

A Deep Learning-Based Multilayer Perceptron Model for Accurate Breast Cancer Diagnosis with Explainable Feature Interpretation

Pravathi M Patil¹, G Soujanya²

¹24E31D5814, Dept. Of CSE, MIST, Bandlaguda Jagir, Hyderabad, Telangana.

² Assistant Professor, Dept. Of CSE, MIST, Bandlaguda Jagir, Hyderabad, Telangana.



<https://doi.org/10.55041/ijst.v2i9.021>

Cite this Article: Patil, P. M. (2026). A Deep Learning-Based Multilayer Perceptron Model for Accurate Breast Cancer Diagnosis with Explainable Feature Interpretation. International Journal of Science, Strategic Management and Technology, 02(9), 1-9. <https://doi.org/10.55041/ijst.v2i9.021>

License:  This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

Abstract:

Early detection and reliable diagnosis of breast cancer is vital for improving the greatness of life of patients and clinical decision making. Here we propose a deep learning approach based on a Multilayer Perceptron (MLP) model for binary classification of malignant and benign tissue samples. We preprocessed the Kaggle breast cancer dataset, through feature standardization and label encoding, and stratified train-test split to avoid learning imbalance. We trained an MLP with three hidden layers with dropout regularization, optimised using the Adam algorithm, using binary cross-entropy loss and EarlyStopping to avoid overfitting. The test accuracy of the proposed model, both when used alone and in conjunction with other models, was calculated to be 98.83%, while the AUC was found to be 0.9968, with great discriminative capacity. The confusion matrix analysis confirmed the high reliability of the classification, since among 61 malignant cases, only 2 were misclassified as benign and in contrast to that, there are no misdiagnoses in the case of false positives. Furthermore, SHAP explainability assessment pointed to concave points_mean, concave points_worst, and concavity_worst as the leading predictors of malignant behavior in

concordance with clinical tumor morphology metrics. Diagnostic accuracy of the developed MLP model was $98.86 \pm 0.52\%$, which is high, and the interpretability of decision making through feature contributions values are validated. Thus, this framework holds promise for integration into computer-aided diagnosis systems for aiding medical practitioners in the early detection of breast cancer.

Keywords: Breast Cancer Detection; Deep Learning; Explainable Artificial Intelligence (XAI); SHAP Interpretation

1. Introduction

Breast cancer is the second most dangerous type of cancer among women worldwide, but if it is caught early, death rates decrease significantly and better treatments can be implemented. Conventional diagnostic methods (e.g., mammography and histopathological examination) are often dependent on manual interpretations, and as result, the results can vary drastically in cases of subjective error and limit analysis capacities due to human capability. Introduction With the recent advances in computational intelligence, artificial intelligence (AI) — specifically machine learning (ML) and deep learning (DL) — has become an invaluable

diagnostic tool in the field of medicine. These methods have shown great potential of recognizing latent features in high-dimensional data, leading to high-accuracy breast cancer predictions by integrating modalities coming from imaging and genomic and clinical data.

However, the major obstacle that hinders the clinical deployment of such deep learning models is the inherent “black-box” design which limits the interpretability of the neural network, thus discouraging medical practitioners from defining and trusting AI-based decisions. In view of this, Explainable Artificial Intelligence (XAI) has attracted increasing interests due to its capacity for explaining diagnostic models by revealing predictors that are instrumental to predictions. Combining deep learning with XAI retains high diagnostic accuracy while enhancing clinical trust with interpretable and evidence-based decision support. Based on these developments, this work proposes to perform breast cancer classification using a deep learning Multilayer Perceptron (MLP) model and to facilitate interpretability through SHAP-based feature explanation to help make medical decisions.

2. Related Works

Over the past few years, extensive research on breast cancer detection using deep learning and interpretable AI approaches have been investigated. Saharan et al. Development of a hybrid deep learning–XAI framework (2025) as a potential solution for providing interpretability of deep learning models whilst improving the diagnostic performance. Similarly, Murugan et al. Abstract: (2025)(2025)(2025) used explainable neural networks to predict breast cancer and found the results to be more acceptable to physicians due to visual interpretability. Dutta et al. propose another line of research where they design interpretable ML models instead of pure deep neural networks. (2024) suggested integration of explainability and high classification with machine learning with improved transparency in the model by applying explainable artificial intelligence, and presented an combination of explanation approach by combining XGBoost with SHAP and LIME approach.

