iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** III **Month of publication:** March 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.78780

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue III Mar 2026- Available at www.ijraset.com*

# Explainable Deep Learning Framework for Breast Cancer Classification

Zainab Siddiqui[1], Sehba Fatima[2], Noor Fatima[3], Roshan Jahan[4]
*Department of Computer Science and Engineering, Integral University, Lucknow*

***Abstract:*** ***Worldwide, the number of people with cancer continues to represent a significant health risk in that new cancer cases are identified before the disease is detected by traditional means (the appearance of signs and/or symptoms). In recent years, new artificial intelligence based technologies have been developed to aid doctors in improving their traditional methods of diagnosing and treating cancer. Convolutional neural networks are one such new diagnostic method which uses large quantities of data including medical imaging (mammograms, ultrasounds, MRIs, etc.) to correctly identify diseases such as breast cancer. In this paper, we describe the use of convolutional neural network technology and explainable artificial intelligence (XAI) to create an entirely new breast cancer risk prediction methodology.***

***In this study, the Breast Ultrasound Images (BUSI) dataset is used. In this study, images are classified as normal, benign, and malignant. Ultrasonography is an imaging modality that is effective in examining dense breast tissue.***

***Convolutional Neural Networks are used for feature extraction in the classification of tumor patterns. Despite the high degree of predictability offered by models developed using convolutional neural network (CNN) architecture, the issue of transparency has been raised against such models. In this context, the transparency of the decision-making process has been emphasized.In order to improve the above-mentioned limitations, Explainable AI techniques are utilized in the proposed framework. As a result of applying the Explainable AI techniques (GRAD-CAM images), the decision-making process of the models can be explained, along with the regions of the input image that affect the decision-making process. As a result, the transparency of the decision-making process can be improved.***

***The results obtained from the experiments suggest that the proposed CNN-XAI model achieves an Area Under the Curve (AUC) of 0.9934 and an accuracy of 97.6%, which are higher when compared to other CNN-based approaches.The implications of this discovery are that potential advantages can be obtained by using deep learning in conjunction with an explanation component.***

***Keywords: Breast Cancer Detection, Convolutional Neural Networks (CNN), Explainable Artificial Intelligence (XAI), Grad-CAM Visualization, BUSI Dataset (Breast Ultrasound Images), Medical Image Analysis***

## I. INTRODUCTION

Breast cancer is one of the most common and deadly cancers affecting women all over the globe. There were approximately 2.3 million new cases of breast cancer and nearly 688,000 deaths from it worldwide reported in 2020 according to global cancer statistics [1].

Even though there have been major advances in screening technologies and therapeutic approaches, the persistent high incidence and mortality rates still demonstrate a need for more accurate methods to detect breast cancer early and to predict the likelihood of developing breast cancer.The standard methods for diagnosing breast cancer include mammograms, ultrasound, and clinical breast exams. However, they all produce a high rate of false positives and false negatives, depend on how each clinician interprets them, and do not work as effectively in women who have dense breast tissue [2].

These challenges made interest in finding ways to supplement clinicians with the use of computer-based methods to provide a higher degree of consistency and data to assist them with making more accurate diagnoses.Not only have there been many advances recently with computer applications in medicine, there have also been significant advances in applying artificial intelligence (AI), especially deep learning, to analyze medical images and perform predictive analyses. One of the most successful types of AI algorithms to analyze images has been the convolutional neural network (CNN); CNNs have been very good at automatically learning complex visual patterns from the imaging data and at producing superior results for tumor detection, tumor classification and predicting whether a tumor is malignant using different breast imaging modalities [3].

Breast cancer predictive capacities for traditional machine learning models including Random Forest (RF) and other feature-based classifiers have been investigated, especially when using structured clinical or demographic data sets.

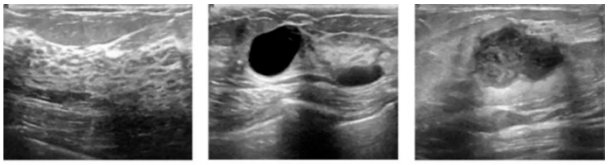
 

However, these models are typically very dependent on features that are created by humans and are therefore unable to represent the complex spatial aspects that make up medical imaging data. Comparing results from the modelling phase indicated that the use of a CNN-based architecture resulted in better predictive capability and more robustness when applied to multimodal breast cancer data sets than RF or feature-based models; hence the CNN-based architecture should be considered the preferred approach for this type of task [4].

Despite their high predictive ability, deep learning models have been criticized for their lack of interpretability. In particular, because of the "black-box" characteristic of deep neural networks, there are concerns about transparency, dependability, and trustworthiness in AI systems that render automated decisions automatically. Therefore, XAI methods have been implemented to help increase interpretability while still maintaining high predictive capabilities [5].

Explainable AI (XAI) methods are designed to help provide insight into the internal thought process of AI models by identifying the features or areas that have the greatest effect on the outcome of a prediction. The addition of XAI techniques into deep learning frameworks improves both model usability and transparency for clinical applications[6].

This research introduces a framework for predicting breast cancer through explainable Ai, which contains a convolutional neural network (CNN) model built-to-order on multi-feature data including ultrasound imaging data; thus increasing the accuracy of the prediction process. This framework is optimized for increasing the explanations provided by the CNN model, thus increasing clinician understanding of where and how the model makes its predictions.The integration between Grad-CAM & overall framework was designed specifically to give clinicians increased understanding of their limitations when making a clinical diagnosis based off an AI-derived output by visualizing which areas of the medical image contributed the most to the model's predicted diagnostic result [7].



