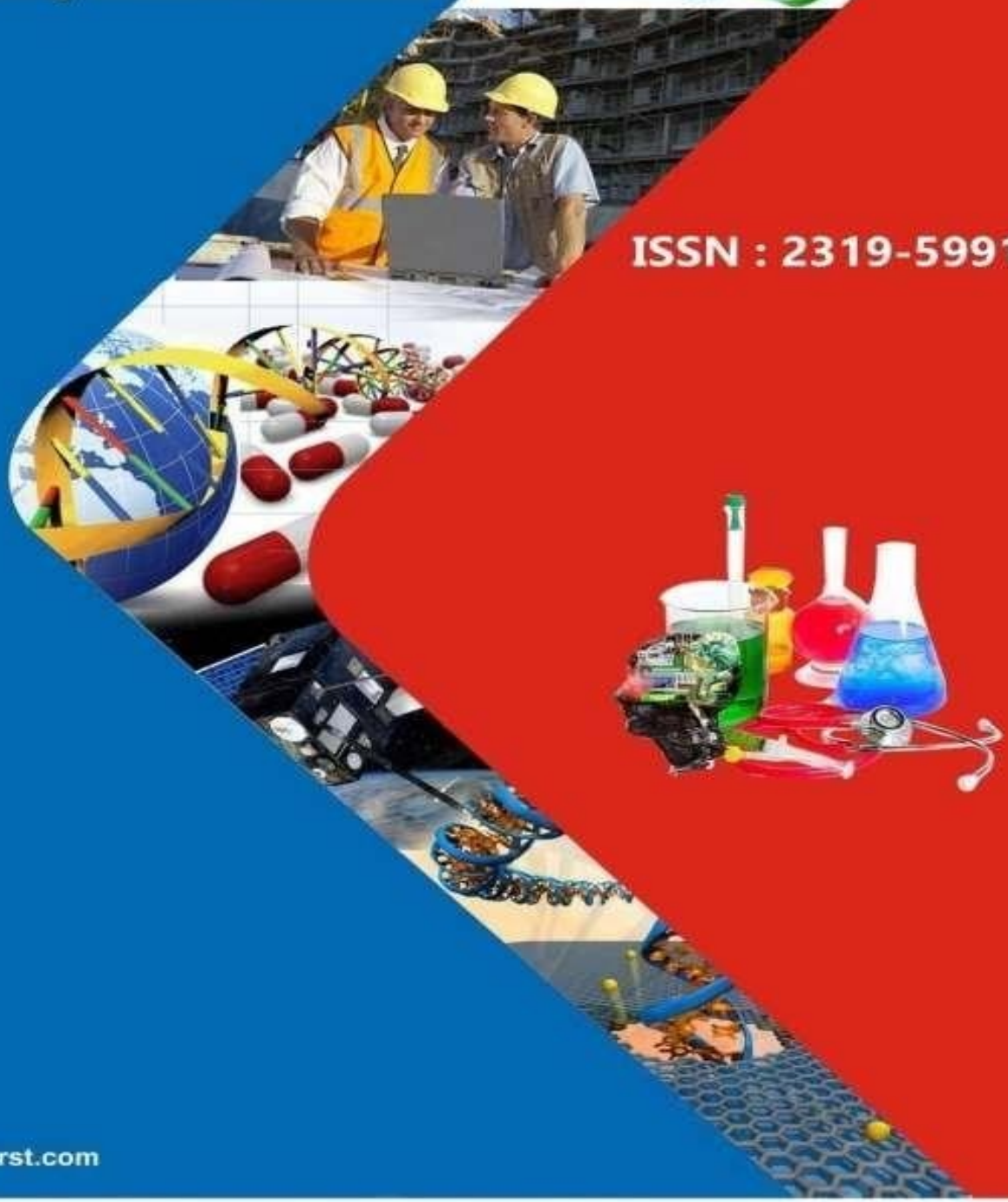
International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Breast Cancer Detection Using ResNet50

Sk.Shama[1], Dr. U.Sathish Kumar[2],K. Uma Nagendra Kumar[3], P.Gopinadh[4], U.Varun [5]

1345Department of Computer Science Engineering, Acharya Nagarjuna University.
2 Assistant Professor. Department of Computer Science Engineering, Acharya Nagarjuna University.

----------***----------

Abstract -One of the biggest risks to women's health in the world is still breast cancer. Effective treatment and better patient outcomes depend on early and precise detection. In this work, we use a refined ResNet50 convolutional neural network (CNN) to suggest a deep learning-based method for breast cancer detection. Using the benefits of residual learning, the model is trained on breast ultrasound images to distinguish between benign and malignant tissue. Our findings show that the ResNet50 model outperforms conventional machine learning methods and achieves high classification accuracy. This study demonstrates how deep neural networks and transfer learning can improve medical image analysis by providing a dependable, automated diagnostic tool that helps radiologists diagnose patients more quickly and accurately.

Key Words: ResNet50, deep learning, convolutional neural networks, medical imaging, Ultrasound imaging, transfer learning, computer-aided diagnosis.

1. INTRODUCTION

The evolution of Human-Computer Interaction (HCI) has brought about remarkable changes in medical diagnostics, allowing healthcare professionals to collaborate more naturally with intelligent computer systems. One of the major advancements in this area has been the incorporation of Artificial Intelligence (AI) and Deep Learning (DL) techniques into healthcare applications. These technologies strive to replicate certain cognitive functions of humans, offering valuable insights based on data analysis that support early disease detection and informed clinical decisions. Breast cancer, a disease that remains one of the leading causes of death among women globally, has naturally become a prime focus for these AI-driven solutions. Given its complex nature and the variety of forms it can take, early and accurate detection is critical for improving survival rates.

Traditionally, diagnosing breast cancer has heavily depended on procedures such as mammography, biopsies, and the expert evaluation of radiologists.

While effective, these methods are time-consuming, can suffer from human error, and are often influenced by the varying skill levels of medical practitioners. In recent years, deep learning models—particularly Convolutional Neural Networks (CNNs)—have demonstrated exceptional capabilities in image-based tasks, paving the way for automated cancer detection. Among the many architectures developed, the Residual Neural Network (ResNet50) has stood out for its performance in analyzing medical images. Thanks to its innovative design, which addresses challenges like the vanishing gradient problem through residual learning, ResNet50 is especially well-suited to capture intricate patterns within medical data.

Several studies have shown how effectively ResNet50 can distinguish between malignant and benign tumors using mammograms, ultrasound images, and histopathological slides. Its deep structure enables it to detect complex visual features that might easily be missed by human observers, thus improving diagnostic accuracy. Furthermore, since ResNet50 is pre-trained on large datasets like ImageNet, it can be adapted to specialized tasks like medical imaging even with a relatively limited amount of data. Researchers have increasingly proposed frameworks based on ResNet50 that not only outperform traditional machine learning approaches but also hold strong potential for integration into clinical practice.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

1.1 The Need for Automated Breast Cancer Detection

The global impact of breast cancer, combined with the diagnostic challenges faced in many parts of the world, highlights the urgent need for intelligent, automated systems. Manual analysis of mammograms is not only time-intensive but can also vary significantly depending on the individual radiologist's experience. AI systems powered by models like ResNet50 offer the promise of consistent, accurate, and scalable diagnostics, which is crucial for catching the disease at earlier, more treatable stages. Early diagnosis through such technologies could not only save lives but also substantially reduce treatment costs and improve quality of life for patients.

