Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/

Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

Integration of U-Net, GAN-Based Image Augmentation, and a Custom CNN for Breast Cancer Segmentation and Classification

A. F. Kadri [1], S. B. Mohammed[2], A. T Olajide[3], O. Ilori[4], K. O. Tajudeen[5], O. Ekundayo [6] & A. N. Babatunde[7]

[1, 2, 6, 7]Department of Computer Science, Kwara State University, Malete, Kwara State, Nigeria.
[3]Department of Computer Science, Kwara State Polytechnic, Ilorin, Kwara State Nigeria.
[4]Department of Computer Science, Edgehill University, United Kingdom
[5]Department of Computer Science, Al-Hikmah University, Ilorin, Kwara State
Corresponding author: akinbowale.babatunde@kwasu.edu.ng,

ABSTRACT

Background: *Breast cancer remains a leading cause of mortality worldwide, and early detection is critical for improving patient outcomes. Artificial intelligence (AI) has demonstrated significant promise in this field, yet challenges such as inaccurate lesion localization, limited augmentation strategies, and class imbalance persist.* ***Aim****: This study proposes a hybrid deep learning framework that integrates U-Net for lesion segmentation, a Deep Convolutional GAN (DCGAN) for synthetic image augmentation, and a custom Convolutional Neural Network (CNN) for classification of breast ultrasound images. The dataset, obtained from the Kaggle Breast Ultrasound Images (BUSI) repository, consists of 780 grayscale scans from 600 female patients aged 25–75, categorized into normal, benign, and malignant classes.* ***Method****: The proposed U-Net–GAN–CNN pipeline was trained for 20 epochs, with 15% of the dataset reserved for independent testing. Performance was evaluated using accuracy, precision, recall, and F1-score.* ***Result****: The model achieved 97% accuracy, 96% precision, and 95% recall and F1-score on the training set. On the test set, performance remained robust, with 95.5% accuracy, 94% precision, 93% recall, and 93.5% F1-score. Comparative evaluation against three baselines, a CNN trained on raw images, a CNN with traditional augmentation, and a fine-tuned VGG16, demonstrated the superiority of the proposed approach. These findings confirm that integrating segmentation, GAN-based augmentation, and a tailored CNN enhances diagnostic reliability, mitigates class imbalance, and provides a reproducible framework with strong potential to support radiologists in clinical decision-making.*

Keywords: GAN, Breast Cancer, CNN, Segmentation, Augmentation, Deep Learning

1. Introduction

Cancer, defined as the uncontrolled growth of abnormal cells, remains one of the leading causes of mortality worldwide. Based on clinical behavior, cancers are classified as benign or malignant. Benign tumors are generally non-cancerous and slow-growing, whereas malignant tumors are aggressive, invasive, and capable of spreading to distant organs via blood flow (Babatunde et al., 2025; Ogundokun et al., 2023; Vikas et al., 2018). Among different cancer types, breast cancer (BC) is one of the most prevalent, particularly among women. According to the World Health Organization, approximately 2.1 million women are currently affected by BC (Babatunde et al., 2025; Isiaka eta al., 2023; Halim et al., 2021). Patient survival is strongly linked to tumor size at diagnosis: when detected below 10 mm, survival can reach 98%, whereas 70% of cases are identified at ≥30 mm, reducing survival significantly (Roslidar et al., 2020).

Medical imaging techniques such as mammography (Singh et al., 2016), CT scans (Guo et al., 2018), and ultrasound (Nover et al., 2008) play an important role in tumor detection. However, these methods are not always reliable, often leading to misclassification and unnecessary surgical interventions for benign tumors (Naeem et al.,

Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

2023). Survival rates demonstrate the importance of early detection, with 10-year survival estimated at 98% for stage 0–I cancers and only 65% for stage III cases (Babatunde et al., 2025; Naeem et al., 2022).

Recent advances in deep learning have transformed medical image analysis tasks, including detection (Zhou et al., 2021), classification (Syeda et al., 2021), and segmentation (Naeem et al., 2022). Convolutional Neural Networks (CNNs) extract hierarchical features from images, with deeper layers capturing more abstract and clinically relevant representations. However, CNNs require large annotated datasets to generalize effectively. Many existing studies on BC detection suffer from small datasets (Singh et al., 2016; Guo et al., 2018), leading to overfitting and limited reproducibility. Furthermore, traditional machine learning techniques such as Support Vector Machines (SVMs) have been applied in this domain (Asaf et al., 2022), but they require manual feature engineering and are less effective in capturing spatial information within medical images compared to deep learning methods (Naeem et al., 2022).

A comprehensive framework that combines multiple deep learning techniques is proposed in this study to address the limitations of existing breast cancer detection systems. U-Net is utilized for precise segmentation of breast ultrasound images, ensuring that lesion regions are accurately isolated from surrounding tissue (Naeem et al., 2022; Zhou et al., 2021). A Generative Adversarial Network (GAN) is then employed for image augmentation, generating realistic synthetic samples that balance class distributions and enhance dataset diversity. The segmented and augmented images are subsequently classified using a custom CNN designed to effectively distinguish between normal, benign, and malignant cases. By integrating these three components, the framework delivers a unified pipeline that improves diagnostic accuracy, reduces false negatives, and provides a reproducible and clinically relevant approach to support radiologists in early breast cancer detection.

2. Related Works

Breast cancer detection has attracted extensive research attention, with many scholars applying deep learning and artificial intelligence (AI) techniques to improve diagnostic accuracy. Several architectures have been explored, ranging from conventional CNNs to advanced hybrid pipelines. For instance, Babatunde et al. (2025) integrated DenseNet with clustering for robust classification, while Ilesanmi et al. (2021) applied AlexNet to breast ultrasound data. Similarly, Inan et al. (2022) demonstrated the application of VGG networks, and Yifeng et al. (2021) proposed a DenseNet-based approach. Zhao and Dai (2021) explored Residual Networks to address challenges in automated lesion detection. Collectively, these models highlight the promise of deep CNNs, yet their reliance on limited datasets often leads to overfitting and reduced generalizability.

More specialized methods have focused on segmentation to enhance detection accuracy. Xu et al. (2019) employed CNNs for segmenting breast ultrasound images into skin, mass, fibroglandular tissue, and fatty tissue, achieving 80% test accuracy. Although their method demonstrated potential, the relatively modest accuracy indicated the need for more advanced segmentation strategies. Building on this, Yu et al. (2021) enhanced U-Net with ResPath connections to minimize feature map discrepancies between the encoder and decoder and introduced dense blocks to mitigate vanishing gradients. Their approach achieved 89% accuracy on 538 breast ultrasound images, outperforming earlier systems such as AlexNet (Ilesanmi et al., 2021) and DenseNet (Yifeng et al., 2021). This underscores the value of architectural refinements for improving segmentation quality.