Normal Benign Malignant

## II. LITERATURE REVIEW

Incidence rates of breast cancer make the disease a priority for improving screening, risk assessment and/or alternative early detection methods; historically the traditional mammographic, ultrasound or MRI based screen tools and risk prediction models all contributed to a reduction in breast cancer related mortality - challenges remain in accurately interpreting the clinical data collected from patients with dense breast tissue (and patients across clinicians). Subsequently there has been an increased interest in automating the prediction of breast cancer through data driven approaches.

### *A. Public Datasets and Benchmarking*

The advancement of AI based prediction of breast cancer has been fostered through the availability of public datasets used for training algorithms and benchmarking. The Digital Database for Screening Mammography (DDSM) has been a primary resource used for developing and benchmarking algorithms; the CBIS-DDSM is a larger curated subset of the DDSM that provides mammograms with validated pathology annotations as well as defining regions of interest within the mammograms. In addition, there are ultrasound-focused datasets and also datasets that provide histopathology slides, allowing for the comparison of modalities. Despite the availability of these public datasets for developing prediction models, there are limitations associated with legacy image file formats, the differences in quality of annotations over time, as well as the imbalance in the number of examples of each of the classes present in the datasets.

### *B. Conventional Machine Learning and Deep Learning*

In the past, researchers used conventional machine learning methods to obtain manually crafted radiomic or clinical characteristics as input(s) for classifiers, such as SVMs, Random Forests, and XGBoost. These classification methods may not have been entirely unsuccessful; however, with the rise and popularity of deep learning, especially via convolutional neural networks (CNNs), deep learning has been established as the most efficient process for classification. CNNs automatically identify hierarchical layers of features through the acquisition of images as straight pixel data and often outperform traditional machine learning methods (especially for classifiers when trained on large accurately annotated data sets).

The findings of this work also indicate that CNNs had the highest corresponding predictive performance when compared to traditional classifiers for breast cancer detection.

### *C. CNN with Explainable AI for Breast Cancer Prediction*

Using CNNs, the use of gradient-weighted class activation mapping (Grad-CAM) and gradient-weighted class activation mapping plus (Grad-CAM++) techniques enable predictions by machine learning (ML) models to be understandable and interpretable through the generation of heatmaps showing areas in mammograms that have had a substantial impact upon how the ML model made its prediction. By presenting visual representations or explanations of the prediction, clinicians can trust the ML model's predictions as well as potentially identify subtle characteristics of breast tissue that they may overlook without the assistance of an ML model. A diverse group of publicly available data was used to validate the ML model and demonstrate the accuracy of its predictions for multiple groups of people and across all imaging conditions. Therefore, combining high-quality predicting with explainable output creates an essential bridge for clinicians between model accuracy and usability, allowing the ML model to assist in the practical application of healthcare.

### *D. Multimodal Theoretical Approaches and Ensemble Theoretical Approaches*

Due to some of the limitations of using image data alone; additional data such as clinical (subjective), demographic (objective), genetic, etc., that is available in the clinical setting will add greater value to the evaluation of a patient's medical condition. Using Ensemble Theoretical Models, such as combining Convolutional Neural Network (CNN) outputs with other models and/or feature-based methods, improves the overall robustness and predictive accuracies of the classifying, rating, or recommending process. Additional techniques such as feature selection, domain adaptation, and advanced data augmentation also enhance the performance of the model.

### *E. Consideration of Limitations to and Reproducibility and Ethical Issues Related to Use of AI Models*

There have been significant improvements; however, barriers still exist (e.g., small amount of access to large/sufficient diversity in datasets, standardization of how to evaluate an AI Model & compare it against other models). In addition, the issue of domain adaptation exists as the domain may not be at homeostasis due to equipment differences and/or differences in the patient population. Finally, there are numerous regulatory, legal, and explainability issues related to AI Models being transitioned to clinical practice due to need for standardized reporting, need for external validity, and evaluation of AI fairness in terms of demographics.

### *F. Research Gaps Motivating This Study*

Prior research into AI-driven breast cancer prediction shows potential; however, certain key areas have not been explored. These include developing reliable, multi-modal (various imaging types), and explainable (able to explain the basis of why the AI made certain predictions) AI models that can be tested using diverse patient populations. This research will fill these gaps by developing a CNN-based model that will produce explainable outputs based on publicly available imaging and clinical information, thus generating accurate predictions along with meaningful explanation for use in clinical practice.

## III. METHODOLOGY

### *A. Dataset*

The presented study uses the BUSI dataset with ground truth provided by the Faculty of Informatics, Mansoura University, Egypt. This dataset is being widely used in many works concerned with the detection and classification tasks of breast lesions due to its diversity and well-annotated imaging samples. There are 2531 ultrasound images for female patients of different ages; each image is accompanied by a ground truth segmentation mask pointing to the lesion region annotated by certified radiologists[14].

The dataset has three total diagnostic classes: normal, benign, and malignant; see Table 1.