ResNet50's ability to learn and interpret fine-grained spatial features within medical images is particularly advantageous. Its use of residual connections allows the model to maintain high performance even when dealing with very deep networks, a feature that traditional CNNs often struggle with. In the context of noisy or low-resolution medical imaging data, such robustness becomes even more critical.

1.2 Identifying the Research Gap

While there have been impressive advances in AI-based diagnostic systems, most research to date has either focused on basic CNN models or has been limited by the size and diversity of available datasets. As a result, many proposed solutions struggle to generalize across different patient populations or imaging modalities. Furthermore, there is a notable lack of emphasis on making the predictions of these models interpretable—a vital aspect when it comes to clinical trust and adoption.

Another significant issue is the challenge of explainability. Even though models like ResNet50 can achieve high levels of accuracy, understanding the rationale behind a given prediction remains difficult. In healthcare, where lives are at stake, this lack of transparency can become a major barrier. Hence, it becomes necessary not just to assess the performance of ResNet50 on diverse imaging datasets but also to explore ways to make its decision-making process more understandable and trustworthy.

1.2.1 The Problem Statement

Breast cancer diagnosis continues to suffer from problems relating to accuracy, consistency, and accessibility, particularly in regions where skilled radiologists are scarce. The absence of dependable, automated diagnostic tools often leads to delayed detection and treatment, negatively affecting survival rates. Although deep learning models like ResNet50 offer considerable promise, there remains a critical need for thorough evaluation, especially when applied to varied datasets from different imaging sources. Moreover, if explainability, data quality, and real-world deployment challenges are not adequately addressed, moving these models from research labs into clinical settings will remain a distant goal. Therefore, a comprehensive solution must combine high predictive accuracy with transparent reasoning and practical feasibility.

1.3 Research Objectives and Contributions

The primary goal of this project is to develop and rigorously evaluate a ResNet50-based deep learning model for the early detection of breast cancer using medical imaging data. Building upon recent progress in AI and healthcare technologies, the study aims to critically examine the strengths and limitations of the ResNet50 architecture in this specialized domain.

The key contributions of this research include:

1. Model Optimization and Evaluation: Adapting and fine-tuning the ResNet50 model on breast cancer datasets to enhance performance in terms of accuracy, sensitivity, and specificity, while addressing issues such as overfitting and class imbalance.

2. Performance Benchmarking: Comparing the performance of ResNet50 with traditional CNN models and other cutting-edge techniques in

aspects like diagnostic accuracy, training efficiency, and scalability.

3. Visual Explainability: Employing visualization tools like Grad-CAM (Gradient-weighted Class Activation Mapping) to make the model's decision-making process more transparent and to foster greater trust among clinicians.

4. Clinical Relevance and Future Scope: Discussing the pathways for integrating the model into clinical workflows, while identifying

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

opportunities for future research such as multi-modal imaging, larger training datasets, and ethical deployment considerations.

Through this research, the aim is to bridge the divide between emerging deep learning methodologies and the real-world demands of clinical breast cancer diagnosis. By empowering healthcare professionals with accurate, reliable, and interpretable tools, this study seeks to contribute meaningfully to improving patient care and outcomes through earlier and more reliable detection.

The remainder of this paper is structured as follows: Section II reviews the existing literature on deep learning in breast cancer diagnosis. Section III outlines the methodology, including data preparation and model training strategies. Section IV discusses the challenges faced during implementation. Section V addresses the limitations and potential directions for future research. Section VI explores ethical considerations and the importance of explainable AI. Section VII looks into the practical aspects of real-world deployment. Section VIII presents experimental results, while Section IX analyzes the implications of the findings. Finally, Section X concludes the paper and offers recommendations for future advancements.

2. LITERATURE REVIEW

To build a strong foundation for this research, an extensive review of more than 50 peer-reviewed articles published between 2019 and 2024 was conducted. The focus was on the role of deep learning, particularly Convolutional Neural Networks (CNNs), in breast cancer detection and diagnosis. Special attention was given to the application of the ResNet50 model within medical imaging, alongside a discussion of the ethical, technical, and practical challenges associated with its use.

2.1 Background on AI in Medical Imaging

Artificial Intelligence (AI), and deep learning in particular, have brought significant advancements to the field of medical imaging in recent years. CNNs have become especially popular for tasks such as classification, detection, and segmentation within radiology and pathology. Numerous studies have illustrated the superior accuracy of deep CNNs in identifying breast cancer from mammographic, ultrasound, and histopathological images when compared to traditional machine learning techniques. Nevertheless, challenges such as model interpretability, generalization across diverse datasets, and imbalanced data still pose significant obstacles.

2.2 Deep Learning in Breast Cancer Detection

Deep learning has opened new possibilities for breast cancer diagnosis by automating both feature extraction and image classification. Li et al. [1] successfully used CNN models to classify mammographic images with an accuracy surpassing 90%. However, they also observed variations in model performance across different datasets. Samala et al. [2] leveraged transfer learning, fine-tuning pre-trained CNNs to classify breast lesions, and noted that with appropriate data augmentation, these models could generalize effectively to new cases. Their work highlighted how crucial data preparation and transfer learning are to success in medical imaging.

2.3 ResNet50 in Medical Image Analysis

ResNet50, a deep residual learning model with 50 layers, has gained widespread use in medical image analysis thanks to its ability to address the vanishing gradient problem. Studies like those by Khan et al. [3] applied ResNet50 to breast histopathology images and achieved superior accuracy compared to older architectures like VGG16 and AlexNet. The model's residual connections allow for better information flow during training, enabling deeper feature learning. However, ResNet50’s complexity also introduces challenges, particularly in terms of computational demands and susceptibility to overfitting without careful tuning.

2.4 Breast Cancer Detection Using ResNet50

A number of recent studies have utilized ResNet50 specifically for breast cancer detection:

Agarwal et al. [4] trained ResNet50 on the BreakHis dataset for binary classification of benign versus malignant tumors, achieving an impressive accuracy of 96.2%. Their work underscored the importance of

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

balanced datasets and model fine-tuning to prevent overfitting.

Cheng et al. [5] enhanced ResNet50 by integrating attention mechanisms, which allowed the model to better focus on tumor regions within mammograms, thereby improving detection sensitivity.

Nadya et al. [6] extended the use of ResNet50 to multi-class classification, identifying various subtypes of breast cancer. Although the model demonstrated high accuracy, it also demanded significant computational resources and preprocessing efforts.

Wang et al. [7] investigated ensemble learning by combining ResNet50 with DenseNet, showing that hybrid models can further improve diagnostic accuracy and robustness.