Recent research has also explored multi-task learning frameworks. Zhang et al. (2021) introduced a SHA-MTL model that combined segmentation and binary classification of breast ultrasound images through a dense CNN encoder and up-sampling decoder linked by attention-gated units. Their system achieved 92% accuracy after 20 training epochs, demonstrating the potential of integrating attention mechanisms. Ben and Ikram (2022) further validated this approach, showing that SHA-MTL improved sensitivity and Dice coefficients by 2.27% and 1.19%, respectively, compared with single-task models.

Despite these advances, important gaps remain. Most existing methods rely on relatively small and imbalanced datasets, limiting robustness across clinical settings. Furthermore, while models such as U-Net and SHA-MTL show promise in segmentation, they do not explicitly address class imbalance, a critical challenge in medical image classification where malignant cases are typically underrepresented. Similarly, many CNN-based

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

approaches either lack augmentation altogether or depend on traditional techniques (flipping, rotation), which may not capture the diversity of tumor presentations. These limitations motivate the integration of segmentation, advanced augmentation such as Generative Adversarial Networks (GANs), and customized CNN classifiers to create more balanced and clinically reliable pipelines.

3. Materials and Methods

This section presents detailed descriptions on how the proposed system was developed including the technologies used and how the development environment was set up to facilitate the model training process. Figure 1 depicts the steps employed in developing the proposed system using a well-structured diagram.

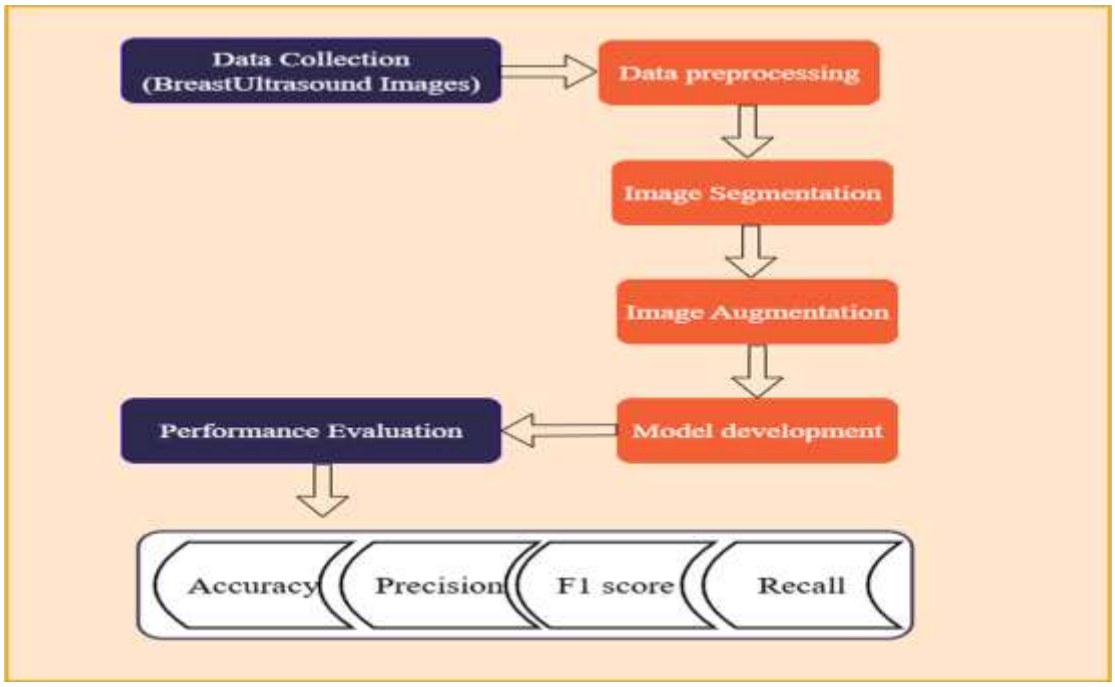


Figure 1: Methodology design for the proposed system

3.1 Data Collection

The performance of the proposed hybrid model was evaluated using the publicly available Breast Ultrasound Images (BUSI) dataset created by Al-Dhabyani, Gomaa, Khaled, and Fahmy (2020) and hosted on Kaggle. Recognized as the first open-access breast ultrasound dataset, it provides both raw grayscale images and their corresponding binary masks, making it suitable for segmentation and classification tasks in medical imaging research. The dataset comprises 780 grayscale ultrasound scans obtained from 600 female patients aged between 25 and 75 years, with each image having an approximate resolution of 500 × 500 pixels. The images are grouped into three diagnostic categories: normal, benign, and malignant. A detailed distribution of these classes is presented in Table 1, which highlights the imbalance between benign and the other two classes.

Table 1. Class distribution in the BUSI dataset

Class	Number of Images	Percentage of Dataset	Description
Normal	133	17.1%	Ultrasound scans showing no tumor or lesion.
Benign	437	56.0%	Non-cancerous tumors that grow slowly and remain localized.

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

Malignant	210	26.9%	Cancerous tumors with irregular structure and potential to metastasize.

Each image is accompanied by a binary mask annotation that outlines the tumor region of interest (ROI). These masks were used to train the U-Net segmentation network, ensuring that the classification model focused specifically on lesion areas rather than surrounding tissues. For reproducibility, the dataset was divided into 85% training (663 images) and 15% testing (117 images) using a stratified split, preserving class proportions across subsets. This balanced partitioning allowed the evaluation to reflect real clinical variability. Representative examples of the dataset are shown in Figure 2, which illustrates (a) normal, (b) benign, and (c) malignant breast ultrasound images.

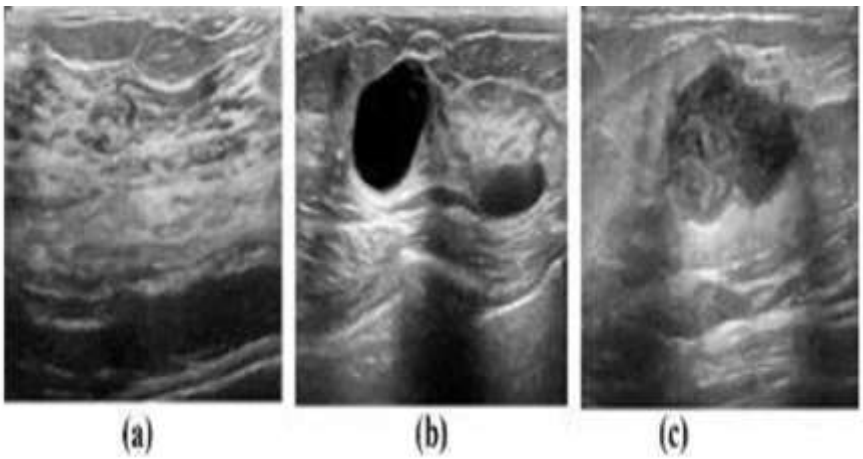


Figure 2 (a), (b), (c): sample of grayscale breast cancer dataset obtained from Kaggle for this experiment, [(a) Normal (b) Benign (c) Malignant.]