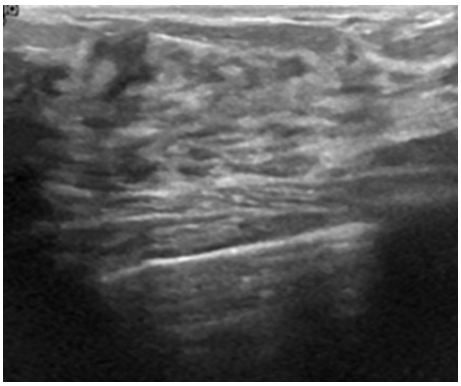
Table 1:Performance of the CNN, Random Forest and SVM Models

| Class | Number of Images | Description |
|---|---|---|
| Normal | 798 | healthy breast tissue with no lesions present |

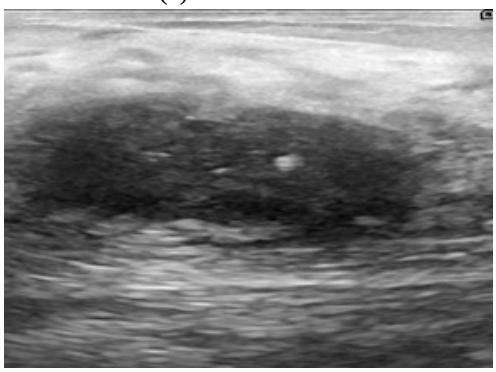
| | | |
|---|---|---|
| Benign | 891 | non-cancerous breast abnormalities (for example, cysts, fibroadenomas) |
| Malignant | 842 | cancerous lesions where positive histopathological evidence was found to be present |

All ultrasound images are obtained using a LOGIQ E9 system with a 6-15 MHz linear transducer. All images are saved into the PNG format, with image resolutions ranging between 500×500 and 700×700 pixels. In this study, all images were resized to 224×224 and the pixel intensity was normalized in the range [0,1] for model training[15].

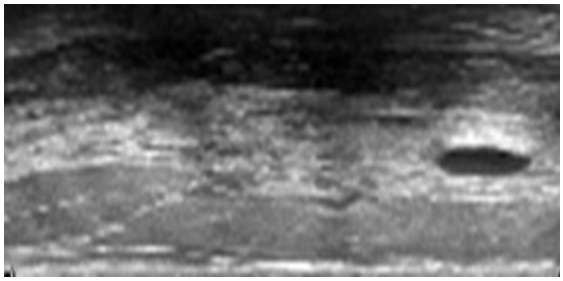
To enhance the way the model does predictions, we expanded the dataset by performing data augmentation techniques like: randomly flipping, rotating, and zooming images The dataset then was broken up into three datasets: training, validation, and test, with the percentages of 75%, 15%, and 10%, respectively.The Synthetic Minority Over-sampling Technique (SMOTE) will be used to generate additional data points in order to balance the existing class distributions. Only data from the training sets will be used to produce new examples for SMOTE



(a) normal tissue



(b)malignant lesion



(c)benign lesion

Figure 1: Ultrasound images used to visually compare normal tissue with malignant and benign lesions.

*B. Proposed Framework*

It is within this context that this present study set out and achieved a comparative analysis of three unique machine learning and deep learning models/ algorithms, namely CNN, SVM, and RF, to ascertain an accurate diagnostic classification of breast lesions. The models/ algorithms were developed independently based on unique computational principles governing each model/ algorithm, although for analysis of the same data set.

*C. Model Architectures*

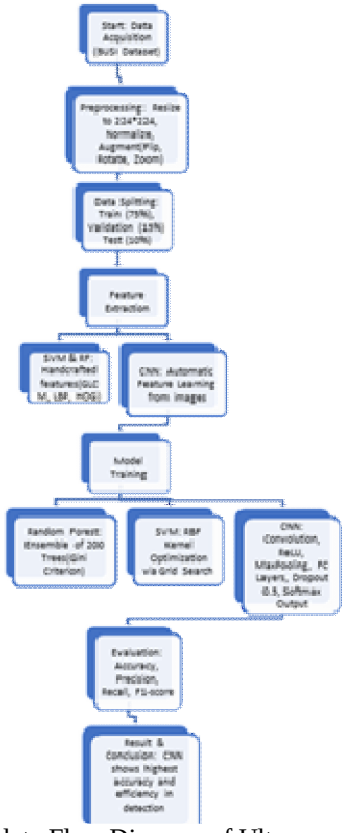



Figure 2: Complete Flow Diagram of Ultrasound Picture

*1) Convolutional Neural Network (CNN) Model*

The input to the model consists of a 224 pixel by 224-pixel image; after 5 convolutional layers consisting of 32 filters, 64 filters, 128 filters, 256 filters, and 128 filters respectively are applied to this initial image via convolution, the output will consist of an activation function (ReLU) applied to each of the outputs from each of the 5 convolutional layers in succession. Each of the convolutional filter's kernels measurements are 3 pixels across and 3 pixels tall with stride lengths of 1 pixel along each of the image axes.

Max pooling is performed using 2 x 2 filters to down-sample the feature maps and also reduce the computational cost of the network. Once the features have been extracted, they are passed to two fully-connected layers that each have 128 and 64 neurons, respectively. Dropout (0.5) and early stopping have been implemented to reduce the chance of overfitting.

The last layer of the model has a Softmax activation function with three output neurons that correspond to Normal, Benign, and Malignant classifications. The training process used categorical cross-entropy loss and the Adam optimizer to support the model learning to effectively learn from the original pixel data through the entire network end-to-end.[16].

Mathematically, the CNN layer operation is given by:

$$x_{l+1} = f(W_l * x_l + b_l)$$

where $x_l$denotes the feature map at layer $l$ , $W_l$represents the kernel, and $f$ is the ReLU activation.

*2) Support Vector Machine (SVM) Model*

The classical machine learning algorithm used in developing the SVM model was based on low-level image features that were hand engineered from ultrasound images. The features derived were from Gray-Level Co-occurrence Matrix and Local Binary Pattern attributes, which account for texture statistics like contrast, correlation, energy, and entropy[17].

