A. Ameh, M. Abdullahi, S. Balarabe, H. Hassan, and H. Chiroma, "Intelligent Systems with Applications Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer)," *Intelligent Systems with Applications*, vol. 14, 2022, https://doi.org/10.1016/j.iswa.2022.200066.

[7] A. U. L. Haq *et al.*, "Detection of Breast Cancer Through Clinical Data Using Supervised and Unsupervised Feature Selection Techniques," *IEEE Access*, vol. 9, 2021, https://doi.org/10.1109/ACCESS.2021.3055806.

[8] H. U. A. Chen, K. Mei, Y. Zhou, N. A. N. Wang, and G. Cai, "Auxiliary Diagnosis of Breast Cancer Based on Machine Learning and Hybrid Strategy," *IEEE Access*, vol. 11, pp. 96374-96386, 2023, https://doi.org/10.1109/ACCESS.2023.3312305.

[9] T. Khater, A. Hussain, S. Member, R. Bendardaf, and S. Member, "An Explainable Artificial Intelligence Model for the Classification of Breast Cancer," *IEEE Access*, p. 1, 2023, https://doi.org/10.1109/ACCESS.2023.3308446.

[10] A. S. Elkorany, M. Marey, K. M. Almustafa, and Z. F. Elsharkawy, "Breast Cancer Diagnosis Using Support Vector Machines Optimized by Whale Optimization and Dragonfly Algorithms," *IEEE Access*, vol. 10, pp. 69688–69699, 2022, https://doi.org/10.1109/ACCESS.2022.3186021.

[11] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, "Computer Methods and Programs in Biomedicine Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Comput. Methods Programs Biomed.*, vol. 223, p. 106951, 2022, https://doi.org/10.1016/j.cmpb.2022.106951.

[12] S. Ara, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms," *International Conference on Artificial Intelligence (ICAI)*, pp. 97–101, 2021, https://doi.org/10.1109/ICAI52203.2021.9445249.

[13] S. L. Verghese, I. Y. I. Liao, T. H. Maul, and S. Y. E. W. Chong, "An Empirical Study of Several Information Theoretic Based Feature Extraction Methods for Classifying High Dimensional Low Sample Size Data," *IEEE Access*, vol. 9, pp. 69157–69172, 2021, https://doi.org/10.1109/ACCESS.2021.3077958.

[14] H. Chiu and T. S. Li, "Breast Cancer – Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine," *IEEE Access*, vol. 8, pp. 204309–204324, 2020, https://doi.org/10.1109/ACCESS.2020.3036912.

[15] T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," *Journal of Robotics and Control (JRC)*, vol. 2, no. 3, pp. 115–118, 2020, https://doi.org/10.18196/jrc.2363.

[16] A. H. Osman, "An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model," *IEEE Access*, vol. 8, pp. 39165–39174, 2020, https://doi.org/10.1109/ACCESS.2020.2976149.

[17] F. Azour and A. Boukerche, "An Efficient Transfer and Ensemble Learning Based Computer Aided Breast Abnormality Diagnosis System," *IEEE Access*, vol. 11, pp. 21199–21209, 2023, https://doi.org/10.1109/ACCESS.2022.3192857.

[18] E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. Elaffendi, "Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis," *IEEE Access*, vol. 11, pp. 85760–85769, 2023, https://doi.org/10.1109/ACCESS.2023.3304242.

[19] A. R. Beeravolu, S. Azam, and M. Jonkman, "Preprocessing of Breast Cancer Images to Create Datasets for Deep-