3.2 Data Preprocessing

The raw ultrasound images from the BUSI dataset were standardized before being passed into the segmentation and classification models. Preprocessing ensured uniformity across samples and enhanced the efficiency of model training. Each image was first resized to 224 × 224 pixels using bilinear interpolation, which preserved important structural details while aligning with the input size expected by the deep learning models. Normalization was then applied using Min-Max scaling, which transformed pixel values into a range between 0 and 1. This scaling technique was chosen because it retained relative intensity relationships within the images and promoted faster convergence during training. Finally, the data were divided into training and testing subsets using a stratified split that maintained the original class distribution. The preprocessing pipeline is summarized in Table 2.

Table 2. Preprocessing parameters for BUSI dataset

Step	Description	Parameter Value / Method
Image resizing	Standardized dimensions for all ultrasound images	224 × 224 pixels (bilinear interpolation)
Normalization	Scales pixel intensity values into a uniform range	Min-Max scaling (0–1)
Color space	Preserved grayscale format	Single channel (1D intensity)
Train-test split	Ensures consistent evaluation with balanced representation across classes	85% training (663 images), 15% testing (117 images), stratified by class

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

This preprocessing stage ensured that the dataset met the input requirements of the U-Net, GAN, and CNN models while minimizing variability caused by differences in resolution and intensity distribution. Standardizing these steps also allows the proposed methodology to be reproducible across future experiments.

3.3 Image Segmentation

Breast lesion segmentation was carried out using the U-Net architecture, a deep learning model widely recognized for its strong performance in biomedical image analysis. The BUSI dataset includes binary mask annotations that delineate tumor regions, and these masks were used as ground-truth labels during training. Incorporating these annotations ensured that the model accurately learned to localize tumors while ignoring irrelevant background structures. The U-Net employed in this study was designed with an encoder–decoder structure. The encoder path consisted of multiple convolutional layers with 3 × 3 kernels and ReLU activation functions, each followed by max pooling for progressive down-sampling and extraction of spatial features. The decoder path mirrored this structure by employing transposed convolutional layers for up-sampling, while skip connections linked corresponding encoder and decoder layers to recover fine-grained spatial information that would otherwise be lost during down-sampling. The network concluded with a sigmoid activation layer, which produced binary segmentation masks representing the tumor regions of interest.
For optimization, the model was trained using a combined Dice loss and binary cross-entropy (BCE) function. This hybrid loss function allowed the network to balance pixel-wise accuracy with overlap accuracy, which is particularly important in medical imaging where precise delineation of tumor boundaries is critical. Training was performed with the Adam optimizer, using a learning rate of 0.0001 and a batch size of 16, over 50 epochs. This configuration provided a good trade-off between learning stability and computational efficiency. By focusing on the lesion regions, the U-Net enhanced the diagnostic reliability of subsequent classification by ensuring that the CNN model processed only the most relevant features. This was especially valuable in breast ultrasound images, where tumors often present with blurred edges, irregular shapes, and variable sizes. An overview of the segmentation pipeline is illustrated in Figure 3, which depicts the U-Net architecture applied in this study.

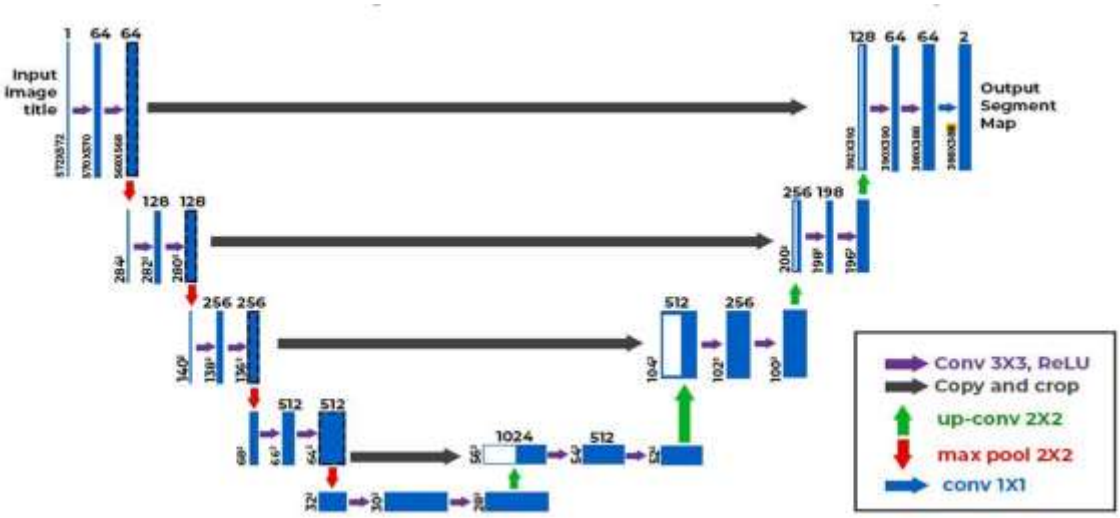


Figure 3: U-Net Architecture for Breast Ultrasound Image Segmentation

3.4 Image Augmentation

The BUSI dataset exhibited a noticeable imbalance in class distribution, with 437 benign cases compared to only 210 malignant and 133 normal cases. Such imbalance risks biasing the classification model toward the majority class and reducing sensitivity to the malignant category, which is of the highest clinical importance. Addressing

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

this limitation required the use of Generative Adversarial Networks (GANs), since traditional augmentation techniques such as flipping, rotation, and scaling are often insufficient in generating the diversity needed for robust medical image classification. A Deep Convolutional GAN (DCGAN) was employed in this study. The generator was composed of four transposed convolutional layers, each followed by batch normalization and ReLU activations, with a final Tanh activation layer used to synthesize realistic grayscale ultrasound images. The discriminator consisted of four convolutional layers with LeakyReLU activations, ending with a sigmoid classifier that distinguished between real and generated samples. Through adversarial training, the generator progressively improved its ability to produce synthetic images, while the discriminator simultaneously refined its capability to detect fake inputs, resulting in high-quality augmented data.