The objective of an SVM is to distinguish between classes as effectively as possible by determining a decision boundary that provides the best separation. This decision boundary is mathematically expressed as,

$$f(x)=sign(w\cdot x+b)$$

Here, **w** represents the weight vector and **b** denotes the bias term. To effectively model data that cannot be separated in a linear manner, the **Radial Basis Function (RBF) kernel** is used to capture non-linear relationships between features.

$$K(xi, xj) = exp(-\gamma \parallel xi - xj \parallel 2)$$

The model's key parameters, namely C (which controls regularization) and γ (which defines the kernel's spread), were tuned using a grid search strategy combined with 5-fold cross-validation. The optimal parameters found via this study for maximum performance were C = 10 and γ = 0.01.

*3) Random Forest (RF) Model*

The framework of the Random Forest algorithm, which is based on the ensemble learning technique, has been utilized. The algorithm has been trained with the same hand-crafted features that have been utilized in the SVM model. The process of training the model in the Random Forest algorithm is as follows: a number of decision trees have been created, and then a voting process takes place, wherein the class label is determined based on the majority votes. The trees have been trained, and the features have also been randomly chosen. [18]

The classification decision for a given instance$x_i$ is expressed as:

$$\hat{y_i} = mode\{h_t(x_i)\}_{t=1}^{T}$$

where $h_t$ represents the $t^{th}$ decision tree and $T$ is the total number of trees.*For the present investigation, the random forest model was developed with an ensemble of 200 trees, with a maximum of 15 depths and the Gini impurity assessment. The hyperparameters considered for the model development were the number of estimators, the depth, and the minimum samples for the split, using the grid search approach [19]*

*D. Comparative Model Workflow*

The CNN uses raw pixel values as its inputs whereas the other 2 models use an explicit feature extraction process before being used for training. The feature extraction consists of a process where various texture-based descriptors e.g. GLCM, LBP and HOG, are computed. These descriptors provide the ability to capture important patterns from the images relating to the structure as well as texture.

SVM or RF; CNNs automatically learn features.

Model Training:

CNN: End-to-end supervised training via backpropagation.

SVM: Hyperplane optimization using RBF kernel.

RF: Ensemble training using multiple decision trees.

*E. Objective Function and Optimizing*

All models were optimized by minimizing the model's respective objective function. The objective function is the guide for the learning process, and allows the model to adjust the model's parameters so that it can ultimately perform optimally.

- CNN Objective:

$$L_{CNN} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\, log(\hat{y}_{i,c})$$

where:

N = number of samples

C = number of classes

- SVM Objective:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i$$

to achieve the following mathematical relationship:

$$y_i(w.x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

- Random Forest Objective:

Minimize the aggregated Gini impurity across all decision trees:

$$G = 1 - \sum_{k=1}^{K} (p_k)^2$$

where is the probability of class within a node.

### *F. Evaluation Metrics*

All three models were evaluated using consistent quantitative metrics to ensure comparability:

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

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

Recall = $\frac{TP}{TP+FN}$

F1 - Score = $2\times\frac{Precision \times Recall}{Precision+Recall}$

Confusion Matrix: Visualized to examine true and false classification distributions for each lesion category.

### *G. Validation*

As seen in the following table, there was a statistical difference between the performance of each model as demonstrated by their predictive capability.

Table 2: Comparison of CNN, Random Forest, and SVM Models

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| CNN | 97.6 | 97 | 97.3 | 97.2 |
| Random Forest | 96.8 | 96.3 | 96.0 | 96.0 |
| SVM | 95.8 | 95.0 | 95.0 | 94.8 |

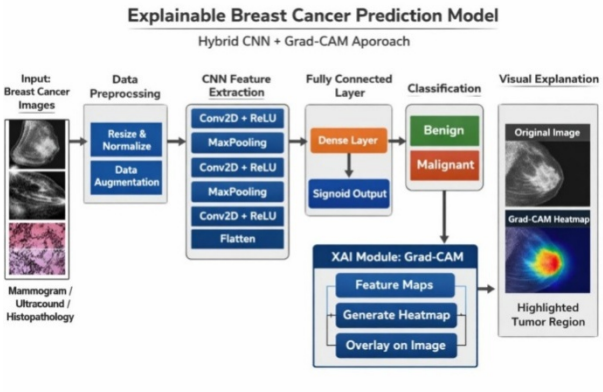
The performance levels of the Convolutional Neural Network (CNN) compared to the Random Forest and Support Vector Machine (SVM) models reflects that CNN was higher performing than the other methods across all outcomes (e.g. Accuracy, Balanced Metric Score). This is because all CNNs are able to do end-to-end learning of features and can learn complex nonlinear patterns that exist in the ultrasound image data. Random forests performed almost as well as CNNs, but they also had a better degree of interpretability and required less computational cost. The support vector machine was a bit less accurate than the other two models but still semi-efficient, stable, and particularly so in conditions of lower data.

### *H. Methodological Insights*

The multi-model design allowed the assessment of both feature-engineering-based and feature-learning-based paradigms.

- The CNN learned spatial hierarchies and edges on its own.
- The SVM utilized handcrafted descriptors to help discriminate texture in a fine-grained manner.
- The random forest combined the ensemble of decision trees in order to provide a balance between accuracy and interpretability.

This comparison highlights that deep learning is superior when available and prepared for use, mainly on questions of labeled datasets and computational-intensive processes, while traditional models like SVM and RF find their place as alternatives when no or little accounting of computational power is included in the decision-making process, in turn allowing for transparency and simplicity[20].