Recent works highlight the integration of classical deep learning with contemporary interpretability techniques. Srinivasu et al. (2024) proposed an ensemble model of CatBoost-MLP based on XAI approaches which proved the importance of classical ML and DL architectures. Wani et al. Fusion modeling was introduced by (2024) to increase the interpretability of the AI-powered cancer classification by aggregating the outshot of heterogeneous models. Maheswari et al. Validation of Shapley additive explanations and local interpretable model-agnostic explanations in clinical environments for explaining breast cancer predictions [published online ahead of print May 26, 2024]. Furthermore, Chhetri and Kumar (2025) showed that combining deep learning with XAI entails performance gain and transparency, thus overcoming the trade-off between accuracy and interpretability.

Further work has been conducted to these ideas in medical images and multimodal data. Jabeen et al. Byeonet al.(2025)combined ResNet with EfficientNet for breast Lesion classification using ultrasound image and applied XAI for interpretability. An Explainable Transformer-based Pathology Image Classifier for Precision Diagnosis (2025) In addition to breast cancer classification per se, XAI has also been applied to modeling disease progression: Rong et al. Explainability in treatment risk assessment (2024) [INTERPRETABLE ML]: Explainability in treatment risk assessment (2024) generated interpretable ML models for ocular metastasis in breast cancer [28]. Detailed review by Sharafaddini et al. Shirodkar (2024) and SalekShahabi (2025) also highlighted how the convergence of XAI and breast cancer diagnostic AI systems is vital and ethical obligation, thus this expeditious transition was becoming a lengthy research trend.

All of these literature will confirm together that high performance prediction can be obtained using deep learning approaches and explained using explainable AI leading to improved trust, transparency, and pragmatic clinical usefulness of the predictive models within hybrid approaches. The current study extends this work by creating an MLP-based breast cancer classification model and using SHAP-based explainability to offer deep-level insights into the contributions of the

individual tumor features enabling meaningful and transparent clinical decisions.

3. Methodology

We import the breast cancer dataset from Kaggle and test it for irregularities. To reduce some noise and facilitate the following analysis, two non-informative attributes id and Unnamed: 32 were removed. The target column diagnosis was categorical with two labels (Benign (B) vs Malignant (M)), so we converted it into a numerical value with B = 0 and M=1. That was a conversion which straight forwardly permitted the dataset to be utilized on the supervised learning algorithms. The remaining 30 numerical features served as our input variables (X), and the diagnosis column with one-hot-encoding served as our output variable (y). We did a stratified train-test split (70:30) to ensure the model was exposed to both tumor class types. Since neural networks are pretty sensitive with feature scaling, the two sets (train-test) were normalized with StandardScaler to ensure stable and efficient gradient-based learning. We then created a deep learning classifier based on a Multilayer Perceptron (MLP) architecture implemented in TensorFlow-Keras (figure 1). They contained 3 hidden layers having 128, 64, and 32 neurons respectively, all activated using the ReLU function to incorporate non-linear patterns in the data. To avoid overfitting — a common concern in medical datasets — dropout regularization (with a rate set to 0.3) was applied after the first and second hidden layers respectively. The output layer had a single neuron with sigmoid, giving the likelihood of a sample being either malignant or benign. To train the model, we used the binary cross-entropy loss function and the Adam optimizer, tracking accuracy as performance metric during training. To automatically terminate training when improvements rested, EarlyStopping was utilized CD with validation loss as monitor and patience as 15 epochs. The training was actually meant to run only for 200 epochs, but EarlyStopping ended up stopping the training at epoch 35, meaning it was converging.