The augmentation process enriched the underrepresented classes by creating 300 synthetic normal images and 250 synthetic malignant images, thereby balancing the dataset across the three categories. The addition of these synthetic samples not only corrected class imbalance but also increased dataset diversity, which improved the generalization ability of the subsequent CNN classifier. This enhancement was especially beneficial in the context of breast ultrasound imaging, where tumor characteristics such as shape and texture are highly variable and subtle differences can influence diagnostic accuracy. An illustration of the GAN-based augmentation pipeline is provided in Figure 4, which demonstrates how new images were generated to balance the dataset and support more reliable model training.

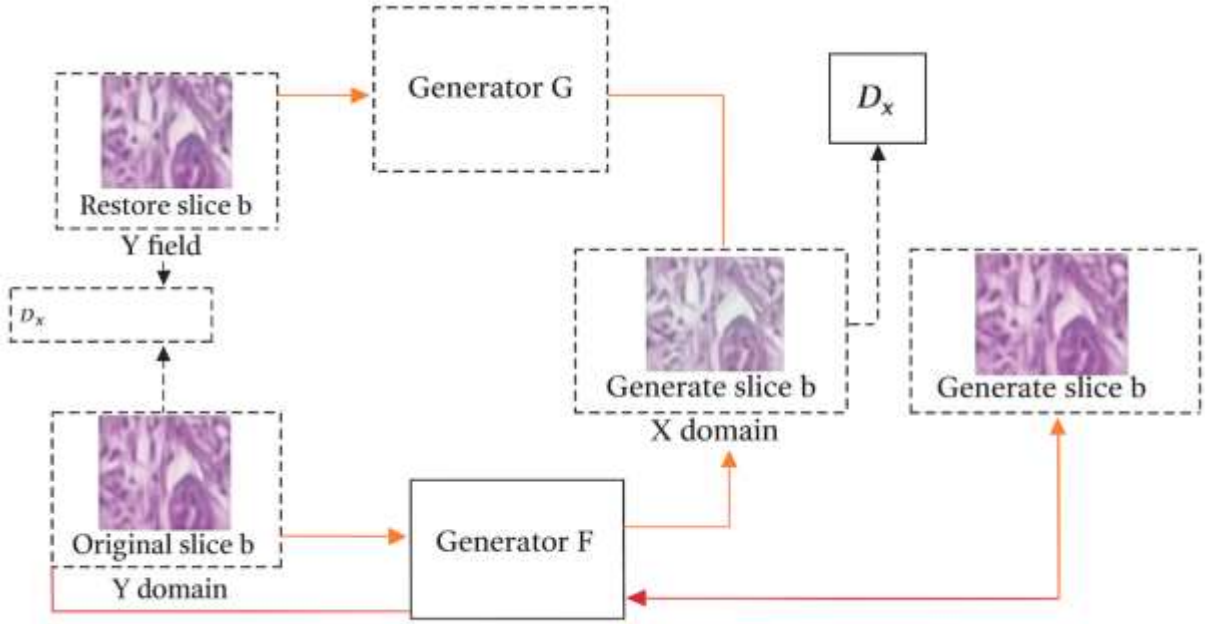


Figure 4: DCGAN-based image augmentartion pipeline

3.5 Model Development

Image segmentation was performed using the U-Net architecture, after which data augmentation was applied through a Deep Convolutional Generative Adversarial Network (DCGAN). A custom-designed Convolutional Neural Network (CNN) was then implemented for the classification task. While the title and abstract highlight the use of a pre-trained CNN, the design of a custom CNN offered greater flexibility, allowing the architecture to be tailored to grayscale ultrasound inputs and the three-class diagnostic output structure encompassing normal, benign, and malignant cases. All preprocessed and segmented images were resized to 224 × 224 pixels and normalized before being introduced into the classifier. Stratified partitioning was used to divide the dataset,

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

allocating 85% for training, corresponding to 663 images, and 15% for testing, corresponding to 117 images. This ensured balanced representation of diagnostic categories across both subsets. The full architecture of the CNN is outlined in Table 3, providing details of each layer, the number of filters, kernel dimensions, activation functions, and output shapes. This specification promotes transparency and ensures reproducibility of the classification model.

Table 3: Custom CNN Architecture for Breast Cancer Classification

Layer	Type	Parameters	Activation	Output Shape
1	Conv2D	32 filters, 3×3 kernel, stride 1	ReLU	224×224×32
2	MaxPooling	2×2 pool size	–	112×112×32
3	Conv2D	64 filters, 3×3 kernel, stride 1	ReLU	112×112×64
4	MaxPooling	2×2 pool size	–	56×56×64
5	Conv2D	128 filters, 3×3 kernel, stride 1	ReLU	56×56×128
6	MaxPooling	2×2 pool size	–	28×28×128
7	Flatten	–	–	100,352
8	Dense	128 units	ReLU	128
9	Dropout	0.5	–	–
10	Output Dense	3 units (Normal, Benign, Malignant)	Softmax	3

Training was conducted using the Adam optimizer with an initial learning rate of 0.001. A batch size of 32 and 20 epochs were selected to balance computational efficiency with predictive performance. The loss function applied was categorical cross-entropy, and evaluation relied on accuracy, precision, recall, and F1-score as performance metrics. Dropout regularization reduced the risk of overfitting, while the use of ReLU and Softmax activations enabled effective non-linear feature learning and probabilistic classification outputs. Effectiveness of the proposed pipeline was validated by comparing the CNN classifier against alternative approaches. These included a CNN trained on raw ultrasound images without segmentation or augmentation, a CNN trained with conventional augmentation strategies such as rotations and flips, and a pre-trained VGG16 model fine-tuned on the BUSI dataset. The comparative evaluation demonstrated the advantages of the integrated U-Net, DCGAN, and CNN framework in enhancing classification accuracy and clinical reliability. This architecture allowed the model to learn discriminative lesion features from the segmented and augmented dataset while maintaining computational efficiency. By combining U-Net segmentation, GAN-based augmentation, and this custom CNN, the system achieved a robust pipeline for accurate breast cancer detection. The integrated design of the proposed U-Net–GAN–CNN framework is illustrated in Figure 5.