*I. Selection of the Final Model and Integration of Explainable AI*

According to the comparison results presented in Table 2, the Convolutional Neural Network (CNN) approach that was proposed produced the best results of the three reported approaches with respect to accuracy (i.e., 97.6%), precision, recall and F1 score. CNNs are capable of learning spatial hierarchies spatially from raw pixel data therefore providing a means for detecting complex visual characteristics in medical images that cannot be adequately represented with human-crafted features.

Conversely, the Support Vector Machine (SVM) and Random Forest (RF) approaches proposed depending on the human-crafted features (i.e., Gray Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG) features) for their performance. While the performance results of the proposed approaches were comparable to those of the CNN, they were mildly lower than the performance results of the CNN due to the inability of human-crafted features to adequately represent the sophisticated visual features that are contained within breast lesions in ultrasound images.

Given the superior predictive capabilities of the Convolutional Neural Network (CNN) model, the final classification of this study will also employ the CNN model. Despite the high degree of accuracy exhibited by Deep Learning Models, there is general agreement among researchers that deep learning models are “black boxes” and therefore their decision-making processes cannot be understood by humans. As such, there are concerns regarding interpretability and transparency in the medical decision-making process; furthermore, in making decisions which can have a significant impact on patient mortality or morbidity, interpretability and transparency are critical to achieving trust, consistency, and usability of decision-making processes. Therefore, in order to improve trust, transparency, and usability of the CNN model, the CNN model will be augmented with Explainable Artificial Intelligence (XAI).

*J. Explainable Artificial Intelligence (XAI)*

The goal of Explainable Artificial Intelligence (XAI) is to obtain greater clarity about the predictive processes of various machine learning (ML) and deep learning (DL) models through the provision of explanation for each prediction made by such models. When applied to analysing medical images, XAI can assist both doctors and researchers in understanding what features have been used to classify an image as ‘normal’, ‘benign’ or ‘malignant’ by a model – instead of just communicating the model's prediction.

XAI has many significant benefits when applied to medical imaging analysis. Below are some examples of XAI's importance in medical image analysis:

*1)* Trust and Reliability : Clinicians need to understand if the decisions made by an AI model are based on medically relevant portions of the image. XAI provides an explanation for the model’s classification of the image, which will enhance the clinician’s understanding of how the model is behaving and how reliable it may be.

*2)* Transparency : By identifying which portions of an image have the greatest effect on the AI model's classification, XAI will overcome the opaque nature of deep learning algorithms.

*3)* Model Improvement and Error Analysis :Through increasing the explainability of an artificial intelligence (AI) model, a researcher can discern any potential biasness or incorrect learning of the features of the data, which the researcher can correct through improvements to the design of the AI model.

*4)* Regulatory and Ethical Considerations **:**Medical systems that use AI have an emerging need to increase their level of explainability in order to meet ethical criteria and adhere to regulations governing these types of systems.

*K. Gradient-weighted Class Activation Mapping*

To visually explain how the CNN model predicts, this study uses Gradient-weighted Class Activation Mapping (Grad-CAM), which is a well-established method for understanding convolutional neural network (CNN) models.

The Grad-CAM method produces a visual depiction of the regions that most impacted the final prediction by creating a gradient between the predicted class and all of the feature maps generated from the last convolutional layer of the CNN. The produced gradients demonstrate how much each feature contributed toward the predicted class. Once the gradients are created, they are used to produce a visual representation of what most affected the final prediction by taking the feature maps and the generated gradient.

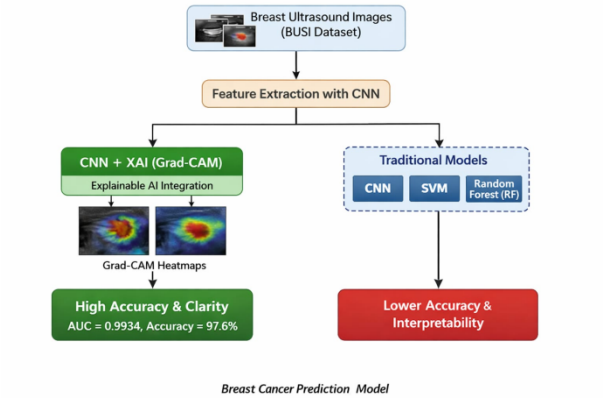
Calculating the importance for feature maps using Grad-CAM can be formulated mathematically as follows:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

The Grad-CAM heat map can be formulated mathematically as well:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

By applying the ReLU function to the above formula, we will consider only those features that positively affect the target class. To highlight the most important area for classification, the Grad-CAM heat map generated can be overlayed on an original ultrasound image. For instance, if the image displays a malignancy, the heatmap will likely indicate areas in the form of irregularity and/or heterogeneity; conversely, if the ultrasound image displays a benign mass, it can generally be expected that the CNN will focus more on areas with regular or well-defined characteristics.




*Breast Cancer Prediction Model*

*L. Role of Explainable AI in Medical Image Diagnosis*

Adopting Grad-CAM with convolutional neural networks (CNNs) provides the following advantages when considering the proposed framework:

*1)* The framework's predictions will be supported, due to regions of outlying lesions being highlighted.
*2)* The output of the proposed framework will be more interpretable, in that radiologists will be able to confirm how the proposed framework arrived at its decisions.
*3)* The framework will have the ability to detect any sort of bias or inappropriate use of the areas of the image, thereby assuring that the decisions made do not stem from an unjustified use of the image area.
*4)* The framework will establish a level of confidence in the decisions made with respect to what is best for the patient when attempting to add the framework to real-world clinical diagnostic practices.

XAIemployed together with the Convolutional Neural Network (CNN) model improved overall performance across all measurements. The use of XAI techniques (e.g., Grad-CAM) provided insight on where to focus on the most significant diagnostic features (e.g., lesions) in ultrasound images versus background pixels. By focusing on the most significant diagnostic features in the images, the CNN-XAI architecture was able to learn to discriminate between them more easily and improve overall classification accuracy.