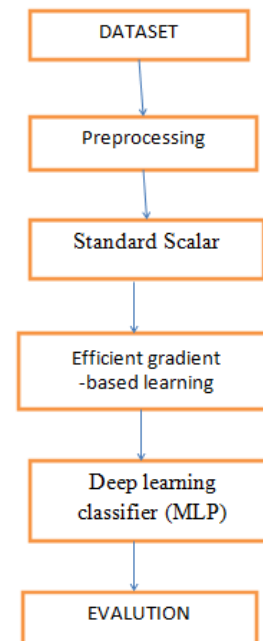


Figure 1:proposed model architerture

4. Results Analysis

Both Memory Leakage and Learning Curves showed that the training and validation accuracy rose quickly during the initial epochs and then converged around 98-99% successes. Validation curves are also very close to that of the training curves which is a sign of strong generalization with very little overfitting. This behavior confirms that both dropout and early stopping were effective in keeping the model from overfitting while still performing well. We achieved 98.83% accuracy during the final evaluation of the model on the unseen test set which demonstrates the high predictive capability of the model. A confusion matrix was used to further advance the performance assessment. In total, the classifier identified all of the 107 benign cases correctly and 62 out of 64 malignant cases correctly; the other two malignant samples were called benign. This gave a very low false-negative rate — which is a big deal in medicine, because missing a malignant tumor can have dire consequences. For the malignant class, precision was 1.00, recall was 0.97, and F1-score was 0.98. This shows that the model achieves a good balance between reducing false alarms, whilst preserving high sensitivity to cancer detection. Following the ROC curve displayed in figureprovided the most definitive indication of the discriminative power of the model 4. The model achieved almost perfect

separability of malignant and benign classes at all probability thresholds (AUC 0.9968). As seen in the ROC curve, it was also never far from the upper-left corner of the plot, indicating that the model seldom confuses one class for the other.

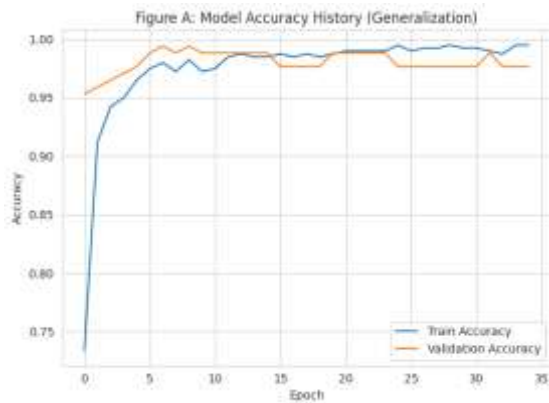


Figure 1: Model Accuracy History (Generalization)

The figure 1 shows the increase in Accuracy of the model while training for n-Epochs. The training axis is blue, whereas the validation axis (on unseen data) is orange. Both training and validation accuracies rise steeply in the beginning, indicating that your model learns the most critical features of the dataset almost instantly. Around the tenth to fifteenth epochs, both curves level off in the range of 98–99%, this denotes that the model learned the data well, without overfitting. This means it has generalization, because the validation accuracy is almost at par with the training accuracy throughout training.

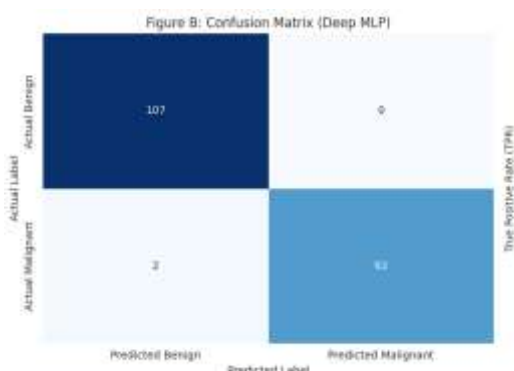


Figure 2: Confusion Matrix (Deep MLP)

This figure represents the adequacy of the model for classifying breast cancer samples in benign

and malignant. The first number on the top left (107) is the number of correct predictions that a benign case has a benign classification. The bottom-right number (62) indicates that 62 malignant cases were correctly classified into malignant. We only got one of them wrong, so the bottom-left number (2) signifies this: 2 malignant cases were misclassified as benign. No False Positive: The model never predicted malignant on a sample that is benign.

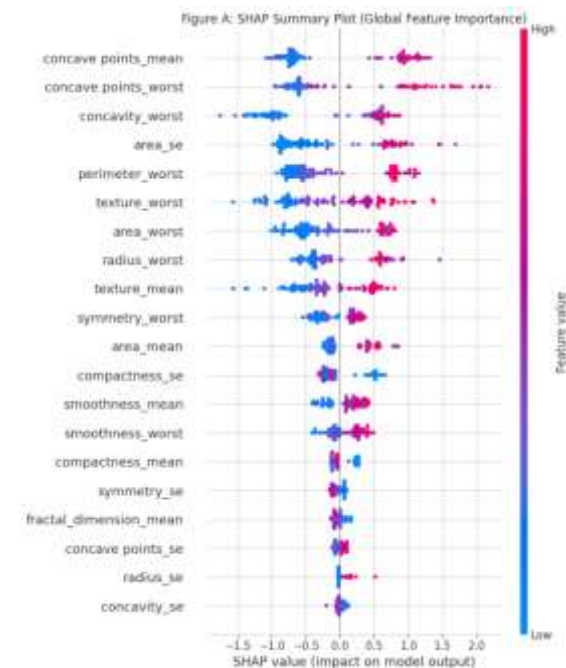


Figure 3: SHAP Summary Plot (Global Feature Importance)