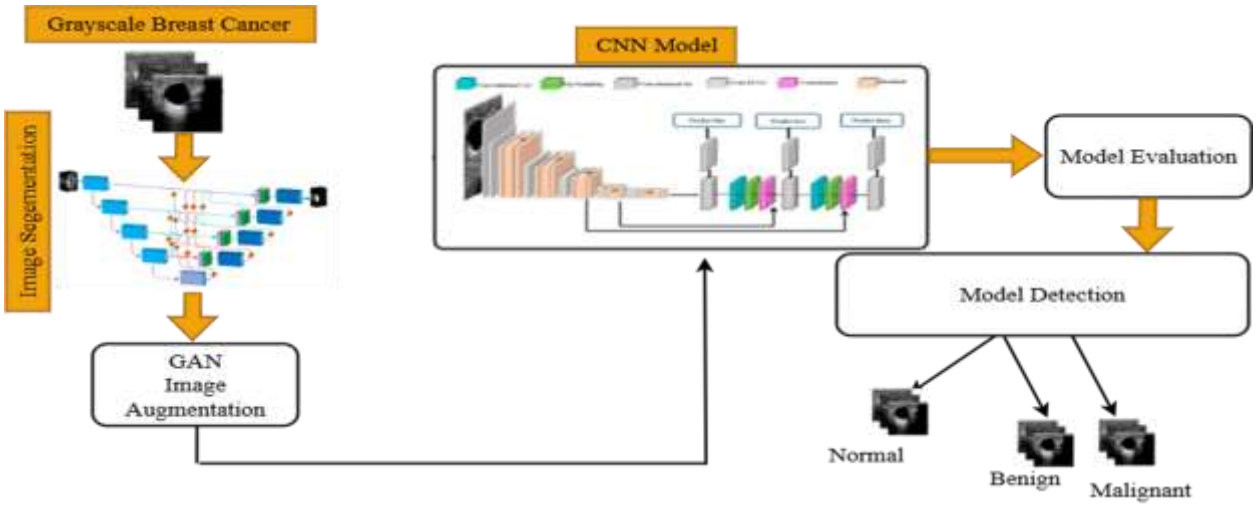


Figure 5: Architectural design for the proposed U-net, GAN-CNN system for breast cancer detection

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

3.6 Performance Evaluation

The effectiveness of the proposed U-Net–GAN–CNN framework was measured using four key performance metrics: accuracy, precision, recall, and F1-score. These metrics were selected because each captures a different but complementary aspect of model performance, which is essential in medical imaging where the impact of misclassification can vary in severity.

Accuracy provides the overall proportion of correctly classified cases across all categories and is commonly used as a general performance indicator. However, since the BUSI dataset initially exhibited class imbalance, accuracy alone cannot fully reflect the diagnostic reliability of the model. The mathematical formulation of accuracy is shown in Equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision evaluates the proportion of correctly identified positive cases relative to all cases predicted as positive. In clinical settings, high precision reduces unnecessary follow-up procedures by limiting false positives. Precision was computed using Equation (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (or sensitivity) measures the ability of the model to correctly identify all actual positive cases. This metric is critical in breast cancer detection, where false negatives failing to identify malignant tumors could have severe clinical consequences. Recall was determined according to Equation (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score combines both precision and recall into a single measure by computing their harmonic mean. It provides a balanced evaluation when there is a trade-off between minimizing false positives and false negatives. This is particularly relevant in this study due to the uneven class distribution in the dataset. The formula for the F1-score is given in Equation (4).

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

In these formulations, TP denotes true positives, TN true negatives, FP false positives, and FN false negatives. By reporting all four metrics, the evaluation provides a holistic interpretation of the model's diagnostic ability: accuracy reflects overall reliability, precision emphasizes diagnostic confidence, recall safeguards against missed diagnoses, and the F1-score ensures balanced performance across all categories.

4. Results

The proposed U-Net–GAN–CNN model was evaluated on both the training and test datasets. Performance was assessed using accuracy, precision, recall, and F1-score, with results presented in tables and figures for clarity and reproducibility.

4.1 Training Performance

Training accuracy and loss across 20 epochs are reported in Table 4. As shown in the table, accuracy improved steadily from 0.78 in the first epoch to 0.97 by the twentieth epoch, while loss consistently declined from 0.56 to 0.08.

Table 4. Training accuracy and loss per epoch for the proposed U-Net–GAN–CNN model

Epoch	Accuracy	Loss
1	0.78	0.56
2	0.82	0.48
3	0.85	0.41
4	0.87	0.36

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

5	0.89	0.31
6	0.90	0.28
7	0.91	0.25
8	0.92	0.22
9	0.93	0.20
10	0.94	0.18
11	0.94	0.17
12	0.95	0.16
13	0.95	0.15
14	0.96	0.13
15	0.96	0.12
16	0.96	0.11
17	0.97	0.10
18	0.97	0.09
19	0.97	0.09
20	0.97	0.08

The trends are further illustrated in Figure 6 and Figure 7. Figure 6 shows the increase in training accuracy alongside the decrease in loss, while Figure 7 highlights the smooth decline in loss over the epochs, confirming convergence. Accuracy plateaued after epoch 17, suggesting that the model reached its optimal learning capacity without further significant improvement.