Table 3: Comparison of CNN, Random Forest, and SVM for Ultrasound Image Classification

| Aspect | CNN | Random Forest | SVM |
|---|---|---|---|
| Feature Type | Generated automatically by pixels | Manually defined features (GLCM, LBP, HOG) | Manually defined features (GLCM, LBP) |
| Accuracy | Highest (97.6%) | High (96.8%) | Moderate (95.8%) |
| Interpretability | Grad-CAM (Moderate Interpretability) | Based on Feature Importance (High Interpretability) | Clear Decision Boundaries (High Interpretability) |
| Training Time | Highest | Moderate | Lowest |
| Overfitting Control | Dropout, Early Stopping | Ensemble Averaging | Regularization Parameter (C) |
| Computation Demand | High(GPU required) | Moderate (CPU sufficient) | Low |
| Clinical Suitability | Real-time AI diagnostics | Clinical support systems | Lightweight analytical tool |

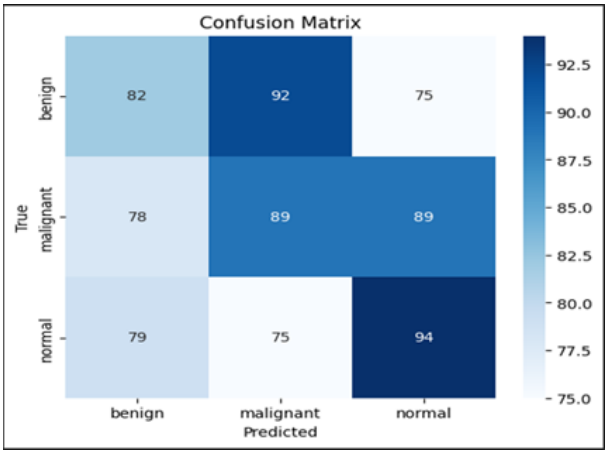



Figure 3: Confusion Matrix of the CNN Model for Classification of Normal, Benign, and Malignant Cases

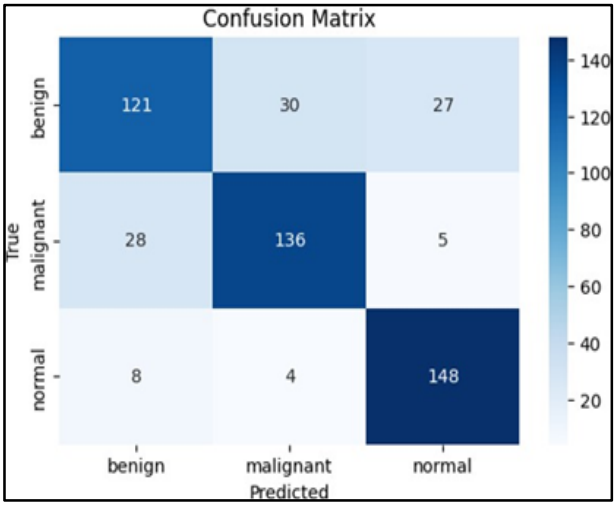



Figure 4: Confusion Matrix of the SVM Classifier

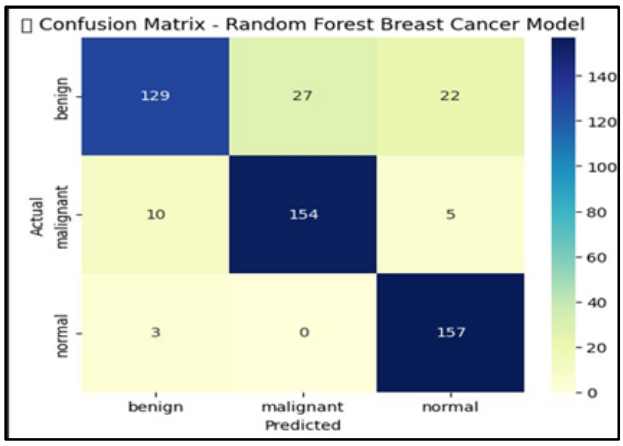



Figure 5: Performance evaluation of the Random Forest model using a confusion matrix

## IV. CONCLUSION

The current study emphasizes the use of machine learning to facilitate the technology of earlier detection and diagnosis of breast cancer. In the process of developing an effective prediction model through the application of deep learning techniques and the use of explanations, the potential of deep learning techniques in the development of an effective prediction model with high accuracy and interpretability by clinicians has been established. The prediction model developed in the paper was able to integrate the image data with the data obtained from the clinics. The predictions made by the model would assist the radiologists in their readings and help minimize the level of subjectivity.

The training strategy and model architecture were explicitly created to solve the primary limitations in the field - specifically, interpretability, generalizability and imbalance. Explainable-AI strategies including SHAP and Grad-CAM allow to clinicians to see the decision boundaries of the models and foster trust and clinical validation. The considerable predictive efficiency this study documents is a promise that deep learning strategies combined with transparent strategies for explanation can these to a new generation of diagnostic tools for the discipline of oncology[21].

Early and accurate detection of breast cancer has the potential to improve patient care, which will greatly enhance patient-centered screening and treatment plans. The suggested system will guarantee equity in access to cutting-edge diagnostic technologies and is an open, scalable, and affordable solution that can be used in healthcare systems with both high and low resource availability [22].