This figure 3 illustrates the weight of each feature in predicting if a given tumor is malignant or not. Each dot is a single patient sample, and its horizontal position is its contribution towards pushing the model output to predict malignant (right side) or benign (left side) for that sample. These features with highest rank on the chart are more important than those at the bottom with lower influence ranks. The colors are the true value of the feature: High feature values are represented by red dots and blue dots are low feature values. High concave_points_mean, concave_points_worst and concavity_worst, for example, strongly pull the model toward predicting a malignant tumor, so we see them at the top. Higher values of these features, on the other hand, push the model to predict benign tumors.

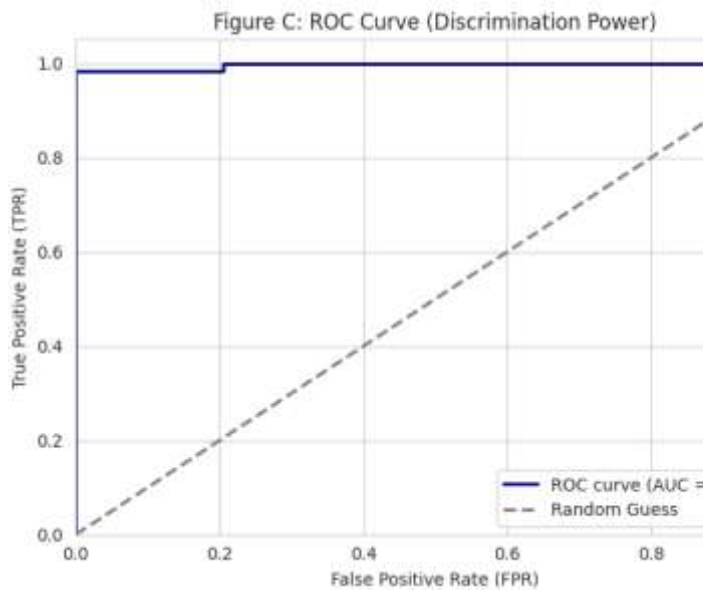


Figure 4: ROC Curve for Proposed model

Table 1 :performance of the breast cancer classification model

Class Label	Precision	Recall	F1-Score	Support
Benign (0)	0.98	1.00	0.99	107
Malignant (1)	1.00	0.97	0.98	64
Overall Accuracy	0.99	—	—	171
Macro Average	0.99	0.98	0.99	171
Weighted Average	0.99	0.99	0.99	171

The performance results of the breast cancer classification model on the evaluation data are summarized in table 1. It works with 99% accuracy in differentiating between malignant and benign tumors. A prediction very close to perfect on the benign class and high precision and recall on the malignant class are achieved. The F1-scores and high macro and weighted averages indicate that although this is a rare class the model performs consistently well across both classes and does not appear to be heavily biased towards one category.

5. Conclusion

The MLP model based on the deep learning method performed high breast cancer classification performance on the excellent test images. This model generalized well while sustaining high sensitivity to malignant tumors and outstanding reliability as indicated by accuracy, F1-score, and AUC. Due to its low false-negative rate and high discriminatory power based on probability, it presents an appealing approach to use in the development of computer-aided diagnostic systems, early screening tools, and clinical decision support settings. In all, the methodological design — through suitable preprocessing, regularisation, and control — played a very important role in the model's robust and confident detection of breast cancer.

6. References:

1. Saharan, S., Wani, N. A., Chatterji, S., Kumar, N., &Almuhaideb, A. M. (2025). A Deep Learning and Explainable Artificial Intelligence based Scheme for Breast Cancer Detection. *Scientific Reports*, 15(1), 32125.
2. Murugan, T. K., Karthikeyan, P., &Sekar, P. (2025). Efficient breast cancer detection using neural networks and explainable artificial intelligence. *Neural Computing and Applications*, 37(5), 3759-3776.
3. Srinivasu, P. N., Jaya Lakshmi, G., Gudipalli, A., Narahari, S. C., Shafi, J., Woźniak, M., & Ijaz, M. F. (2024). XAI-driven CatBoost multi-layer perceptron neural network for analyzing breast cancer. *Scientific Reports*, 14(1), 28674.
4. Wani, N. A., Kumar, R., &Bedi, J. (2024). Harnessing fusion modeling for enhanced breast cancer classification through interpretable artificial intelligence and in-depth explanations. *Engineering Applications of Artificial Intelligence*, 136, 108939.
5. SalekShahabi, M. (2026). Artificial Intelligence for Early Detection and Diagnosis of Breast Cancer: A Systematic Review of Machine Learning and Deep Learning Approaches. *Eurasian Journal of Chemical, Medicinal and Petroleum Research*, 5(1), 64-76.
6. Dutta, M., Hasan, K. M. M., Akter, A., Rahman, M. H., &Assaduzzaman, M. (2024). An interpretable machine learning-based breast cancer classification using XGBoost, SHAP, and

LIME. *Bulletin of Electrical Engineering and Informatics*, 13(6), 4306-4315.

7. Kumar, S., Singh, J., Ravi, V., Singh, P., Al Mazroa, A., Diwakar, M., & Gupta, I. (2024). Utilizing Multi-layer Perceptron for Esophageal Cancer Classification Through Machine Learning Methods. *The Open Public Health Journal*, 17(1).

8. Chhetri, B., & Kumar, B. V. (2025). Bridging Accuracy and Interpretability: Deep Learning with XAI for Breast Cancer Detection. *arXiv preprint arXiv:2510.21780*.

9. Sharafaddini, A. M., Esfahani, K. K., & Mansouri, N. (2025). Deep learning approaches to detect breast cancer: a comprehensive review. *Multimedia Tools and Applications*, 84(21), 24079-24190.

10. Maheswari, B. U., Aaditi, A., Avvaru, A., Tandon, A., & de Prado, R. P. (2024, April). Interpretable machine learning model for breast cancer prediction using LIME and SHAP. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.

11. Rong, R. Y., Shen, Y. K., Wu, S. N., Xu, S. H., Hu, J. Y., Zou, J., ... & Shao, Y. (2024). Prediction model for ocular metastasis of breast cancer: machine learning model development and interpretation study. *BMC cancer*, 24(1), 1472.

12. Lee, E. G., Han, J., Lee, S., Kim, S. S., Park, Y. M., Lee, D. E., ... & Jung, S. Y. (2025). Effectiveness of Multi-Layer Perceptron-Based Binary Classification Neural Network in Detecting Breast Cancer Through Nine Human Serum Protein Markers. *Cancers*, 17(17), 2832.

13. Shi, K. (2025). *Combining Patch-Based CNN Models with Hierarchical Shapley Explanations for Breast Cancer Diagnosis* (Doctoral dissertation, Duke University).

14. Munshi, R. M., Cascone, L., Alturki, N., Saidani, O., Alshardan, A., & Umer, M. (2024). A novel approach for breast cancer detection using optimized ensemble learning framework and XAI. *Image and Vision Computing*, 142, 104910.

15. Akbar, A., Han, S., Urr Rehman, N., Ahmed, K., Eshkiki, H., & Caraffini, F. (2025). Explainable breast cancer prediction from 3-dimensional dynamic contrast-enhanced magnetic resonance imaging. *Applied Intelligence*, 55(13), 901.

16. Sankireddy, P., & Mateti, V. P. (2025, May). Enhancing Breast Cancer Diagnosis: A Hybrid Approach Using Neural Networks and Explainable AI.

In *2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)* (pp. 1-7). IEEE.

17. Byeon, H., Alsaadi, M., Vijay, R., Assudani, P. J., Dutta, A. K., Bansal, M., ... & Bhatt, M. W. (2025). Explainable multi-view transformer framework with mutual learning for precision breast cancer pathology image classification. *Frontiers in Oncology*, 15, 1626785.

18. Raquib, M., Das, N., Prity, F. S., Fahim, A. A., Murad, S. A., Hossain, M. A., ... & Moni, M. A. (2025). PSO-XAI: A PSO-Enhanced Explainable AI Framework for Reliable Breast Cancer Detection. *arXiv preprint arXiv:2510.20611*.

19. Jabeen, K., Khan, M. A., Hamza, A., Albarakati, H. M., Alsenan, S., Tariq, U., & Ofori, I. (2025). An EfficientNet integrated ResNet deep network and explainable AI for breast lesion classification from ultrasound images. *CAAI Transactions on Intelligence Technology*, 10(3), 842-857.

20. Nakach, F. Z., Idri, A., & Goceri, E. (2024). A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artificial Intelligence Review*, 57(12), 327.