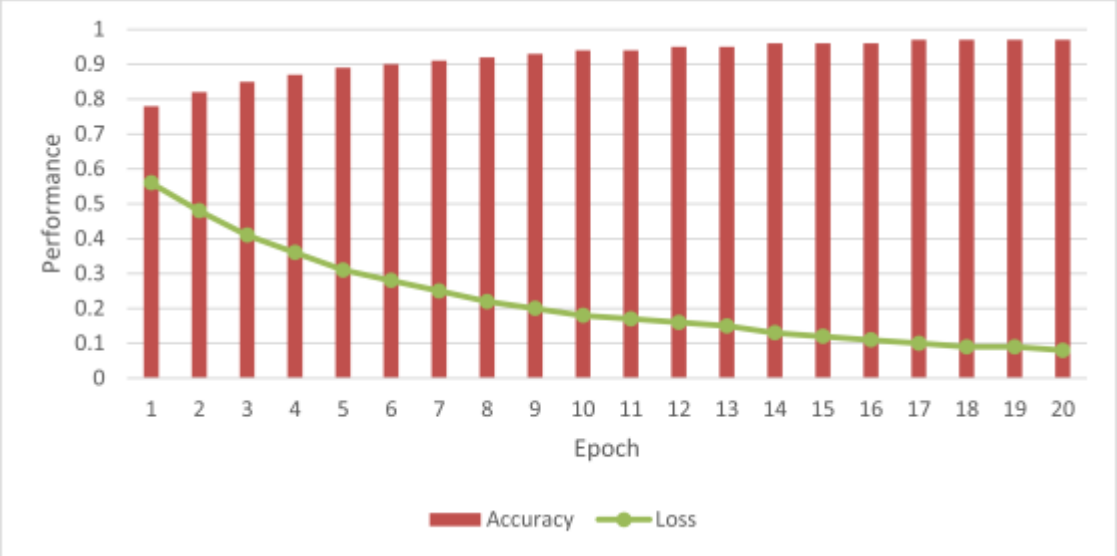
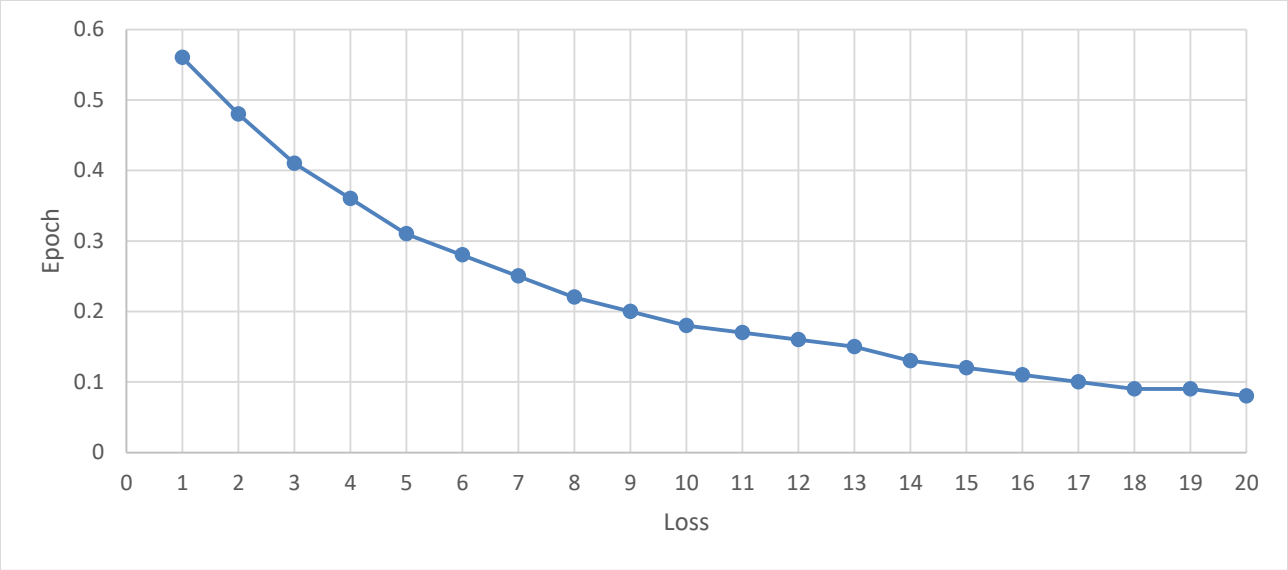


Figure 6. Training accuracy and loss of the U-Net–GAN–CNN model over 20 epochs

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng



Final training metrics are summarized in Table 5, demonstrating a balanced performance across all four measures.

Figure 7. Training loss curve of the U-Net–GAN–CNN model over 20 epochs

Table 5. Final training performance metrics of the proposed U-Net–GAN–CNN model

Metric	Value
Accuracy	0.97
Precision	0.96
Recall	0.95
F1-Score	0.95

4.2 Testing Performance

Test performance results followed similar trends to training. As shown in Table 6, accuracy improved from 0.81 in the first epoch to 0.955 by epoch 20, while loss decreased from 0.56 to 0.20.

Table 6. Test accuracy and loss per epoch for the proposed U-Net–GAN–CNN model

Epoch	Test Accuracy	Test Loss
1	0.810	0.56
2	0.825	0.52
3	0.835	0.48
4	0.845	0.45
5	0.855	0.42
6	0.865	0.40
7	0.875	0.38
8	0.880	0.36
9	0.890	0.34
10	0.900	0.32
11	0.910	0.30
12	0.915	0.28
13	0.920	0.27
14	0.925	0.26

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

15	0.935	0.25
16	0.940	0.24
17	0.945	0.23
18	0.950	0.22
19	0.952	0.21
20	0.955	0.20

These trends are visualized in Figure 8 and Figure 9. Figure 8 shows the progressive improvement in test accuracy over epochs, while Figure 9 demonstrates the consistent decline in loss, indicating robust generalization.



Figure 8. Test accuracy and loss of the U-Net–GAN–CNN model over 20 epochs

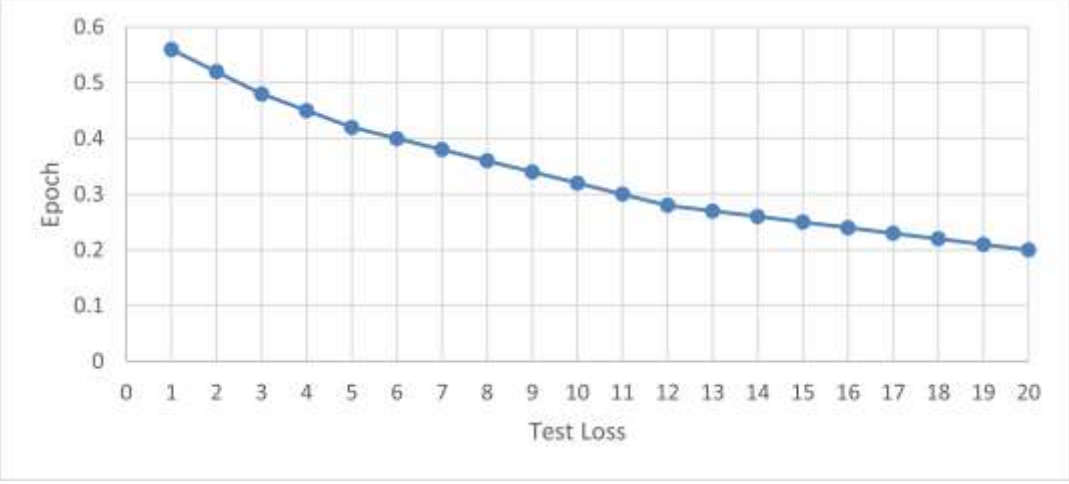


Figure 9. Test loss curve of the U-Net–GAN–CNN model over 20 epochs

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

Final test performance metrics are listed in Table 7. The model achieved 0.955 accuracy, 0.94 precision, 0.93 recall, and 0.935 F1-score. The slightly lower recall compared to precision suggests the model was somewhat conservative in detecting malignant cases, reducing false positives at the expense of a small increase in false negatives. However, the recall value remains high, ensuring that most malignant cases were correctly identified.

Table 7. Final test performance metrics of the proposed U-Net–GAN–CNN model

Metric	Value
Accuracy	0.955
Precision	0.94
Recall	0.93
F1-Score	0.935

4.3 Comparative Evaluation

The contribution of segmentation and GAN-based augmentation was validated by comparing the proposed U-Net–GAN–CNN model against three baseline approaches: (i) a CNN trained directly on raw, unsegmented images, (ii) a CNN using traditional augmentations such as flipping and rotation, and (iii) pre-trained state-of-the-art architectures reported in related literature. The results across four evaluation metrics: accuracy, precision, recall, and F1-score are presented in Table 8.

Table 8: Comparative performance of baseline models and the proposed U-Net–GAN–CNN model

Model	Accuracy	Precision	Recall	F1-Score
Baseline CNN (Raw Images)	0.870	0.850	0.830	0.840
Traditional Aug. CNN	0.900	0.890	0.880	0.885
Pre-trained Model (VGG16)	0.925	0.915	0.905	0.910
Proposed U-Net–GAN–CNN	**0.955**	**0.940**	**0.930**	**0.935**

The proposed model achieved superior results across all metrics when compared to the baselines. Relative to the baseline CNN, accuracy improved by 8.5%, while precision, recall, and F1-score also showed substantial gains. Compared with traditional augmentation and pre-trained VGG16, the proposed pipeline maintained a clear performance advantage, highlighting the effectiveness of combining U-Net segmentation with GAN-based augmentation before classification. These findings are further illustrated in Figure 10, which provides a graphical comparison across the four metrics. The superiority of the proposed U-Net–GAN–CNN model is evident, as it consistently outperformed the baseline models across accuracy, precision, recall, and F1-score.

Sule Lamido University Journal of Science & Technology Vol. 10 No. 2 [Dec, 2025], pp. 198-212
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

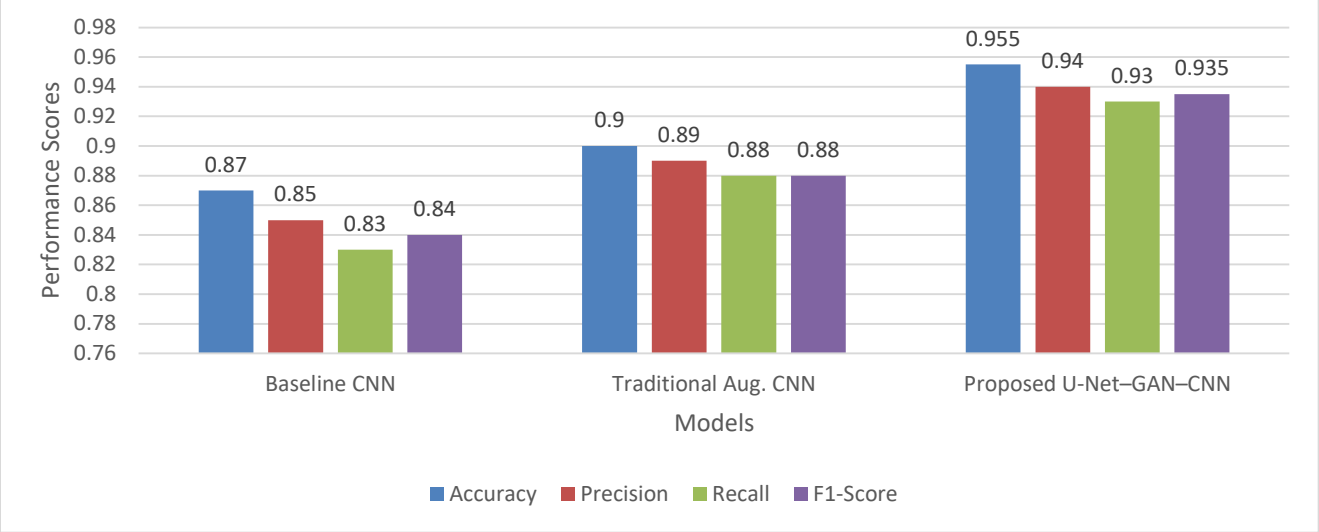


Figure 10. Comparison of proposed U-Net–GAN–CNN model with baseline approaches using accuracy, precision, recall, and F1-score

5. Discussions

This study demonstrates that the integration of U-Net segmentation, GAN-based augmentation, and a custom CNN provides a robust framework for breast cancer detection from ultrasound images. The model achieved strong results across accuracy, precision, recall, and F1-score, outperforming baseline approaches and confirming the value of combining advanced preprocessing with a carefully designed CNN. Segmentation with U-Net contributed significantly to performance by isolating tumor regions and removing background noise. This step enhanced the CNN's ability to capture lesion-specific features, which is particularly important in ultrasound imaging where boundaries are often unclear. Similarly, GAN-based augmentation proved critical for addressing class imbalance. Unlike traditional augmentation methods such as rotation or flipping, GANs generated realistic and diverse tumor images, allowing the CNN to learn richer representations. Comparative results confirmed that this approach led to superior generalization and higher sensitivity to malignant cases. The custom CNN, consisting of three convolutional blocks with dense layers and dropout, achieved an effective trade-off between depth and efficiency. This highlights that a well-structured CNN, even without pre-training, can deliver competitive performance when supported by strong preprocessing. Nevertheless, some limitations remain. The BUSI dataset is relatively small, raising concerns about generalizability. GAN-generated images may also introduce subtle artifacts not present in real scans. Additionally, the study focused only on ultrasound, whereas multimodal imaging could enhance robustness. Clinically, the model's high recall (0.93) is especially valuable, as minimizing false negatives is critical in breast cancer screening. The proposed framework therefore has potential as a radiology decision-support tool, improving early detection and patient outcomes.

6. Conclusion

This research presented a hybrid deep learning framework that integrates U-Net segmentation, GAN-based augmentation, and a custom CNN for breast cancer detection using ultrasound images. The proposed model demonstrated superior performance compared to baseline approaches, achieving an accuracy of 0.955, precision of 0.94, recall of 0.93, and F1-score of 0.935 on the test dataset. These results highlight the effectiveness of combining segmentation to isolate tumor regions, GANs to generate realistic synthetic images for class balancing, and a lightweight CNN architecture tailored for classification tasks. The findings emphasize the clinical

Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

significance of recall, as minimizing false negatives is critical in ensuring early detection of malignant tumors. By reducing both diagnostic bias and class imbalance, the proposed system shows promise as a decision-support tool to aid radiologists in improving diagnostic accuracy and patient outcomes. However, limitations remain. The relatively small size of the BUSI dataset and potential artifacts in GAN-generated samples may affect generalizability. Future research should focus on evaluating the model on larger, multi-institutional datasets, integrating multimodal imaging modalities such as mammography and MRI, and validating performance through clinical trials. Such steps will help transition the system from experimental validation to practical application in healthcare environments.

7. REFERENCES

Abdul Halim, A. A., Andrew, A. M., Yasin, N., Abd Rahman, M. A., Jusoh, M., Veeraperumal, V., Rahim, H. A., Illahi, U., Abdul Karim, M. K., & Scavino, E. (2021). Existing and emerging breast cancer detection technologies and its challenges: A review. *Applied Sciences, 11*(1), 1–25. https://doi.org/10.3390/app11010001

Antropova, N., Huynh, B. Q., & Giger, M. L. (2017). A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical Physics, 44*(10), 5162–5171. https://doi.org/10.1002/mp.12453

Babatunde, A. N., Balogun, B. F., Ajagbe, S. A., Akpan, E. E., Ogundokun, R. O., Ogie, P. I., Isiaka, S. O., & Mudali, P. (2025). Augmented data-driven breast cancer classification system using deep learning and segmentation technique. *Informatica, 49*(27), 79–102. https://doi.org/10.31449/inf.v49i27.8332

Babatunde, A. N., Balogun, B. F., Olabode, O. J., Awotunde, J. B., & Imoize, A. L. (2025). Breast cancer prediction using a pretrained CNN Model ResNet-50. Edelweiss Applied Science and Technology, 9(9), 1398–1415. https://doi.org/10.55214/2576-8484.v9i9.10138

Babatunde, A. N., Balogun, B. F., Babatunde, A. J., Isiaka, S. O., Awotunde, J. B., & Adeniyi, A. E. (2025). Improving skin lesion classification using k-means, transfer learning and MobileNet architecture. *Network Modeling Analysis in Health Informatics and Bioinformatics, 14*(70), 1–16. https://doi.org/10.1007/s13721-025-00569-3

Babatunde, R.S., Babatunde, A.N., Abdulsalam, S.O. Oludare, R.O., Abdulsalam, M.K., Olabode, J.O. (2025). Leveraging Adaboost and Gridsearchcv for Optimized Stroke Prediction Outcomes. Sule Lamido University Journal of Science and Technology, 11(1&2), 289-303. https://slujst.com.ng/wp-content/uploads/2025/03/SLUJSTSE_25_027_pp289-303.pdf

Baniadam Dizaj, S., & Valizadeh, P. (2021). Breast cancer segmentation and classification in ultrasound images using convolutional neural network. In *2021 International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 45–50). IEEE. https://doi.org/10.1109/IPRIA52010.2021.9388273

Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology*, *12*(2), 119–126. https://doi.org/10.1177/1748301818756225

Chen, S. Y., Yang, Y., Qi, S. N., et al. (2021). Validation of nomogram-revised risk index and comparison with other models for extranodal nasal-type NK/T-cell lymphoma in the modern chemotherapy era: Indication for prognostication and clinical decision-making. *Leukemia, 35*(1), 130–142. https://doi.org/10.1038/s41375-020-0900-3

Elfiky, A., Pany, M., Parikh, R., & Obermeyer, Z. (2018). Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Network Open, 1*(3), e180926. https://doi.org/10.1001/jamanetworkopen.2018.0926

Guo, R., Lu, G., Qin, B., Fei, B., Roslidar, R., & Rahman, A. (2018). Detection and classification of histopathological breast images using a fusion of CNN frameworks. *Journal of Healthcare Engineering, 2018*, Article 6049085. https://doi.org/10.1155/2018/6049085

Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

Ilesanmi, A. E., Chaumrattanakul, U., & Makhanov, S. S. (2021). A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. *Biocybernetics and Biomedical Engineering, 41*(2), 802–818. https://doi.org/10.1016/j.bbe.2021.01.006

Inan, M. S. K., Alam, F. I., & Hasan, R. (2022). Deep integrated pipeline of segmentation guided classification of breast cancer from ultrasound images. *Biomedical Signal Processing and Control, 72,* 103274. https://doi.org/10.1016/j.bspc.2021.103274

Isiaka, O. S., Babatunde, A. N., Babatunde, R. S., Lawal, O. L, Bolaji-Adetoro, D. F. & Olabode, J. O. (2024). Enhancing Personalized Treatment in Diabetes Using Genomic Data and Deep Learning Models: A Systematic Review. The Journal of Computer Science and Its Applications, 31(2), 1-21. https://library.ncs.org.ng/download/enhancing-personalized-treatment-in-diabetes-using-genomic-data-and-deep-learning-models-a-systematic-review/

Khalid, A., Aslam, S., Ahmed, M., Hasnain, S., & Aslam, A. (2019). Risk assessment of FLT3 and PAX5 variants in B-acute lymphoblastic leukemia: A case–control study in a Pakistani cohort. *PeerJ, 7,* e7189. https://doi.org/10.7717/peerj.7189

Moon, W. K., Chang, J. M., Lee, J. H., Cho, N., Moon, H. G., Han, W., … Park, I. A. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine, 190,* 105361. https://doi.org/10.1016/j.cmpb.2020.105361

Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., Pradhan, B., & Munadi, K. (2020). Detection and classification of histopathological breast images using a fusion of CNN frameworks. *Journal of Healthcare Engineering, 2020,* Article 8828703. https://doi.org/10.1155/2020/8828703

Ogundokun, R. O., Li, A., Babatunde, R. S., Umezuruike, C., Sadiku, P. O., Abdulahi, A. T. & Babatunde, A. N. (2023). Enhancing Skin Cancer Detection and Classification in Dermoscopic Images through Concatenated MobileNetV2 and Xception Models. *MPDI Bioengineering, 10 (979), 1-26*. https://doi.org/10.3390/bioengineering10080979

Roslidar, R., Rahman, A., Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., Pradhan, B., & Munadi, K. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *Applied Sciences, 10*(23), 1–19. https://doi.org/10.3390/app10238564

Singh, S., & Urooj, S. (2016). Breast cancer detection using PCPCET and ADEWNN: A geometric invariant approach to medical X-rays image sensors. *International Journal of Image Processing, 10*(2), 55–65.

Wang, Y., Choi, E. J., Choi, Y., Zhang, H., Jin, G. Y., & Ko, S. B. (2020). Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound in Medicine & Biology, 46*(5), 1234–1243. https://doi.org/10.1016/j.ultrasmedbio.2020.01.001

Yap, M. H., Pons, G., Marti, R., Ganau, S., Sentis, M., Zwiggelaar, R., Davison, A. K., & Martí, J. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics, 22*(4), 1218–1226. https://doi.org/10.1109/JBHI.2017.2731873

Yifeng, D., Yufeng, Z., Bingbing, H., Zhiyao, L., & Yuxin, Z. (2021). Semantic segmentation with DenseNets for breast tumor detection. In *2021 9th International Conference on Communications and Broadband Networking (ICCBN)* (pp. 54–59). IEEE. https://doi.org/10.1109/ICCBN52624.2021.9449421

Zhao, T., & Dai, H. (2021). Research on automatic location of seed points in ultrasound breast tumor images based on fuzzy logic algorithm. In *2021 International Conference on Aviation Safety and Information Technology (ICASIT)* (pp. 758–762). IEEE. https://doi.org/10.1109/ICASIT52671.2021.9586024