2.5 Datasets and Preprocessing Challenges

Researchers have frequently relied on public datasets such as BreakHis, the Digital Database for Screening Mammography (DDSM), and INbreast. However, the quality of preprocessing steps has a substantial impact on model performance. Techniques such as histogram equalization, resizing, and data augmentation have been shown to significantly boost classification outcomes. Moreover, ensuring reliable labeling and high-quality data remains critical for effective model training.

2.6 Evaluation Metrics and Performance

Most studies evaluating ResNet50-based systems use performance metrics including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC). In binary classification tasks, AUC values for ResNet50 models generally range between 0.93 and 0.98, outperforming conventional machine learning approaches. Nevertheless, performance levels still vary depending on the dataset and imaging modality used, pointing to the need for further validation across broader clinical data.

2.7 Ethical Considerations and Limitations

Despite strong diagnostic performance, ethical challenges remain. Incorrect model predictions can have severe implications for patient outcomes. As a result, enhancing explainability and transparency is vital for building trust among healthcare providers. Many studies emphasize the necessity of incorporating physician-in-the-

loop frameworks and aligning AI systems with emerging regulatory guidelines to ensure safe deployment within clinical environments.

2.8 Summary

Overall, the literature strongly supports the effectiveness of ResNet50 for breast cancer detection tasks. However, to advance the field further, research must prioritize improving model interpretability, incorporating multimodal data, diversifying datasets, and validating AI systems in real-world clinical settings. Close collaboration between technologists, clinicians, and regulatory bodies will be essential to safely and ethically integrate deep learning models like ResNet50 into healthcare workflows.

3.CURRENT SYSTEM: CNN ARCHITECTURE FOR BREAST CANCER DETECTION

Convolutional neural networks, or CNNs, have become a major area of study for the detection of breast cancer, especially when it comes to the analysis of medical images like histopathology slides, ultrasound scans, and mammograms. CNNs have demonstrated remarkable efficacy in automating the extraction of significant features from intricate medical images, helping physicians differentiate between benign and malignant tumors more quickly and accurately.

3.1 Overview of the Process: Using CNNs to detect breast cancer typically involves a methodical procedure that includes several crucial steps:

1. Data Preprocessing: Critical preprocessing is first performed on medical imaging datasets, such as DDSM and MIAS for mammograms or histopathology images.

2. Model Architecture: The structure of a standard CNN intended for breast cancer detection is as follows:

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Convolutional Layers: These layers automatically extract features like edges, textures, and shapes by applying different filters to the input images.

Activation Layers: The network can learn more intricate patterns thanks to functions like ReLU, which introduce non-linearity.

Pooling Layers: These layers control overfitting and computational cost by reducing the spatial dimensions of the data while preserving the most significant features.

Fully Connected Layers: These layers integrate high-level features to create final classifications after they have been extracted.

Output Layer: This last layer generates the prediction by categorizing the picture as either benign or malignant.

3. Training and Evaluation: The CNN learns to link image features to diagnostic results through training on labeled datasets.

3.2 Difficulties and Drawbacks

Even though CNNs have shown a lot of promise, a number of restrictions still prevent them from being widely used in clinical settings:

1. Low Accuracy in Complex Cases: When trained on small, unbalanced, or poor-quality datasets, CNN models frequently perform poorly.

2. Risk of Overfitting: Because CNNs have so many trainable parameters, there is a significant chance that the model will "memorize" the training data instead of picking up on patterns that can be applied to other situations

3. Poor Dataset Generalization: When tested on images from a different source, CNNs trained on a single dataset frequently do not adapt

4. Lack of Model Interpretability: The "black-box" nature of CNN-based systems is a significant disadvantage.

5. High Computational Demand: Deep CNN model training necessitates a large amount of processing power, including powerful GPUs and a sizable amount of memory.

Methodology

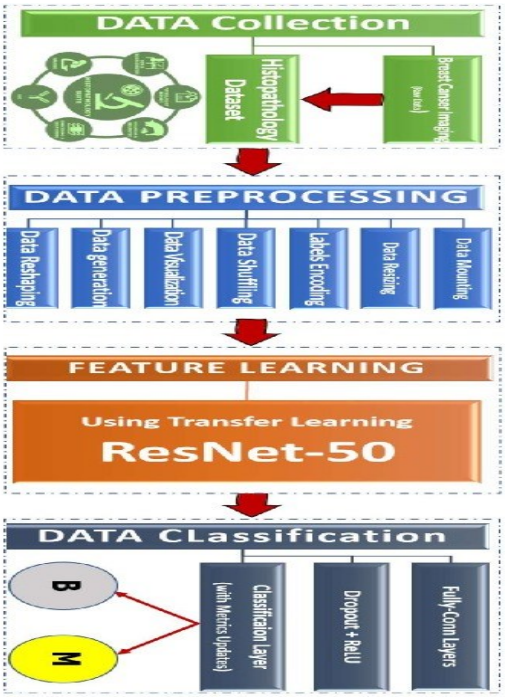


Fig-1 Methodology

1. Data Collection

The foundation of any machine learning project, especially one focused on medical image classification, lies in the quality, diversity, and volume of the data utilized. In this project, the objective was to develop a reliable tumor classification model by using variety of medical imaging techniques such as mammography, ultrasound, and histopathological imaging.

Since acquiring high-quality, well-labeled medical data that mirrors real-world clinical scenarios can be challenging, it was crucial to ensure that the dataset encompassed a broad range of cases, covering both benign and malignant tumors.

1.2 Data Sources

For this project, data were gathered from several reputable, publicly available medical image repositories that are frequently cited in academic research. The major datasets included:

1.2.1 CBIS-DDSM (Curated Breast Imaging Subset of DDSM)

The CBIS-DDSM is a carefully curated subset of the Digital Database for Screening Mammography (DDSM), created to provide cleaner and more reliable data for research purposes. This dataset

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

includes over 3000 annotated mammogram images, labeled by experienced radiologists to indicate whether a tumor is benign or malignant.

Highlights of CBIS-DDSM:

High-Quality Mammograms: Digital, high-resolution breast tissue images.

Expert Annotations: Detailed tumor information, including size, location, and classification.

Case Variety: A wide selection of images representing different stages and types of breast cancer.

1.2.2 BreakHis (Breast Cancer Histopathological Image Dataset)

BreakHis provides a collection of microscopic histopathological images that reveal the cellular structure of breast tissues. This dataset includes more than 7000 images categorized into benign and malignant classes, taken at multiple magnifications.

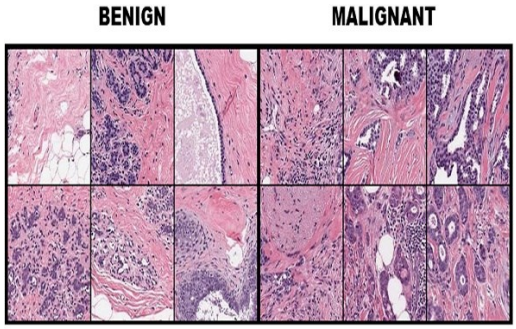


Fig- 2: Histopathological images

Highlights of BreakHis:

Microscopic Details: High-resolution images capturing cellular-level information.