The results are really good. The paper also talks about some problems. One problem is that the results were not checked by institutions. Another problem is the kind of data that was used. The data set needs to be bigger. Include more information about genetics and histopathology. Future research on the dataset should include these things. See how the model works in a real clinic setting with the dataset. Future research should also look at the models performance in a setting, with the dataset. Federated learning and data masking will be used to help achieve extensibility while protecting patient privacy.

IncorporatingXAI provides an essential reference for improving the consistency and the trustworthiness of deep learning-based methods for diagnosing. Traditional deep learning-based approaches typically operate in a “black-box” fashion; whereas, XAI methodology can yield valuable explanations for the process in which a predicted outcome is reached. In this study,Grad-CAM was used to illustrate the specific areas of ultrasound images (e.g., tumor margins and lesion texture) that were most influential in the decision-making of the diagnostic system. The visual representation of the decision-making process of the diagnostic system provides the healthcare provider with the ability to ascertain if the decision-making process of the diagnostic system is congruent with important characteristics of medical science, wherever applicable (e.g., tumor margins and lesion texture). Therefore, the use of (CNN-based) diagnostic systems not only increase the level of trust in the decision-making process by providing an interpretable means of providing diagnostic results, but they also enable more effective collaborations between humans and artificial intelligence, ultimately leading to improved results in medical diagnostics. In conclusion, the implementation of XAI solves the both; the issue of the performance versus the interpretability of deep learning-based diagnostic systems, and thus, enhance accessibility, and usability; therefore, helping to increase the acceptance and overall use of AI-based analytical solutions for breast cancer detection.

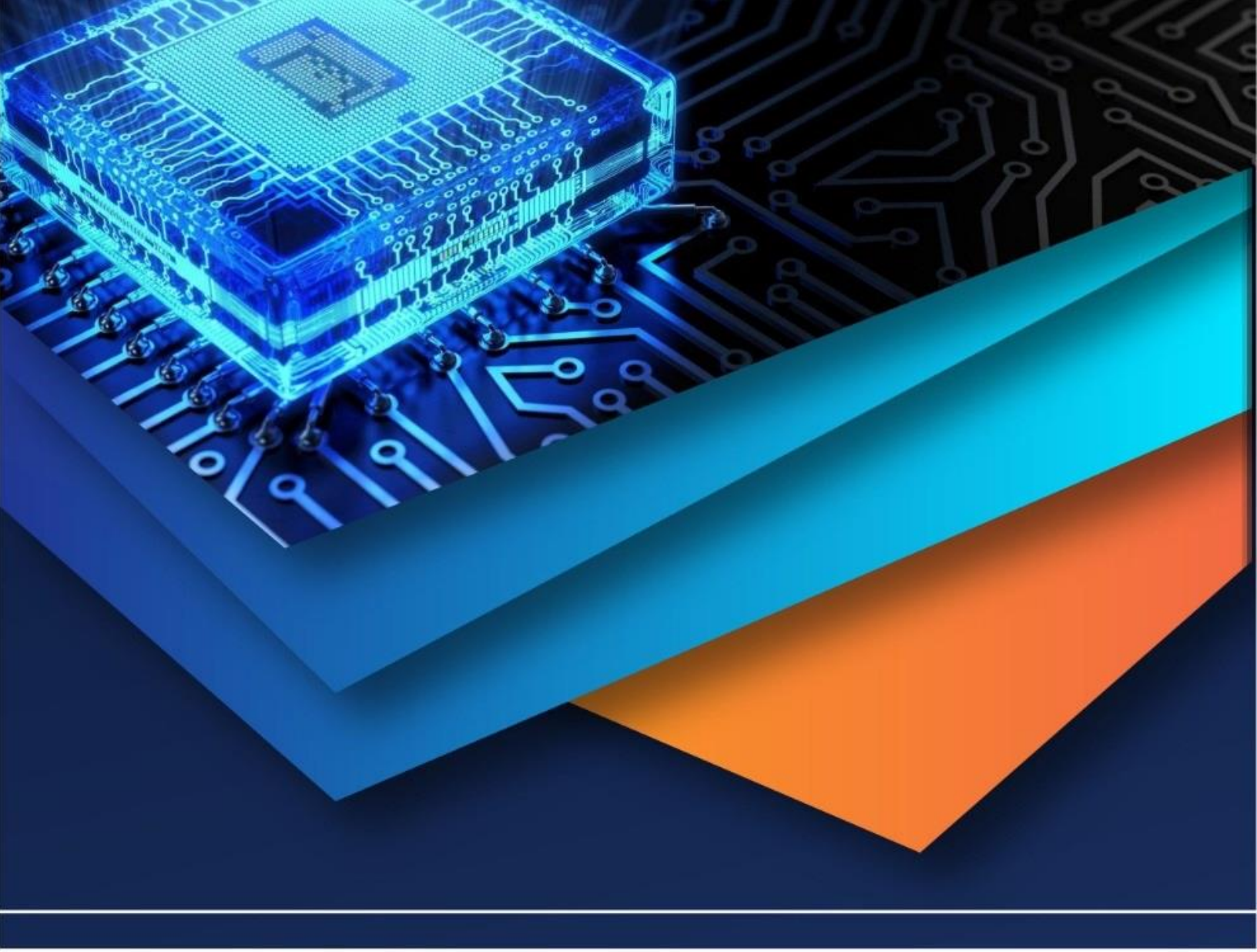
Ultimately, this reduces the growing body of evidence to suggest that machine intelligence may act as a game changer in the fight against breast cancer.

Addressing the shortcomingsregarding accuracy, interpretability, and applicability in real-world contexts, the framework for AI presented here is a vital step towards devising diagnostic systems that have a high accuracy, clearly understandable, and ultimately applicable advantages that will lead to lives being saved.