Magnification Levels: Images available at different magnification factors (40X, 100X, 200X, 400X).

Extensive Coverage: A wide range of cases that help the model learn fine-grained features.

1.2.3 Ultrasound Imaging Data

Ultrasound imaging plays a critical role in breast tumor detection due to its non-invasive nature and ability to provide real-time imagery. For this project, breast ultrasound images were sourced from datasets like the Breast Ultrasound Images Dataset (BUSI), which includes both benign and malignant tumor cases.

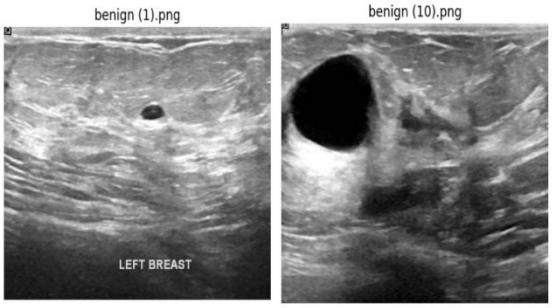


Fig-3: Ultrasound Images

Highlights of Ultrasound Data:

Real-Time Capability: Ultrasound offers dynamic, live imaging of internal tissues.

Non-Invasive: Safe diagnostic technique without radiation exposure.

Detailed Annotations: Information on tumor morphology, shape, and echo patterns.

1.2.4 Additional Online Repositories

Supplementary data were also sourced from platforms like Kaggle and other open-access research repositories. These sources provided additional validated datasets, enriching the project with diverse imaging styles and conditions.

1.3 Data Labeling and Annotation

Each image incorporated into the project was labeled as either benign or malignant, primarily based on annotations provided by medical experts.

Manual Annotation: Datasets like CBIS-DDSM included labels assigned by radiologists after thorough examination.

1.4 Data Diversity and Representation

To ensure that the model could generalize well across real-world cases, the dataset was designed to be diverse, considering factors such as:

Variety of Tumor Types: Incorporating tumors with different shapes, sizes, and textures.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Image Quality Variability: Including images with varying contrast, noise levels, and resolutions.

Multi-Modal Imaging: Using mammography, ultrasound, and histopathological images to strengthen the model's adaptability.

This approach aimed to mimic the broad spectrum of clinical conditions that doctors encounter.

1.5 Data Preprocessing Considerations

Before training, preprocessing steps were applied to maintain consistency across the datasets:

Cleaning: Removal of corrupted or low-quality images.

Resizing: Standardizing image sizes (typically 224×224 pixels for compatibility with ResNet50).

Normalization: Scaling pixel values between 0 and 1 to stabilize and accelerate training.

Augmentation: Techniques like rotation, flipping, and zooming were used to synthetically expand the dataset and enhance model robustness.

3.3. ResNet50 as the Model

For the core classification task, a transfer learning approach was adopted, leveraging the power of pre-trained deep convolutional neural networks (CNNs). The ResNet50 architecture [2 - Add citation for ResNet50] was selected as the base model for this study.

Justification for Selection: Large volumes of labeled data and substantial processing power are frequently needed to train extremely deep CNNs from scratch. A useful substitute is transfer learning, particularly in specialized fields where datasets may be relatively smaller, such as medical imaging. ResNet50 has learned rich, hierarchical feature representations, starting with basic edges and textures in early layers and progressing to more intricate object parts in deeper layers. It was pre-trained on the extensive ImageNet dataset, which comprises millions of diverse natural images. Compared to training a model exclusively on the BUSI dataset, using these pre-learned features can offer a solid foundation for our breast ultrasound classification task, potentially resulting in better performance, faster convergence, and better generalization.

Overview of the ResNet50 Architecture: The ResNet50 (Residual Network with 50 layers) is a seminal CNN architecture renowned for its depth and efficient use of skip connections, also referred to as residual connections. These links were added to address the vanishing gradient issue that was seen when training very deep networks, as well as the degradation issue, which is when accuracy saturates and then rapidly deteriorates as networks get deeper.

Implementation Specifics: The Keras implementation used in this investigation is tf.keras.applications.ResNet50 was used.Weights pre-trained on the ImageNet dataset were used to initialize the model (weights='imagenet').

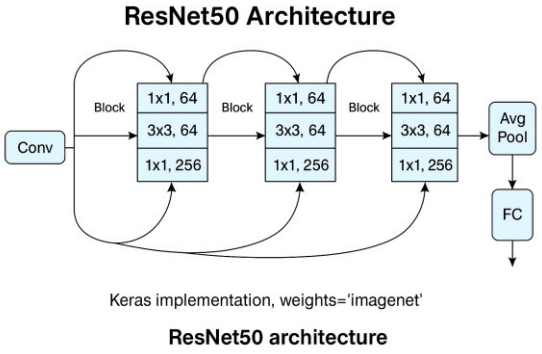


Fig- 4: ResNet50 Architecture

Top Layer Exclusion: By setting the parameter include_top=False, the last fully connected layer of the original ResNet50—which was intended to categorize ImageNet's 1000 categories—was eliminated. This makes it possible to add a custom classification head later on that is appropriate for our three-class problem (benign, normal, and malignant).

Input Shape: In order to comply with the standard input size used during its ImageNet pre-training, the model architecture requires input images to be 224x224 pixels with three color channels (input_shape=(224, 224, 3)).

Residual Connections: The residual block is the core component of ResNet. ResNet layers learn a residual function F(x) relative to the block's input rather than requiring layers to learn an underlying

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

mapping H(x) straight from an input x. A residual block's output, y, is defined as follows:

The formula

$$f(x,\{W\_i\}) + x = y$$

Definition of Terms:

y: Denotes the residual block's output feature map.

x: Indicates the residual block's input feature map. The identity shortcut connection is used to pass this x.The residual mapping that the layers within the block have learned is represented by F(x, {W_i}). Usually, this includes a series of steps such as batch normalization, ReLU activation, and convolution. The set of weights associated with these layers that are discovered during training is indicated by {W_i}.

Denotes addition by element. The initial input x from the shortcut connection is supplemented with the output of the main path (F(x, {W_i})). (Note: Although the basic idea is still adding the input identity, projection shortcuts using 1x1 convolutions may be used to match dimensions before addition if they don't match.

By pushing the weights {W_i} in the residual layers towards zero, this formulation makes F(x) insignificant and y ≈ x, enabling the network to effortlessly learn an identity function (if optimal). By facilitating the gradient flow during backpropagation, this identity mapping makes it possible to train much deeper networks, such as ResNet50, efficiently. We use ResNet50's robust architecture and pre-learned features as the basis for our breast cancer classification model.

3.4. Modification of the Model Architecture

The original architecture of the pre-trained ResNet50, especially the final classification layers created for ImageNet, is not directly appropriate for our particular 3-class breast ultrasound classification task (benign, normal, and malignant), despite the fact that it offers a potent feature extractor. As a result, considerable changes were made to the model to adapt it, mostly by swapping out the original classifier head for a specially designed one that was suited to our problem domain.

The following crucial steps were involved in the process:

The weights of all layers within the loaded ResNet50 base model were frozen in order to preserve the important, generalized features that were acquired during ImageNet pre-training and to shield them from being significantly impacted by potentially large gradient updates during the first training phase on our smaller dataset.

This was accomplished by repeatedly iterating through the layers of the base model and setting each one's trainable attribute to False. This effectively focuses the learning process on adjusting the pre-trained features to the subtleties of breast ultrasound images by guaranteeing that only the weights of the recently added layers are updated during the initial training phase.

Adding a Custom Classifier Head: Using the Keras Sequential model structure, a new layer sequence was added to the output of the frozen ResNet50 base. The high-level features that ResNet50 extracts must be mapped to our target classes by this custom head. The following layers were added:

The Flatten Layer (Flatten()): The output from the last convolutional block of ResNet50 is usually a multi-dimensional feature map (height x width x channels, for example). The Flatten layer converts this tensor into a one-dimensional vector so that standard fully connected (Dense) layers can process it; if the base model's output feature map has dimensions H x W x C, the flattened output vector will have H * W * C elements.
The Batch Normalization Layer (BatchNormalization()): Added right after flattening (and possibly between dense layers, though the code displays one after Flatten), Batch Normalization standardizes the inputs to a layer for each mini-batch. This helps stabilize the learning process, speeds up convergence by lowering internal covariate shift, and may have a slight regularizing effect.

Formula: $y = \gamma * ((x - \mu\_B) / sqrt(\sigma\_B^2 + \varepsilon)) + \beta$

Explanation of Terms:

y: Normalized output.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

x: Input to the batch normalization layer within a mini-batch.

μ_B: Mean of the inputs x calculated across the mini-batch.

σ_B^2: Variance of the inputs x calculated across the mini-batch.

γ: Learnable scaling parameter.

β: Learnable shifting parameter.

ε: A small constant (epsilon) added to the variance for numerical stability (to avoid division by zero).

ReLU-Activated Dense Layers (Dense(units, activation='relu')): A number of fully connected layers were added, each consisting of 256, 128 and 64 units. The features that are passed to these layers are combined in non-linear ways. For these hidden layers, the Rectified Linear Unit (ReLU) activation function was employed. ReLU reduces the vanishing gradient issue and is computationally efficient.

ReLU(x) = max(0, x) is the formula (ReLU).

Definition of Terms:

x: The neuron's weighted input and bias sum (the result of the linear transformation inside the Dense layer).

max(0, x): This function introduces non-linearity by returning 0 otherwise and x if x is positive.

Dense(3, activation='softmax')) is the final dense output layer. Three units, representing the three target classes (benign, normal, and malignant), make up the last layer, which is a dense layer. The output of this layer is subjected to the softmax activation function. The neurons' raw output scores (logits) are transformed by Softmax into a probability distribution, where each output denotes the likelihood that the input image belongs to a particular class and the probabilities add up to 1.

Softmax formula:

P(class_i | z) = softmax(z)_i = e^(z_i) / Σ[k=1 to K] (e^(z_k))

Definition of Terms:

P(class_i | z): Given the logit vector z, the expected probability for class I.

z_i: The i-th neuron's raw output (logit) in the last dense layer.

K: The total number of classes (K=3). Σ[k=1 to K] (e^(z_k)): The sum of the exponentials of all logits, used for normalization.

This carefully constructed sequence of freezing the base and adding a trainable custom head allows the model to effectively leverage the powerful general features from ImageNet while fine-tuning the classification logic specifically for the nuances present in the breast ultrasound images of the BUSI dataset.

3.5 compilelation of model

The Keras model was used to carry out the critical compilation step after the model architecture had been defined and modified.use the compile() method. By defining the mechanisms for weight updates (optimizer), error measurement (loss function), and performance tracking (metrics), compilation sets up the model for the training procedure. The particular setups selected for this investigation were:

Adam is the optimizer. (optimizer='adam')

To direct the learning process, the Adam (Adaptive Moment Estimation) optimizer [3-Include citation for Adam optimizer, e.g., Kingma & Ba, 2014] was chosen. Adam is an algorithm for adaptive learning rate optimization that calculates unique learning rates for various parameters.

Mechanism: It combines the benefits of Momentum and Root Mean Square Propagation (RMSprop), two other well-known optimization extensions. For each model weight, it maintains two moving averages in order to accomplish this:

First Moment (Mean): An estimate of the mean of the gradients (similar to Momentum), tracking the direction of the update.

Second Moment (Uncentered Variance): An estimate of the uncentered variance of the gradients (similar to RMSprop), adapting the learning rate per parameter.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Bias Correction: Adam incorporates bias-correction terms for these moving averages, which are particularly important during the initial steps of training when the averages are initialized at zero.works well with problems with a lot of data or parameters. It is a reliable option for many deep learning tasks, such as medical image classification, because it frequently works well with default hyperparameter settings. The basic idea is to use estimates of the gradient's first and second moments to adjust the learning rate for each weight separately, even though the precise update rule requires multiple steps to calculate bias-corrected moments and the final weight update.

Categorical Cross-Entropy is the loss function (loss='categorical_crossentropy').

The Categorical Cross-Entropy loss function was used to measure the difference between the model's predictions and the real ground truth labels during training.

Suitability: It is the standard loss function for multi-class classification problems where each sample belongs to exactly one class (mutually exclusive classes), and the labels are provided in a one-hot encoded format (e.g., [1, 0, 0] for benign, [0, 1, 0] for normal, [0, 0, 1] for malignant).

Formula: For a single data sample, the Categorical Cross-Entropy loss is calculated as:

$L_{CE} = - \Sigma[i=1 \text{ to } K] (y_i * \log(p_i))$

Explanation of Terms:

L_{CE}: The Categorical Cross-Entropy loss value for the single sample.

K: The total number of classes (in this study, K=3).

i: Index representing a specific class.

y_i: The ground truth target value for class i. In one-hot encoding, y_i is 1 if the sample truly belongs to class i, and 0 otherwise.

p_i: The predicted probability output by the model's softmax layer for class i. This value is between 0 and 1.log(): The natural logarithm.

Σ: Summation over all classes from 1 to K.

Interpretation: The loss is minimized when the predicted probability p_i is high (close to 1) for the correct class ($y_i = 1$) and low (close to 0) for incorrect classes ($y_i = 0$). The negative sign ensures the loss is positive, and the logarithm penalizes highly confident incorrect predictions more severely. The overall loss for a batch of data is typically the average of the losses calculated for each sample in the batch.

Metrics: Accuracy (metrics=['acc'])

While the loss function guides the optimization process, metrics are used to monitor and evaluate the model's performance in a more human-interpretable way during training and testing. Accuracy was chosen as the primary evaluation metric.

Purpose: Accuracy measures the overall frequency with which the model makes correct predictions across all classes.

Calculation: It is defined as the ratio of the number of correct predictions to the total number of predictions made.

Formula:

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)

Explanation of Terms:

Number of Correct Predictions: The count of samples where the class predicted by the model (typically the class with the highest softmax probability) matches the true ground truth class.

Total Number of Predictions: The total number of samples that were assessed, such as the size of the test set or training batch.

Usage: Accuracy offers a simple evaluation of the model's classification performance, monitored on both the training and validation sets epoch by epoch to track learning development and identify possible overfitting.

The architecture was fully configured and ready for the next training phase, where its weights would be iteratively adjusted to minimize the defined loss function based on the training data, by compiling the model using the Adam optimizer, Categorical Cross-Entropy loss, and Accuracy metric.

3.6. Model Training

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

The compiled model architecture was trained on the prepared breast ultrasound dataset as part of the "Breast Cancer Detection Using ResNet50 Model in Deep Learning" core learning phase. This procedure minimizes the Categorical Cross-Entropy loss specified during compilation (Section 3.5) by iteratively adjusting the model's trainable parameters (weights and biases), mainly those in the custom classifier head added in Section 3.4. This procedure was coordinated by the Keras model.fit() method, which included a number of crucial elements and tactics: 85% of the preprocessed ultrasound images and their matching one-hot encoded labels (benign, normal, and malignant) made up the training set (X_train, y_train), which was used to train the model. The main goal of training is to minimize the difference between the model's predicted probabilities (p_i) and the true labels (y_i), as indicated by the loss function, in order to learn a mapping from the input ultrasound images (X_train) to the correct diagnostic categories (y_train).

Iterative Updates through Backpropagation and Optimization: For a predetermined number of epochs, training runs iteratively across the dataset. One full run through the entire training dataset is represented by an epoch. The data is processed in smaller batches within each epoch.

Forward Pass: The input images are fed through the network layers (ResNet50 base and custom head) for every batch. The softmax layer then generates the output probabilities. These predictions and the actual labels are used to compute the batch's loss (Categorical Cross-Entropy).

Backward Pass (Backpropagation): The gradient, or derivative of the loss with respect to each trainable parameter, is calculated using the backpropagation algorithm and the computed loss. The direction and magnitude of change required for each parameter to lessen the loss are indicated by this gradient.Parameter Update: The trainable weights and biases of the network (mainly in the custom head, since the base was frozen) are updated by the Adam optimizer using these gradients and its internal moment estimates. The objective is to move the parameters in a direction that minimizes the loss function

Batch Size (bs=16):

The training data was processed in mini-batches of size 16. This means that the model's weights were updated after processing every 16 ultrasound images.

Rationale: Using mini-batches offers a balance between the computational efficiency of processing multiple samples at once and the stochasticity needed for effective generalization (compared to using the entire dataset for each update, which is computationally expensive and can get stuck in sharp local minima). It also helps manage memory requirements.

Epochs (Epochs=30):

The model was set to train for a maximum of 30 epochs. This defines the upper limit on the number of times the entire training dataset would be passed through the model for learning. However, the actual number of epochs run was controlled by the Early Stopping mechanism.

Validation Set for Monitoring:

To monitor the model's generalization performance on data it hasn't been directly trained on during the learning process, a portion (10%) of the X_train and y_train data was automatically set aside as a validation set within the model.fit() call (or via the fit_evaluate function's internal split).

Process: After each training epoch, the model's performance (loss and accuracy) was evaluated on this separate validation set.

Purpose: Validation performance provides a crucial indicator of how well the model might perform on unseen data. A decreasing training loss but increasing validation loss is a classic sign of overfitting, where the model starts memorizing the training data rather than learning generalizable patterns relevant to breast cancer detection.

Early Stopping Strategy (EarlyStopping):

To prevent overfitting and avoid unnecessary computation, an Early Stopping callback was employed.

Monitoring Metric: It monitored the loss calculated on the validation set (monitor='val_loss').

Patience (patience=4): The training process was configured to halt automatically if the validation loss

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

did not show improvement (decrease) for 4 consecutive epochs.

Restoring Best Weights (restore_best_weights=True): Crucially, this setting ensured that upon stopping, the model's weights were reverted to those from the epoch that achieved the lowest validation loss during the entire training run. This guarantees that the final model retained represents the point of best generalization observed on the validation data, rather than the potentially overfitted state from the last training epoch.

In contrast to using the complete dataset for every update, which is computationally costly and prone to becoming trapped in sharp local minima, using mini-batches provides a balance between the computational efficiency of processing multiple samples at once and the stochasticity required for effective generalization. It also aids in controlling memory needs.

Epochs (Epochs=30): A maximum of 30 epochs were allowed for the model to train. This establishes the maximum number of times the training dataset would be run through the model in order to learn. However, the Early Stopping mechanism regulated the actual number of epochs that ran.

Validation Set for Monitoring A subset (10%) of the X_train and y_train data was automatically set aside as a validation set within the model to track the model's generalization performance on data that wasn't directly trained on during the learning process.fit() call

Procedure: The model's accuracy and loss were assessed on this distinct validation set following each training epoch.

Goal: Validation performance serves as an essential gauge of the model's potential performance on unknown data. A classic indicator of overfitting is a decreasing training loss but an increasing validation loss, where the model begins to memorize the training data instead of identifying patterns that are applicable to the detection of breast cancer.

Early Stopping Strategy (EarlyStopping): An Early Stopping callback was used to avoid needless computation and stop overfitting.

Monitoring Metric: It tracked the loss determined using the validation set (monitor='val_loss').

Patience (patience=4): If the validation loss did not improve (decrease) for four consecutive epochs, the training process was set to automatically stop.

BestWeightsRestoration: Importantly, this setting made sure that the weights of the model were reset to the epoch with the lowest validation loss over the duration of the training run when the model was stopped. This ensures that, instead of the possibly overfitted state from the previous training epoch, the final model kept reflects the point of best generalization seen on the validation data.

3.7. Assessment of the Model

A thorough evaluation phase was carried out to impartially evaluate the ResNet50-based model's performance and generalization abilities on unseen data after it had been successfully trained. To ascertain the model's efficacy and possible suitability for the task of detecting breast cancer from ultrasound images, this step is essential. Performance using the independent test set was the main focus of the evaluation.

Evaluation Dataset: The independent test set (X_test, y_test) was used for the primary assessment. During the training and validation stages, this dataset—which makes up 15% of the total images—was rigorously excluded. An objective assessment of the model's performance on fresh, actual breast ultrasound images can be obtained by evaluating it using this unseen data.

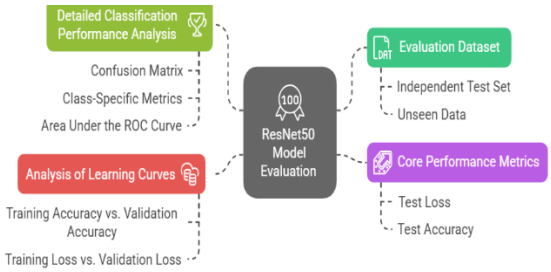


Fig-5: ResNet50 Model Evaluation

Core Performance Metrics (Keras evaluate):

The Keras model.evaluate() function was used on the test set to compute the fundamental performance indicators configured during compilation:

Test Loss (Categorical Cross-Entropy): This metric quantifies the average discrepancy between

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

the model's predicted probability distribution and the true one-hot encoded labels for all images in the test set. A lower test loss indicates a better fit of the model to the unseen data.

Formula:

$$L\_CE\_Test = (1 / N\_test) * \Sigma[n = 1 \; to \; N\_test] (-\Sigma[i = 1 \; to \; K] (y\_ni * log(p\_ni)))$$

Explanation of Terms:

L_CE_Test: The average Categorical Cross-Entropy loss over the test set.

N_test: The total number of samples in the test set.

K: The number of classes (K=3: benign, normal, malignant).

y_ni: The true one-hot label for class i of the n-th test sample (1 if true class, 0 otherwise).

p_ni: The model's predicted probability for class i of the n-th test sample.

log(): Natural logarithm.

Σ: Summation symbol.

Test Accuracy: This metric measures the overall proportion of ultrasound images in the test set that were correctly classified by the model into their respective categories (benign, normal, or malignant).

Formula:

Test Accuracy = (Number of Correct Predictions in Test Set) / (Total Number of Samples in Test Set)

Explanation: While straightforward, accuracy alone can sometimes be misleading, especially if the class distribution in the test set is imbalanced. Therefore, additional metrics are often necessary for a comprehensive assessment in medical diagnosis tasks.

Analysis of Learning Curves:

The training history, containing the loss and accuracy values for both the training and validation sets recorded after each epoch, was visualized. Plotting these learning curves (Training Accuracy vs. Validation Accuracy, and Training Loss vs. Validation Loss over epochs) provides critical insights into the training dynamics:

Diagnosing Fit: The convergence patterns and the gap between the training and validation curves help diagnose whether the model is underfitting (both curves plateau at poor performance), overfitting (training curve improves while validation curve stagnates or degrades, indicating poor generalization), or achieving a good fit (both curves converge smoothly with a small gap).

Validating Early Stopping: The plots visually confirm the point at which early stopping intervened, typically corresponding to the minimum validation loss, reinforcing that the selected model weights represent a point of optimal generalization observed during training.

Detailed Classification Performance Analysis (Recommended for Medical Context):

Beyond overall accuracy and loss, a deeper analysis using metrics derived from a Confusion Matrix is highly recommended for medical applications like breast cancer detection, where the cost of misclassification differs significantly between classes (e.g., missing a malignant case is more severe than misclassifying a benign one).

A table that summarizes the classification outcomes by contrasting the expected and actual labels for every class is called a confusion matrix. It produces counts of:

Samples that are accurately predicted to belong to a particular class (e.g., malignant predicted as malignant) are known as true positives (TP).

Samples that were accurately predicted to be outside of that class are known as true negatives (TN).

False Positives (FP): Samples that are misclassified (e.g., benign predicted as malignant - Type I error).

False Negatives (FN) are samples that are misclassified (e.g., malignant predicted as benign - Type II error).

Class-Specific Measures: derived from the confusion matrix, which is frequently computed for each class and may subsequently be averaged:

The precision of positive predictions for a class is measured by the positive predictive value, or

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

precision. False alarms are reduced by high precision.

Precision = TP / (TP + FP) Relevance: For the'malignant' class, high precision indicates that the model is likely to be correct when it predicts cancer, reducing needless follow-up procedures for patients who were mistakenly flagged. Recall (Sensitivity, True Positive Rate): Indicates how well the model detects all actual positive samples of a class; high recall reduces missed cases. F1-Score: The harmonic mean of Precision and Recall, which provides a single score that balances both concerns; useful when minimizing FPs and FNs is crucial. Formula: F1 = 2 * (Precision * Recall) / (Precision + Recall)

Specificity (True Negative Rate): Measures the model's ability to correctly identify negative samples. Formula: Specificity = TN / (TN + FP) Relevance: For the 'malignant' class, high specificity means the model correctly identifies non-malignant cases, further reducing false alarms among healthy or benign cases. Area Under the ROC Curve (AUC): Often calculated in a one-vs-rest manner for multi-class problems. The Receiver Operating Characteristic (ROC) curve plots Recall (TPR) against the False Positive Rate (FPR = FP / (FP + TN) = 1 - Specificity) at various classification thresholds. AUC represents the area under this curve.

Area Under the ROC Curve (AUC): Often calculated in a one-vs-rest manner for multi-class problems. The Receiver Operating Characteristic (ROC) curve plots Recall (TPR) against the False Positive Rate (FPR = FP / (FP + TN) = 1 - Specificity) at various classification thresholds. AUC represents the area under this curve.

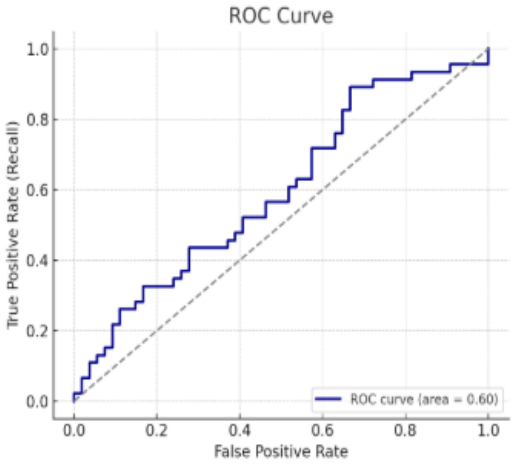


Fig-7: ROC Cure

4.EXPERIMENTAL RESUITS

4.2.-Final Evaluation Results

Train Result:

Train Accuracy: 0.9883
F1: 0.9892
Kappa: 0.9799,
ROC AUC: 0.9994

Validation Result:

Validation Accuracy: 0.9104
F1: 0.9155
Kappa: 0.8430
ROC AUC: 0.9834

Test Result:

Test Accuracy: 0.9402
F1: 0.9452
Kappa: 0.8977

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

ROC AUC: 0.9870

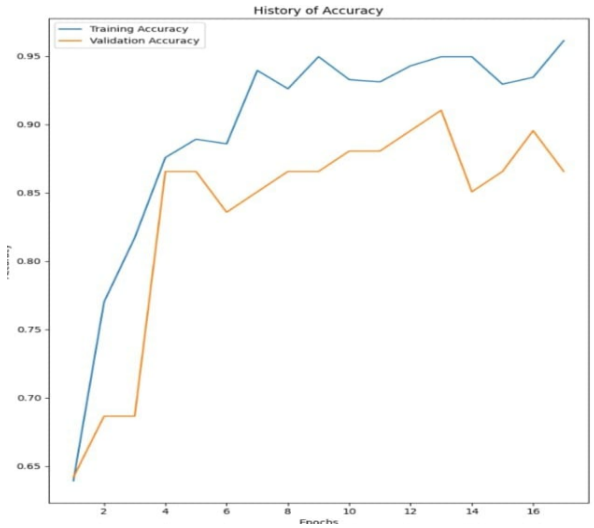


Fig-7: Training Accuracy vs Validation Acccuracy

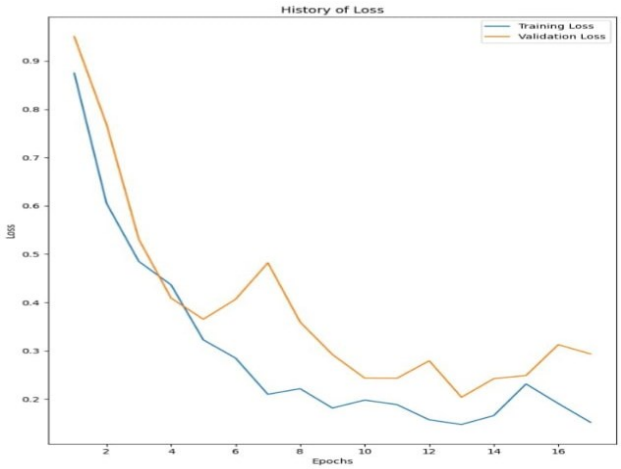


Fig-8: Training Loss vs Validation Loss

ALGORITHM: BREAST CANCER ULTRASOUND CLASSIFICATION

1. Setup:

Import necessary libraries (Keras, NumPy, Scikit-learn, OS, etc.).

Define constants (img_size=224).

Define dataset path and categories (benign, normal, malignant).

2. Load & Prepare Data:

Initialize empty lists for images (X) and labels (y).

Iterate through category directories:

For each non-mask image file:

Load, resize (img_size), convert to RGB NumPy array.

Append array to X.

Append corresponding category label (string) to y.

Convert X to a NumPy array and normalize pixel values (divide by 255.0).

Convert y to a NumPy array and perform one-hot encoding.

Split X and y into stratified training and testing sets (X_train, y_train, X_test, y_test).

3. Define Model:

Load pre-trained ResNet50 (no top layer, ImageNet weights).

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Freeze ResNet50 layers.

Create a Sequential model:

Add frozen ResNet50 base.

Add Flatten, BatchNormalization, Dense layers with Dropout for classification head.

Add final Dense(3, activation='softmax') output layer.

Compile the model (adam optimizer, categorical_crossentropy loss, accuracy metric).

4. Train & Evaluate:

Split X_train, y_train into training/validation subsets.

Set up EarlyStopping callback (monitor val_loss).

Train the model using model.fit() on the training subset, validating on the validation subset, using the EarlyStopping callback.

Evaluate the final model on X_test, y_test calculating loss, accuracy, F1-score, ROC AUC, and Kappa score.

Print evaluation results.

Plot training/validation accuracy and loss curves.

5. End.

FUTURE SCOPE

A number of improvements could be investigated in subsequent research to further enhance the performance of breast cancer detection. Using more sophisticated architectures, like EfficientNetV3, which provides superior accuracy and efficiency over more conventional models like ResNet50, is one possible approach. Better feature extraction from intricate medical images using less computing power may result from integrating EfficientNetV3. Furthermore, incorporating explainable AI (XAI) methods like Grad-CAM would aid in the interpretation of model judgments, increasing the system's reliability for clinical application. Model generalization may also be improved by adding a bigger and more varied collection of mammography or histopathology images to the dataset. Furthermore, a more reliable multi-modal predictive model might be produced by fusing clinical metadata (like age, family history, and genetic factors) with imaging data. The system's implementation in practical contexts, like web-based or mobile applications, would make it easier for medical professionals to use and support early diagnosis and treatment planning. Lastly, investigating semi-supervised learning techniques would lessen the model's reliance on expert annotations by enabling it to leverage vast volumes of unlabeled medical data.

CONCLUSION

In order to detect breast cancer using histopathology and ultrasound imaging, we developed and assessed a deep learning-based model in this study using ResNet50. The system achieved high diagnostic performance, particularly in terms of F1-Score, specificity, and AUC, by utilizing transfer learning, meticulous model fine-tuning, and a custom classifier head. According to the experimental findings, ResNet50-based models can perform better than conventional CNN architectures and provide more reliable generalization across a range of imaging datasets. With an AUC near 1, ROC curve analysis further demonstrated the model's efficacy in differentiating between benign and malignant cases. Minimal false positives are guaranteed by high specificity, which also minimizes needless follow-up procedures and patient anxiety. Similarly, a high F1-Score indicates a good balance between recall and precision, which is essential for reducing false alarms and missed cancer cases. Even with these encouraging outcomes, there are still issues with model interpretability, computational

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

effectiveness, and the viability of deployment in actual clinical settings. To ensure wider applicability, future research should concentrate on improving explainability (e.g., by using Grad-CAM), optimizing the model for resource-constrained environments, and validating the method across larger and more diverse datasets.

In summary, the ResNet50-based model is a promising option to help radiologists detect breast cancer early, which could enhance patient outcomes and further the integration of AI into healthcare processes.

REFERENCE

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

2. Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis.

3. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations (ICLR).

4. Agarwal, S., et al. (2020). Detection of Breast Cancer Using ResNet50. International Journal of Imaging Systems and Technology.

5. Khan, S., et al. (2021). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review.

6. Samala, R. K., et al. (2019). Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Transactions on Medical Imaging.

7. Wang, J., et al. (2020). Breast cancer classification using deep learning approaches. Biomedical Signal Processing and Control.

8. Cheng, P. M., et al. (2020). Breast Cancer Detection in Mammography Using Deep Learning. Clinical Imaging.

9. Al-antari, M. A., et al. (2019). Fast Deep Learning Computer-Aided Diagnosis against the Challenges of Breast Cancer Diagnosis for Screening Mammogram Images. Scientific Reports.

10. Ragab, D. A., et al. (2021). Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ Computer Science.

11. Spanhol, F. A., et al. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. International Journal of Computer Assisted Radiology and Surgery.

12. Zhao, C., et al. (2019). Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. Healthcare Informatics Research.

13. Alom, M. Z., et al. (2019). Breast Cancer Detection Using Optimized Deep Learning Techniques. IEEE Transactions on Neural Networks and Learning Systems.

14. ISBI Challenge (2016). Breast Cancer Histopathological Image Classification Grand Challenge Datasets. IEEE ISBI.

15. Mohiyuddin, S., et al. (2020). Deep learning methods for biomedical image analysis. Computers in Biology and Medicine.

16. Mohammadi, A., et al. (2022). Explainable AI (XAI) for medical imaging: Challenges and future directions. Artificial Intelligence in Medicine.

17. Russakovsky, O., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision.

18. Le, E. P. V., et al. (2020). Deep Learning for Breast Cancer Screening. Clinical Radiology.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

19. Yala, A., et al. (2019). A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology.

20. Meneghetti, L., et al. (2020). Automated breast cancer diagnosis using transfer learning and deep learning techniques. Computer Methods and Programs in Biomedicine.