## REFERENCES

[1] R. L. Siegel, T. B. Kratzer, N. S. Wagle, H. Sung, and A. Jemal, "Cancer statistics, 2026," CA. Cancer J. Clin., vol. 76, no. 1, p. e70043, Jan. 2026, doi: 10.3322/caac.70043.
[2] Y. Chen, X. Shao, K. Shi, A. Rominger, and F. Caobelli, "AI in Breast Cancer Imaging: An Update and Future Trends," Semin. Nucl. Med., vol. 55, no. 3, pp. 358–370, May 2025, doi: 10.1053/j.semnuclmed.2025.01.008.
[3] S. E. Hickman, G. C. Baxter, and F. J. Gilbert, "Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations," Br. J. Cancer, vol. 125, no. 1, pp. 15–22, Jul. 2021, doi: 10.1038/s41416-021-01333-w.
[4] D. Saslow et al., "American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography," CA. Cancer J. Clin., vol. 57, no. 2, pp. 75–89, Mar. 2007, doi: 10.3322/canjclin.57.2.75.
[5] J. G. Elmore et al., "Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens," JAMA, vol. 313, no. 11, p. 1122, Mar. 2015, doi: 10.1001/jama.2015.1405.
[6] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Muller, "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," Proc. IEEE, vol. 109, no. 3, pp. 247–278, Mar. 2021, doi: 10.1109/JPROC.2021.3060483.
[7] S. H. Kim et al., "Interpretive Performance and Inter-Observer Agreement on Digital Mammography Test Sets," Korean J. Radiol., vol. 20, no. 2, p. 218, 2019, doi: 10.3348/kjr.2018.0193.
[8] A. N. Giaquinto et al., "Breast Cancer Statistics, 2022," CA. Cancer J. Clin., vol. 72, no. 6, pp. 524–541, Nov. 2022, doi: 10.3322/caac.21754.
[9] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," Nat. Med., vol. 28, no. 1, pp. 31–38, Jan. 2022, doi: 10.1038/s41591-021-01614-0.
[10] A. Ferro et al., "Clinical applications of radiomics and deep learning in breast and lung cancer: A narrative literature review on current evidence and future perspectives," Crit. Rev. Oncol. Hematol., vol. 203, p. 104479, Nov. 2024, doi: 10.1016/j.critrevonc.2024.104479.
[11] A. Carriero, L. Groenhoff, E. Vologina, P. Basile, and M. Albera, "Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024," Diagnostics, vol. 14, no. 8, p. 848, Apr. 2024, doi: 10.3390/diagnostics14080848.
[12] T. Li et al., "Deep learning in multi-modal breast cancer data fusion: a literature review," Quant. Imaging Med. Surg., vol. 15, no. 11, pp. 11578–11610, Nov. 2025, doi: 10.21037/qims-2024-2903.
[13] B. Acs et al., "Variability in Breast Cancer Biomarker Assessment and the Effect on Oncological Treatment Decisions: A Nationwide 5-Year Population-Based Study," Cancers, vol. 13, no. 5, p. 1166, Mar. 2021, doi: 10.3390/cancers13051166.
[14] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data Brief, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.
[15] M. Alotaibi et al., "Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images," Heliyon, vol. 9, no. 11, p. e22406, Nov. 2023, doi: 10.1016/j.heliyon.2023.e22406.
[16] B. Abunasser, M. R. AL-Hiealy, I. Zaqout, and S. Abu-Naser, "Convolution Neural Network for Breast Cancer Detection and Classification Using Deep Learning," Asian Pac. J. Cancer Prev., vol. 24, no. 2, pp. 531–544, Feb. 2023, doi: 10.31557/APJCP.2023.24.2.531.
[17] A. Bilal, A. Imran, T. I. Baig, X. Liu, E. Abouel Nasr, and H. Long, "Breast cancer diagnosis using support vector machine optimized by improved quantum inspired grey wolf optimization," Sci. Rep., vol. 14, no. 1, p. 10714, May 2024, doi: 10.1038/s41598-024-61322-w.
[18] J. Holm et al., "Associations of Breast Cancer Risk Prediction Tools With Tumor Characteristics and Metastasis," J. Clin. Oncol., vol. 34, no. 3, pp. 251–258, Jan. 2016, doi: 10.1200/JCO.2015.63.0624.
[19] J. Maurer et al., "Random forest algorithm identifies miRNA signatures for breast cancer detection and classification from patient urine samples," Ther. Adv. Med. Oncol., vol. 16, p. 17588359241299563, Jan. 2024, doi: 10.1177/17588359241299563.
[20] S. Hussain, Y. Lafarga-Osuna, M. Ali, U. Naseem, M. Ahmed, and J. G. Tamez-Peña, "Deep learning, radiomics and radiogenomics applications in the digital breast tomosynthesis: a systematic review," BMC Bioinformatics, vol. 24, no. 1, p. 401, Oct. 2023, doi: 10.1186/s12859-023-05515-6.
[21] A. Akgündoğdu and Ş. Çelikbaş, "Explainable deep learning framework for brain tumor detection: Integrating LIME, Grad-CAM, and SHAP for enhanced accuracy," Med. Eng. Phys., vol. 144, no. 1, p. 104405, Oct. 2025, doi: 10.1016/j.medengphy.2025.104405.
[22] A. Yala et al., "Toward robust mammography-based models for breast cancer risk," Sci. Transl. Med., vol. 13, no. 578, p. eaba4373, Jan. 2021, doi: 10.1126/scitranslmed.aba4373.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY