© April 2026 | IJIRT | Volume 12 Issue 11 | ISSN: 2349-6002

Ultrasound-Based Breast Cancer Diagnosis with Deep Learning

Megha Saha[1], Soni N[2], Thanu Shree M N[3], Swathi P[4]
[1,2,3,4] *AMC Engineering College*

Abstract—**Breast cancer ranks among the leading causes of death from disease among women across the world and prompt detection is imperative to effective treatment. In this study, we proposed a deep learning approach aimed at automatically classifying breast ultrasound pictures into benign, malignant, and normal groups. The model was built using the DenseNet121 pretrained convolutional neural network, which was then intricately fine-tuned using additional layers to accomplish accurate classification outcomes. Modular image preprocessing was applied on the ultrasound images such as normalization, as well as data augmentations were implemented to benefit the training process. The model was trained using the Adam optimizer which was then evaluated on a test set using multiple accuracy and performance measurements. The proposed system achieved 94.94% test accuracy resulting in affirming the proposed approach as a possible reliable and efficient guide for radiologists in diagnosis of breast cancer cases.**

Index Terms—**Breast Cancer, Breast Ultrasound, Deep Learning, DenseNet121, Transfer Learning, Computer-Aided Diagnosis.**

Objective of the Study

Breast cancer is a significant threat to women's health globally, and early diagnosis can significantly save lives. Unfortunately, the manual interpretation of ultrasound images can be laborious, requires an expert, has the potential for human error, and may be misrepresented.

The main purpose of this study is to create a deep learning model that can automatically classify breast ultrasound images into one of three classes:

- Benign – non-cancer tumors
- Malignant – cancer tumors
- Normal – healthy tissue (no tumors)

By classifying ultrasound images, the model can help radiologists or doctors with a decrease in their workload, quick diagnosis outcomes, and better accuracy, all of which are greatly beneficial to health care providers with limited access to designated radiologist where they are from. Moreover, this research emphasizes building a model that is efficient and simple to use so that one day, it can be used in healthcare settings such as clinics and hospitals.

Method Used in Classification

We used a deep learning model called DenseNet121 to address this classification problem. DenseNet121 is a deep convolutional neural network (CNN) pretrained on ImageNet, which is a massive dataset of general images. Since DenseNet121 is pretrained on images, it has some relevant knowledge of images (edges, shapes, textures, etc.). Transfer learning, which is what we used in our study, saved us data and time, since we had reused DenseNet121's knowledge and simply added some layers on top of DenseNet121 to finetune it for breast ultrasound images.

I. INTRODUCTION

Breast cancer is one of the most common cancers in women worldwide and is an important public health problem. As prognosis and treatment options greatly improve with early detection, this is an important issue. Ultrasound imaging is an accepted method for evaluating breast findings. Ultrasound imaging is the preferred imaging in many clinical settings, given its low cost, safety, and high sensitivity for tumor detection, particularly in younger women or women with dense breast tissue. Despite all of the benefits, interpreting ultrasound imaging can be a difficult task. Interpretation is often based on the expertise of the radiologist and can lead to variation in diagnosis and human error. This challenge is even greater in areas of limited qualified personnel, such as radiologists.

Artificial intelligence has shown positive results in the automation of medical diagnostics due to the rapid expanding of the technology. Convolutional Neural

Networks (CNNs), a type of deep learning, have believed to be a powerful tool for medical image processing. These types of models can pick up features and patterns in images that are invisible to the human eye. However, there are usually both large computational costs and a need for much labeled data which are generally lacking in the medical research literature for building such models from scratch. Transfer learning has emerged as a well-known solution to overcoming these limitations. In some cases, like medical image classification, transfer learning fine-tunes pre-trained models from large scale image datasets. This strategy is useful because it is fast, efficient, and often yields higher accuracy rates, even when access to medical imaging data is limited. In this study, a pre-trained DenseNet121 architecture was chosen to classify breast ultrasound images into one of three categories: normal, malignant, and benign. The DenseNet121 architecture was selected, because it uses and reuse specific features more effectively which leads to a faster training speed, and better performance.

The aims of this work were to create a reliable and automated system to determine whether there are breast cancer indications in ultrasound images, which will allow radiologists to work faster and with higher accuracy. The goal of the project is to lessen the workload of medical staff while increasing rates of early detection of breast cancer by integrating deep learning with medical imaging and ultimately improving patient outcomes. The aims of this work were to create a reliable and automated system to describe whether there are indications of breast cancer in ultrasound images, which will allow radiologists to work faster and with a higher standard of accuracy. The ultimate goal of the project is to reduce the workload of medical staff while increasing the rates of early detection of breast cancer by integrating deep learning with medical imaging and ultimately improving patient outcomes.

II. LITERATURE SURVEY

Huang et al [1] introduced Dense Net, which changed effects for convolutional neural networks. They did not connect the layers one after the other. rather they connected every subcaste to all the layers. This small change helped the model produce features and it also helped the model to avoid the evaporating grade problem. The model could exercise what it learned without demanding a lot of parameters. Because of these benefits Dense Net came veritably popular in imaging especially for classifying bone cancer.

Kermany et al [4] showed that deep literacy models that were trained on individual images could diagnose medical conditions directly. They showed that these models are not just good at tasks; they can also handle complex visual details in medical imaging.

Ragab et al [8] took an approach. They combined convolutional neural networks with support vector machines. First, they used CNNs to find the features in the images, and they also used these features for classification with an SVM. The result was delicacy in opinion. Their study showed that combining literacy with traditional machine learning indeed makes bone cancer analysis more.

Shen et al [12] concentrated on using literacy to ameliorate bone cancer discovery in mammography wireworks. Their automated system helped radiologists find cancer easily and with lower misapprehensions. This is proof that AI- driven tools are veritably helpful in bone cancer screening.

Khan et al. [14] proposed a deep learning-based framework specifically for the detection and classification of breast cancer using ultrasound images.This exploration is related to the design showing that these models can handle ultrasound data well and find excrescences veritably directly.

Zhang and Patel [23] worked on the problem of having limited annotated ultrasound data. They used tone-supervised knowledge for bone cancer opinion. Their system showed that representation literacy ways make a big difference helping individual systems do a better job indeed when labeled data is limited.

Almazroi et al [21] reviewed the field. set up that deep literacy- grounded computer- backed opinion systems are veritably effective for bone cancer discovery. CNNs, transfer literacy and automated fabrics =constantly get results and offer good support for decision- timber. Their review makes a case, for developing smarter ultrasound- grounded bone cancer opinion systems.

2.1 Motivation

Medical diagnostics are high-stakes and time-dependent, and nowhere is this more true than in cancer care where early detection can have a life-changing impact on a patient's prognosis. Ultrasound imaging is an invaluable part of the breast cancer diagnostic process, however expert interpretation of ultrasound scans is required for it to be clinically useful. In many low-resource situations, the availability of skilled specialists is limited and patients could face delays resulting in late-stage diagnosis, or the possibility of undergoing unnecessary procedures due to mis-interpreted imaging results.

This study is powered by pressing concerns in the need for improved and timely access to breast cancer detection that is delivered in a consistent and standard way. The three motivations are:

1. High mortality due to diagnosis at disease stage: Many women diagnosed with breast cancer will die of breast cancer, and the vast majority of these cases will have been diagnosed at late-stage disease, when the efficacy of treatments is reduced. Early classification is essential in guiding appropriate therapy, particularly in distinguishing benign from malignant cancers.
2. Cognitive overload to interpreting radiologists: Radiologists routinely manage a considerable (increasing) volume of imaging data on a daily basis. Fatigue and human error can inhibit accurate recognition. An AI-based assistant to support in the interpretation of breast ultrasound will help to reduce cognitive overload.
3. Healthcare inequities: Many resource-limited regions, especially in low-and-middle income countries (LMICs) do not have the healthcare infrastructure or personnel to provide timely screening for breast cancer. AI models can help democratize access to the provision of diagnostic support.
4. Enhanced Decision Support: Even in well-resourced clinical settings, AI models provide a valuable second opinion and assist clinicians in confirming their interpretations.

2.2 Research Gap

There has been a significant increase in interest in AI in medical imaging, however, previously published models primarily use the classification task solely as a binary decision (cancer vs. no, in most studies), without distinguishing differentiating between the types of tumors (benign and malignant). Bypassing the distinction of benign versus malignant tumor risk oversimplifies the critical diagnostic distinctions that are needed for effective treatment planning. By extension, most studies do not fully validate against multiple and real-world datasets; therefore, they might be overfitted and unable to generalize. Only few studies have used more powerful architectures, like DenseNet121, which retain a greater number of features through the principle of dense connection. Furthermore, the following have not received adequate attention:

- Multi-class classification, which is important to clinical diagnosis
- Potential for advanced data augmentation to help with generalization
- Robust evaluation, with confusion matrices and classification reports
- Real-world deployment, in environments with limited resources
- Explainability and interpretability of predictions to earn clinician trust

This presents an opportunity to improve and develop a more comprehensive and reliable deep learning model for breast cancer image classification.

2.3 Research Contribution

We have provided a definite approach to multi-class classification of breast ultrasound images. We have been able to add value to the field in the following areas:

Multi-class differentiation- Our model is able to classify images into three clinically significant classes—benign, malignant, and normal. Multiclass classification conveys much richer diagnostic information than binary models used in existing breast cancer research.

DenseNet121 usage- We use a powerful, state-of-the-art pre-trained model. The DenseNet architecture is known for its effective design of feature reuse and gradient flow, which increases the learning ability of even small medical datasets.

Validation of methods- The dataset is separated into training, validation, and test sets, and augmentation strategies are employed to simulate real-world differences and to prevent overfitting on the training data when evaluating test performance. Performance indicators are reported on a whole scale.

Scale and usability- The workflow can directly translate to practice as a secure product for use in clinics, mobile health units, or telehealth setting.

End-to-end workflow- The way it's constructed here considers an end-to-end workflow from ingestion of data to visualization of performance. This means it can be reproduced in future research.

III. METHODOLOGY

The purpose of this study is to create and utilize a deep learning model to classify breast ultrasound images in three categories, benign, malignant, and normal. To complete the study we will detail a methodology that includes defining the dataset preparation and image preprocessing, building the model using transfer learning, training, and evaluation. Each of these stages will designed and executed to obtain the best model for accurate classification, reliability, and usability for medical work.

The first step on our methodology is to identify a suitable dataset. For the purpose of our study, we used the Breast Ultrasound Images (BUSI) dataset for training and evaluation of the deep learning model. This openly accessed, publicly available BUSI dataset contained a broad range of breast ultrasound images that are classified in categories benign, malignant, or normal. Those ultrasound images offered a range of useful cases that allowed the model to differentiate between different types of breast tissues. Pre-processing the datasets prior to running them through a deep learning model involved many processes. Within the pre-processing stage, the first is to standardize all the images in the datasets to have dimensions of 256×256 pixels. This step was vital to ensure that the images were all the same dimensions within either dataset. The next step was to normalize pixel values to a range of 0 to 1. Pixel normalization was important for making the model learn faster and more effectively. The next step was to augment the datasets. Data augmentation involved numerous techniques applied on the Image Data Generator function provided in Keras (image flip, rotation). Data augmentation artificially expands the amount of training data while also providing variety of the images, to help reduce overfitting while also allowing the model to generalize to unseen images. For the classification problem, the base of the system used DenseNet121 as the base system. DenseNet121 is considered a robust convolutional neural network, and has been shown to be a viable and efficient deep feature extractor. Instead of training the model from scratch, the model used transfer learning. Transfer learning is defined as taking the original version of DenseNet121 which had been trained on the ImageNet dataset, and then reused the model to build on it by adding additional layers to allow a model to classify breast ultrasound images. The additional layers included Flatten and dense layers with dropout in order to reduce overfitting and improve accuracy. The last layer used softmax activation to classify images into one of the three classes. The model was compiled with the Adam optimizer, which supports faster convergence during training. The loss selected was categorical crossentropy, which is appropriate for multi-class classification problems. Training was performed on the training set with validation performed on a separate validation set used to monitor the validation performance of the model. The training phase lasted 10 epochs with a batch size of 16. A ModelCheckpoint callback was used to save the best performing version of the model during training.

Once training was complete, the model was evaluated using an independent test set, containing unseen images. Performance of the model was evaluated using accuracy, confusion matrix, and classification report methods, to summarise how the model performed in successfully interpreting the three classes. Image by image predictions were made on some selected examples to visually validate the predictive capacity of the model.

3.1 Description of Dataset

The dataset utilized for this research study is Breast Ultrasound Images (BUSI) Dataset, which is publicly and widely available in the medical imaging community. The data were accumulated and subsequently classified by medics to ensure the reliability and quality of the data. Overall, the dataset comprised 780 ultrasound pictures sorted into three classes:

- Benign: Non-cancerous tumors or lesions (437 images)
- Malignant: Cancerous tumors (210 images)
- Normal: Healthy breast tissue with no abnormalities (133 images)

All images in the dataset are in PNG format and have ground truth labels that correspond with histological diagnoses. The BUSI Dataset also contains the mask images that actually segment the tumor area, although we only used the raw images for classification in this study.

Features of the dataset:

- The images are grayscale but converted to RGB to conform with the DenseNet121 model input.
- The images vary in resolution therefore they were resized to a uniform 256x256 pixel format during pre-processing.

There is class imbalance, particularly between malignant and normal, mitigated by the use of data augmentation, and balanced sampling during training. The BUSI dataset was used for its easy access, practicality in clinical settings, and well-annotated data quality, therefore this dataset was adequate for testing deep learning models for breast cancer identification.

IV. PROPOSED METHOD

The proposed method is a comprehensive deep learning pipeline using a pre-calibrated model in the form of a DenseNet121 architecture to classify breast ultrasound images as benign, malignant, or normal. It is based on the principle of transfer learning, where a pre-trained model is adjusted to the domain of medical imaging. By freezing the convolutional portion of the DenseNet121 pre-trained model, the enhanced deep neural network was able to use its extensive number of learned, relevant features, while only retraining the newly created dense layers to the task of breast image classification. The data processing pipeline also included numerous steps for preprocessing and augmentation, as well as splitting for the purpose of a greater sense of robustness and generalization. The training strategy focused on validation monitoring and model checkpointing, in order to save the best model identified during the training process. Once trained, the model was thoroughly evaluated using multiple metrics, and after which it was tested in so-called real-time scenarios to assess its readiness for clinical use.

Aspects of the Recommended Process:

1. Transfer Learning using DenseNet121:
- Use of the DenseNet121 and applying transfer learning from this model (pretrained on ImageNet) to allow for deep feature extraction in an efficient manner.
- Convolutional layers (feature extraction layers of the model) were frozen during training to avoid undoing what the original training on ImageNet had previously taught the model, which lead to better overall results and saved training time.

2. Model Modification:
- Added several custom dense layers to properly tailor the model to the issue of Breast Cancer Classification:
- Flatten Layer that converts the 2D features to one dimensional.
- Two separate dense layers. The first of which has 1024 units and uses the ReLU activation function. The second has 512 units and similarly uses ReLU as the activation function.
- Two dropout layers to prevent overfitting of the model, one with a dropout value of 50% and one with a dropout rate of 30%.
- A Softmax output layer for multi-class classification.

3. Data Preparation and Data Augmentation:
- The training image data sent to the model for training were normalized and transformed to the same dimension (256 x 256 pixels).
- Data Augmentation was performed with the training image data to produce new training data while improving generalization. Augmentation benefited the model by introducing, rotating, zooming, flipping and changing brightness.

4. Training procedures:
- Compiled using Adam and used categorical cross-entropy to represent loss.
- Trained by batching size of 16 for 10 epochs.
- The best model was picked based on validation accuracy on the validation set and stored in the Model Checkpoint for later analysis.

5. Evaluation as a metric:
- Determined model performance using accuracy, precision, recall and F1- score.
- As a part of building the model I produced a confusion matrix and classification report to investigate results.
- The final test accuracy achieved was 94.94%. This indicates a high level of accuracy and reliability.

FLOWCHART

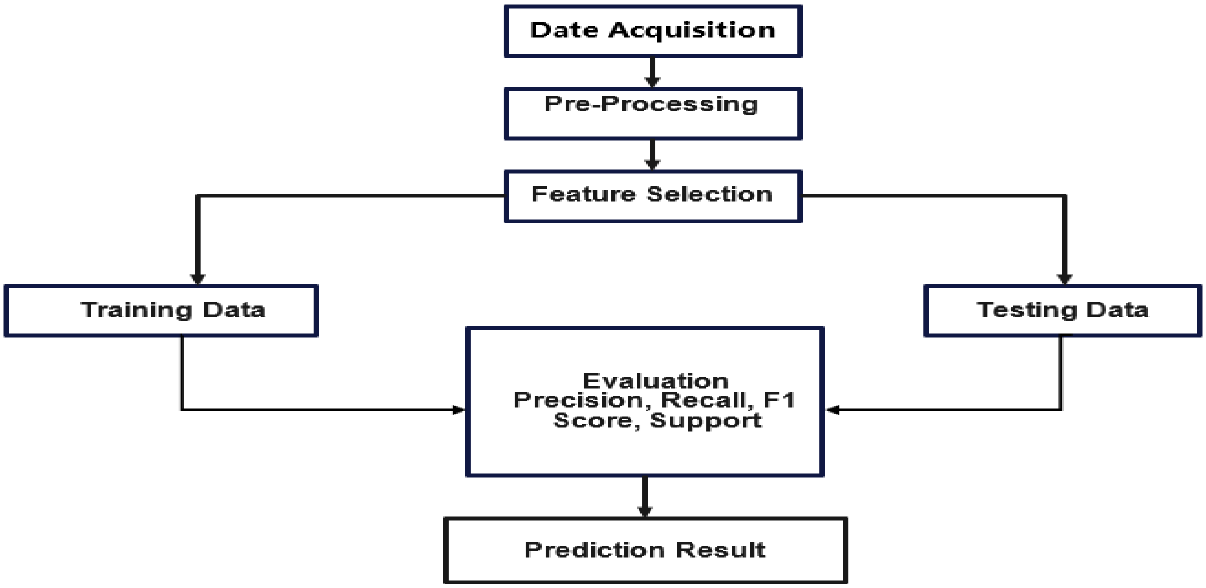


Fig 1: Proposed Methodology

OBJECTIVE

This project primarily aims to develop a deep learning-based classification system for breast ultrasound images that can identify benign, malignant, and normal tissues. The project will focus on the following objectives:

1. To automate breast cancer diagnosis in ultrasound images

- Manually diagnosing breast cancer in ultrasound images is an time-consuming effort in addition to the human error risk. The outcome of this project will have an AIsystem that aims to:

- Automate the diagnosis.
- Assist radiologists and their workloads.
- Provide fast diagnosis with consistent, and repeatable results.

2. Use deep learning and transfer learning to help improve the accuracy of the classification process

The project makes use of a pre-trained model, DenseNet121 instead of training a neural network from scratch. DenseNet121 has been validated widely in many machine learning classification tasks as it is:

- Great for feature extraction,
- Efficient with vanishing gradient issues, and
- A high performing model in visual classification tasks.

With transfer learning, it is expected that classification accuracy will improve, even when only a relatively small dataset is available.

3. To assess the effectiveness of DenseNet121 with respect to medical imaging

- Another objective is to see how well DensNet121, which was trained on natural images, can generalize to medical imaging. Questions that the assessment will help answer include:

- Does DenseNet121's feature space and classification generalize well to medical imaging?
- What sort of features need to be modified (layer freezing, dropout, etc)?

4. Build a Strong Training and Validation Framework

In the code, there are steps to split the data into train, validation, and test datasets. The goal is to:

- Train the model to perfection and see what accuracy it can get with 80% of the data,
- Validate the model's performance on 10% of data during training,
- Use the remaining 10% to assess the model's performance during testing with no bias.

Understanding how to split data into training, validation, and test data contributes to a balanced model that can generalize.

5. Provide Informative Evaluation and Visualization Metrics

Using the project's confusion matrices, classification report, and accuracy/loss plots it achieves the following to produce:

- Visual and quantitative representations of the models ability to produce correct predictions,

- Identify tendencies toward large or small bias for different classes,
- Confirm that predictions did progress through benign, malignant, and normal images with fairness.

6. Allow Real World Testing Through Prediction on Custom Images
The model is also given a chance to get tested through a few of the ultrasound images carefully selected and uploaded manually, and it's outcome was observed as to:

- To approach how this model would be used in the world,
- To confirm that the model works just fine with custom image inputs,
- To see how accurate the prediction is when image inputs are not tested in a batch process.

7. Prepare the Groundwork for Clinical Deployability
Even though this is only a prototype, the long term vision includes:

- Saving the best model with ModelCheckpoint.
- Producing clean architecture visuals.
- Creating a pathway to create a deployable model in a real-world diagnostic tool using tools such as Streamlit or a web API.

8. Make Breast Cancer Screening Accessible
Wherever experienced radiologists are in short supply, deploying an automated AI solution will:

- Increase clinical utility as a second-opinion tool,
- Provide assistance to less experienced clinicians
- Aid early breast cancer detection to increase patient survival rates.

V. RESULT AND DISCUSSION

Performance Metrics
Training Accuracy ≈ 93.22%
Validation Accuracy ≈ 91.87%
Test Accuracy ≈ 94.94%
The performance results above tell us that we have a good-performing model, and it has strong generalization ability.
Confusion Matrix & Classification Report:
A confusion matrix is a summary of the predictive performance of a classification model. It indicates how many of each class were predicted correctly and incorrectly. In this case (multi-class classification with 3 classes: benign, malignant, normal), it is a 3x3 matrix.

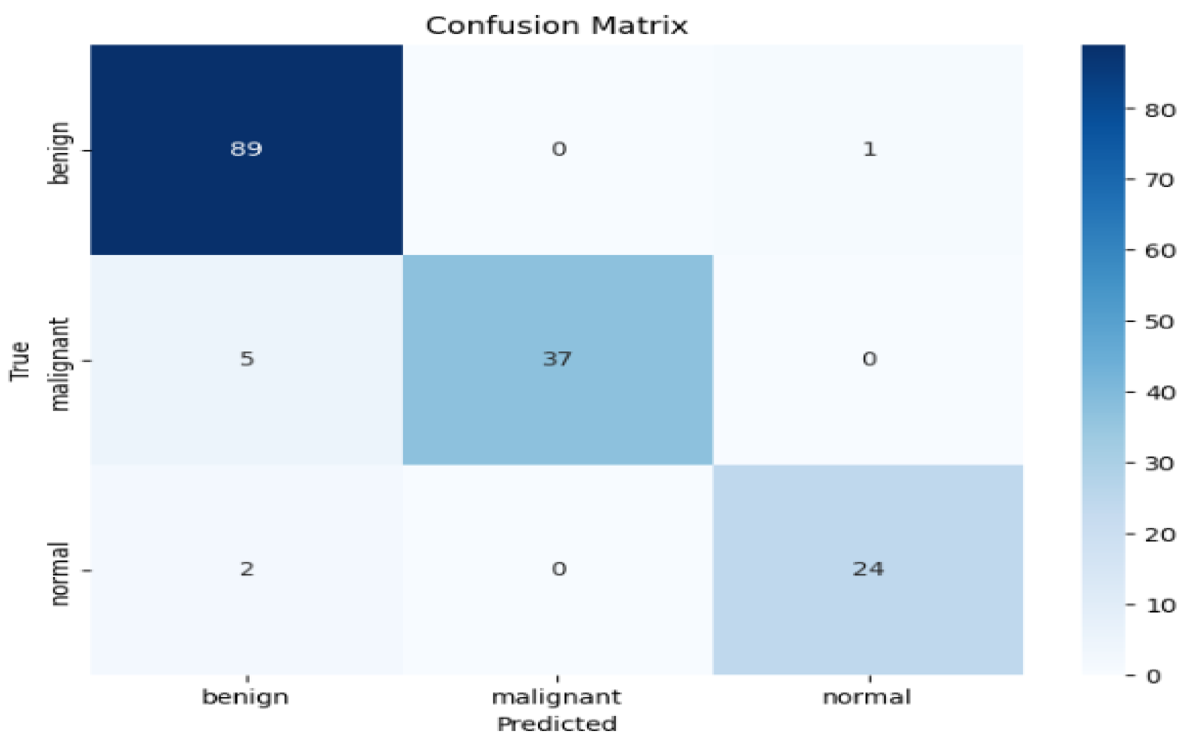


Fig 2: Confusion matrix of DenseNet121

Table 1: Classification report of DenseNet121

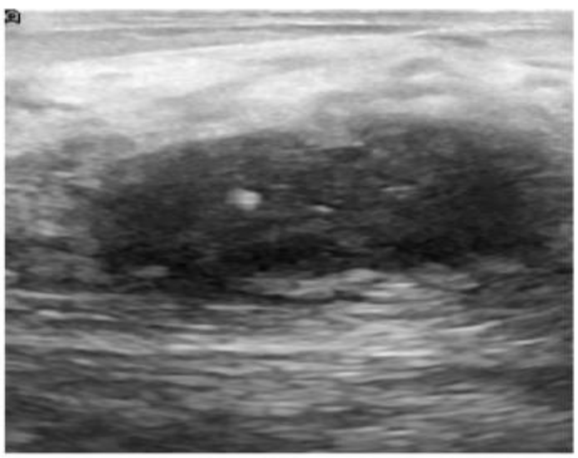
	Precision	Recall	F1-score	Support

Benign	0.93	0.99	0.96	90
Malignant	1.00	0.88	0.94	42
Normal	0.96	0.92	0.94	26
Accuracy			0.95	158
Macro average	0.96	0.93	0.94	158
Weighted average	0.95	0.95	0.95	158

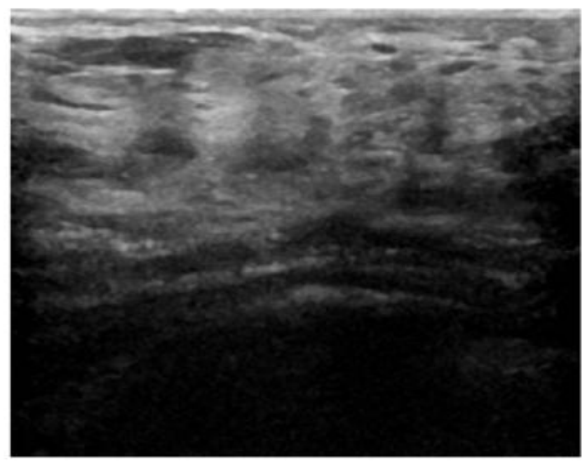
Classify Images

The model makes predictions on each image in the data set, selects the class with the highest predicted probability as the predicted class, and maps this prediction to the class label using a dictionary. The real class of the image is inferred from the directory name in the path, and is compared to the predicted label, to determine whether the model classified the image correctly. Whether or not the model classified the image correctly (correct/incorrect) is printed, while the script displays each image. Finally, the script reports on the total number of test images that were classified correctly out of the total number of test images processed, but also keeps track of the number of correct predictions made.

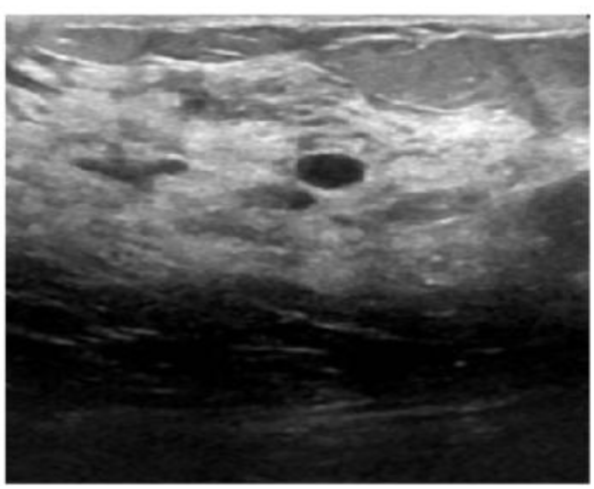
```
Real Class: malignant
Predicted Class: malignant
Correct
```



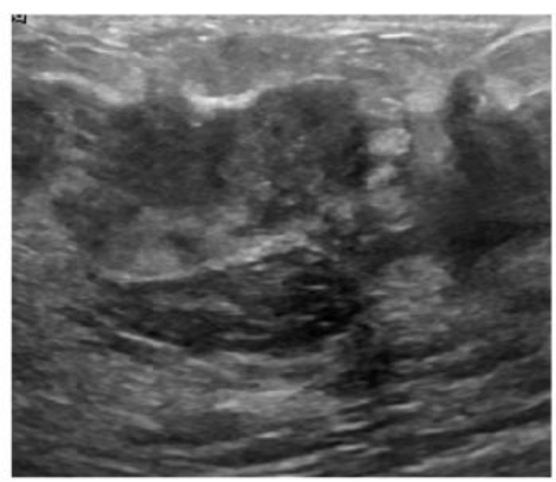
```
Real Class: normal
Predicted Class: normal
Correct
```



```
Real Class: benign
Predicted Class: benign
Correct
```



```
Real Class: malignant
Predicted Class: malignant
Correct
```



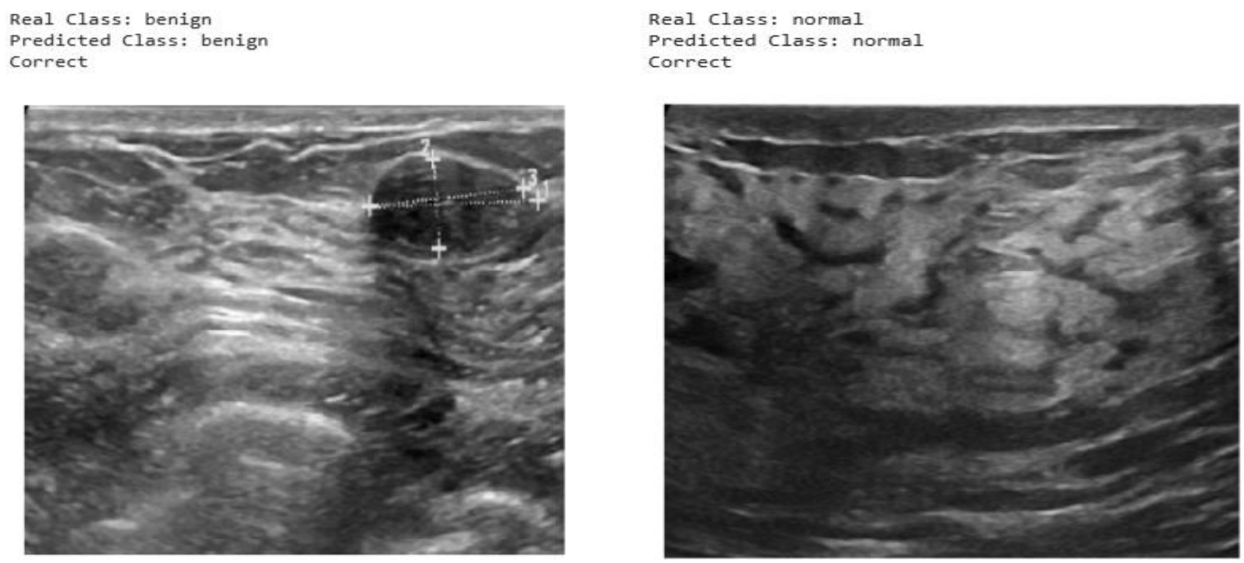


Fig 3: Classification of images

Date Visualization

The practice of visualizing data to enable communication, interpretation, and understanding is essentially the act of representing data visually. This example is shown as a bar chart, in which each bar represents the different categories of the data or "classes" in the "Label" column—as in the case of "Benign" or "Malignant" classes for the types of data columns in a medical dataset. The height of each bar represents how many of the instances belong to an individual category, in which more height means more instances from that specific category. Thus, anyone looking at this bar chart can immediately see the distribution of classes or categories of this particular data, and what the categorical instances were organized as. The actual numbers are placed above each column so anyone viewing the chart has a numerical reference, making it even more helpful, both as visual representation and a numeric perspective. The x-axis class notes are rotated so they are more readable, and a title is added which shows how many files are in the dataset overall. The visual representation provides the viewer with an indication of whether the data is unbalanced or skewed toward a particular class, since this is notably helpful when it comes to model training, for instance in the context of machine learning.

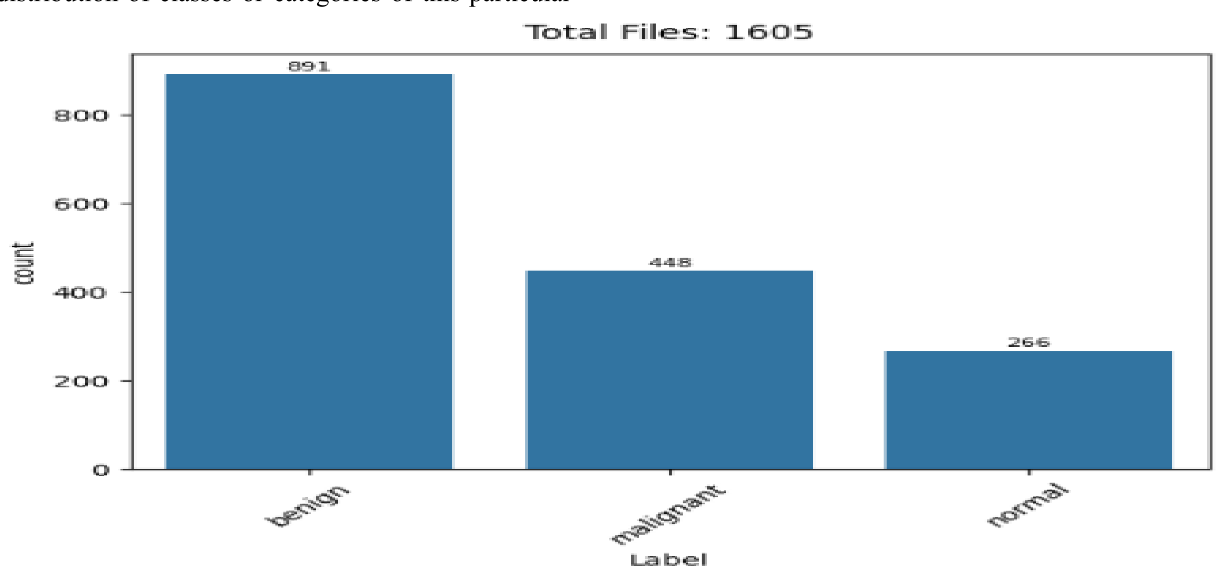


Fig 4: Graph of total file

Image Grid Visualization

The grid-based visualization approach was useful for developing an understanding of the dataset and visual verification of images, with respect to quality and structure. There isn't reporting of each training image, we create a grid with a selection of sample images from the training set and the accompanying labels. The images are in two rows with four columns for a total of eight images. Each image is loaded from the dataset using the Python Imaging Library (PIL), converted into NumPy array format for compatibility with plotting tool compatibility, and images are displayed using Matplotlib's subplot capabilities. Labels were included below each image to identify its associated class, such as "Benign" or "Malignant" in the breast cancer case study, for a more accurate view of the distributions of classes and it is an additional step to try to look for data that is labeled and formatted correctly prior to starting a model training process. The Axis markers were removed for the display of the images for a clean presentation without distractions, so that manufacturers could view the images in a visual sense and interpret them quickly.

This method of exploratory data analysis (EDA) was useful for a quick assessment of quality and structure of the dataset. It allowed us to provide early warning for potential problems such as incorrect labels, wrong file paths, or different image resolutions. When combined with other visual inspection techniques in the early stages of a computer vision project, we can assist in determining the quality of the input data, which will lead, in turn, to improved model performance, and a reliable and accurate output.

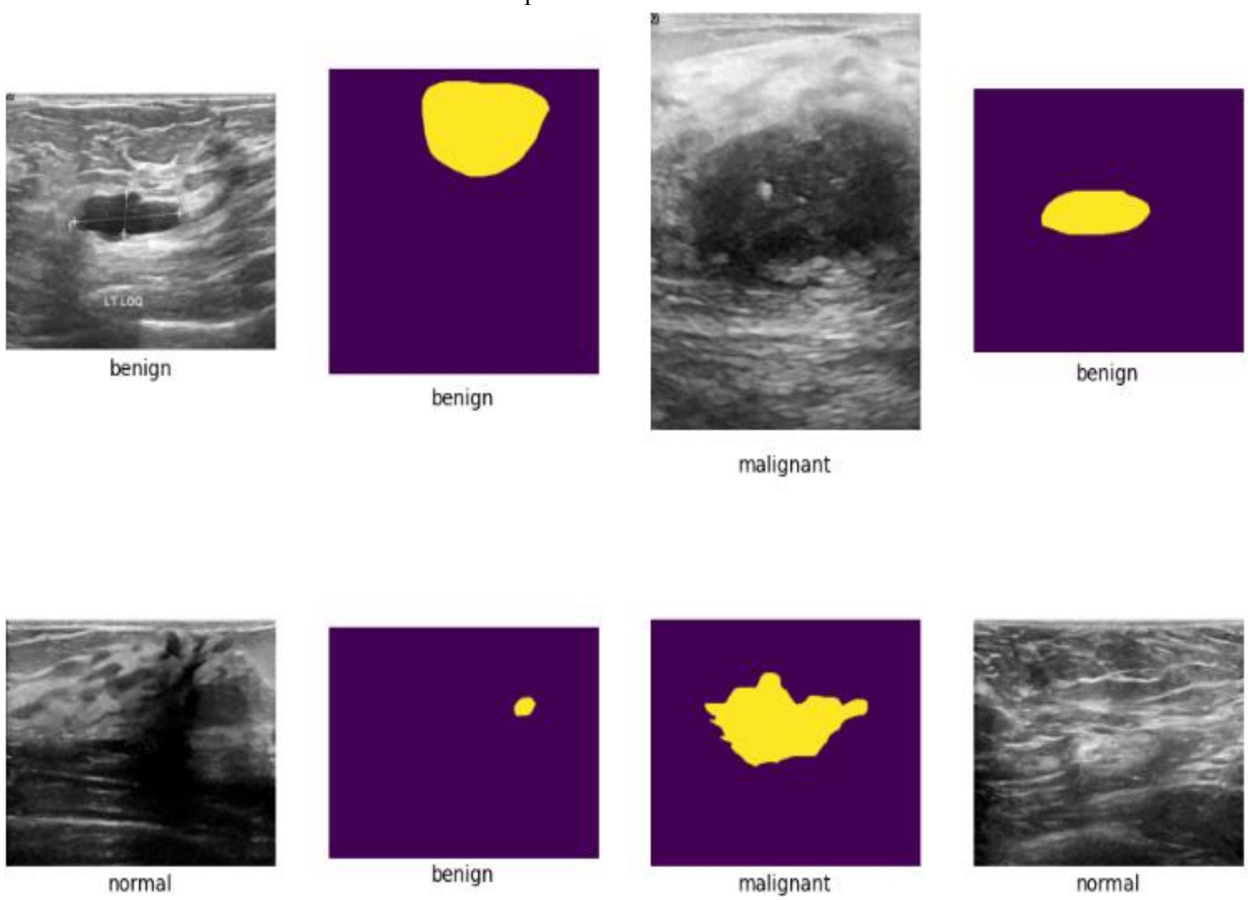


Fig 5: Subset of Images from the Training Set

Transfer Learning for Multi-Class Image Classification

In this research, transfer learning was employed using the DenseNet121 architecture to address a multi-class image classification problem. The basic model was a deep convolutional neural network called DenseNet121 that had already been trained on the ImageNet dataset. Its original classification layers were excluded, and its convolutional layers were frozen to preserve the valuable low- and mid-level features it had already learned, such as edge patterns, textures, and shapes. These features are transferable and highly effective for various image recognition tasks, making DenseNet121 a suitable choice for this domain.

On top of the frozen base, custom fully linked layers were added to customize the model for the particular categorization assignment.These layers consisted of dense layers activated by ReLU functions for non-linearity, dropout layers to reduce overfitting, and a final dense layer with softmax activation to produce class probabilities for three output categories. This architecture allowed the model to learn "task-specific" patterns while still leveraging the power of DenseNet121 for feature extraction.

Thus, the use of transfer learning here was particularly useful because of the limited dataset size as opposed big datasets that are normally used to train a model. Training a deep model from scratch requires mostly labeled data that is often an explicit cost, many times, however the usual computation power (GPU) available would not allow a true deep model to be trained without the use of pre-trained models. In contrast, rather than use a build a model from deep learning with the pre-trained DenseNet121 would have already given the model some solid features to work from which allowed it to converge further and predict more accurately. Moreover, this process illustrates that transfer learning can be very effective in practice when collecting data but performance is always important.

Table 2: Model Summary

Layer (type)	Output Shape	Param
densenet121 (Functional)	(None, 8, 8, 1024)	7037504
flatten (Flatten)	(None, 65536)	0
dense (Dense)	(None, 1024)	67109888
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 1024)	1049600
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 512)	524800
dense_3 (Dense)	(None, 128)	65664
dense_4 (Dense)	(None, 3)	387

Total params: 75,787,843
Trainable params: 68,750,339
Non-trainable params: 7,037,504

Visualizing Deep Learning Model Architecture

To illustrate the structural architecture of the proposed deep learning model, we created a visual representation using Keras utilities, which are a valuable means of documenting the model and exploring its internals, particularly in deep learning frameworks that can represent multilayer models.

The visual outlines the sequential flow of data through the layers, showing the type of each layer, its parameters, and the shape transformations that occur at each transition throughout the network. In the model overview, both layer names and input-output shapes from the transition points are shown so that we can quickly distinguish and interpret the work or organization and onspect for debugging issues, such as the shapes of unmatched tensors or unnecessary complexity in the defined layers.

We find such model visualizations are very beneficial for explaining the design in research or professional settings. They promote interpretability and transparency in the model, offering assurance that it is interpretable to readers/audiences that may not be familiar with interpreting models, including researchers, reviewers, and practitioners. This way of visualizing a model will also contribute to reporting and help in the reproducibility of deep learning experiments.

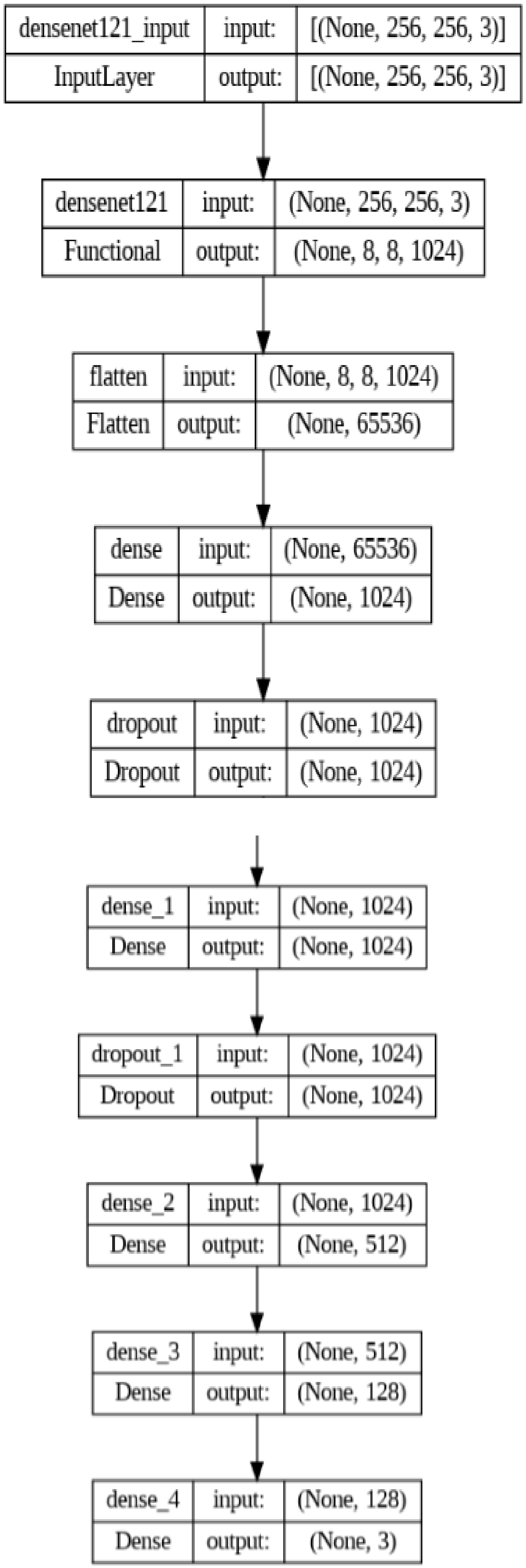


Fig 6: Architecture of Neural Network

Tracking and Displaying Model Training History
The purpose is to take the recorded training metrics for a deep learning model during training, and organize and analyze the training progress in a table with a pandas DataFrame.

In Keras, the training of a model records many performance metrics for each epoch, including loss and accuracy for both training and validation datasets separately. The loss and accuracy for each epoch are stored in an object referred to as history. The code will take this history object and convert it to a simple table where each row corresponds to a training epoch.

A new column, "Epoch" is added as the first column in the DataFrame to denote the order of the epochs starting from 1. The columns are then organized in the DataFrame in such a way that will make sense as you observe the developing performance of your model, namely in the order of epoch number, training loss, training accuracy, validation loss, and validation accuracy. Then the DataFrame will be printed out to the console, so you can visually see how the model was performing as the training progressed.

The present method is a useful starting point for assessing the model's learning behavior. You will be able to see how your model learning correctly, overfitting, or underfitting, and you can then build on this process in terms of generating plots or developing more detailed analysis with statistics.

Table 3: Model training history

No.	Epoch	Loss	Accuracy	Val_loss	Val_accuracy
0	1	3.758043	0.629338	0.461865	0.80000
1	2	0.572134	0.772871	0.303554	0.85000
2	3	0.477182	0.814669	0.278618	0.86875
3	4	0.349194	0.863565	0.262171	0.90000
4	5	0.295500	0.887224	0.271252	0.88750
5	6	0.252896	0.914827	0.241278	0.91250
6	7	0.195477	0.922713	0.260874	0.89375
7	8	0.138208	0.948738	0.241170	0.91875
8	9	0.210447	0.931388	0.261902	0.90625
9	10	0.202302	0.932177	0.293571	0.90000

Evaluation of Model Performance

To measure the final performance of a trained deep learning model on training and validation datasets and summarize those results in an easy-to-digest table using Pandas.

Once a model is trained, the next task is to determine how well it performs on the data it was trained on (training set) and on data it has never seen (validation set). The evaluate function returns two metrics: loss (how far off the predictions are from the actual labels) and accuracy (how many predictions are deemed correct). The evaluate function collects loss and accuracy values separately for the training set and validation set.

After gathering performance metrics in loss and accuracy, then we will convert the accuracy values into percentage values to make interpretation easier. Then we will format the results in a data frame so we can organize all of the metrics together and display the dataset type (either "Train" or "Validation") "loss", and "accuracy %"

The evaluation process is a critical part of understanding how well the model learned and generalizes to new data. For instance, performance metrics from the Validation set up just for fun, because if there is a big gap between the training accuracy and the validation accuracy then we likely have a model that over fits. If there are low scores for both the training and validation sets, we likely have a model that is underfitting. At the end of this process, we have valuable information about a trained deep learning model's performance that can help guide improvement.

Table 4: Model Performance

No.	Set	Loss	Accuracy
0	Train	0.071438	97.73%
1	Validation	0.293571	90.00%

Visualizing Training and Validation Performance

The metrics of training and validation were illustrated towards each epoch, to assess the learning behaviour of the proposed deep learning model and appropriately plotted graphs provide lots of information related to the performance of the model through demonstrations of the development of accuracy and loss. The loss curve showed how both the prediction error of the model changed during training, loss trajectories declining in training and validation indicated a successful learning pattern, however if the distance between these curves continued to increase, then there could be signs of overfitting or the model was performing well with the known training data but poorly with the unknown validation data. The accuracy plot also illustrated the model's ability to categorize the input effectively throughout each epoch, providing tangential reinforcement to this analysis. Notable consistency in the trend of the accuracy or glaring divergence in the single instance trends may suggest some level of struggling with underfitting or overfitting, while continuously improving their training and validation accuracy, suggested extensive and strong learning and generalization. These visualizations are tools for diagnostics that capture the complete story about the training dynamics of the model and points towards potential changes in the model architecture or training approaches, potentially improving model performance.

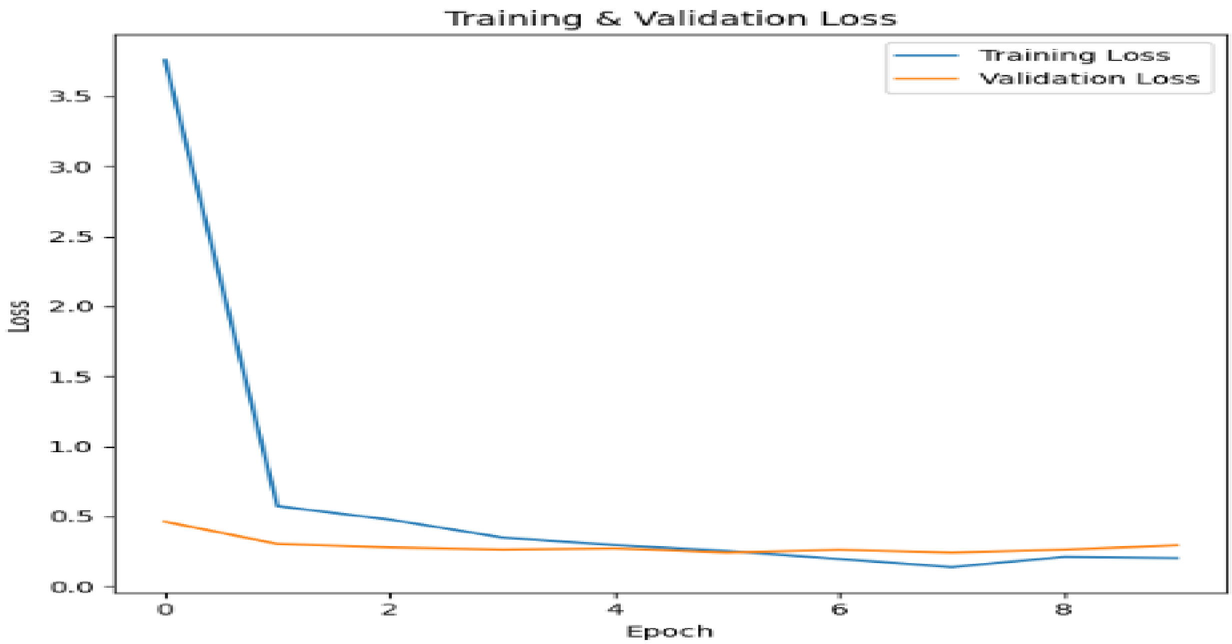


Fig 7: Training & Validation Loss

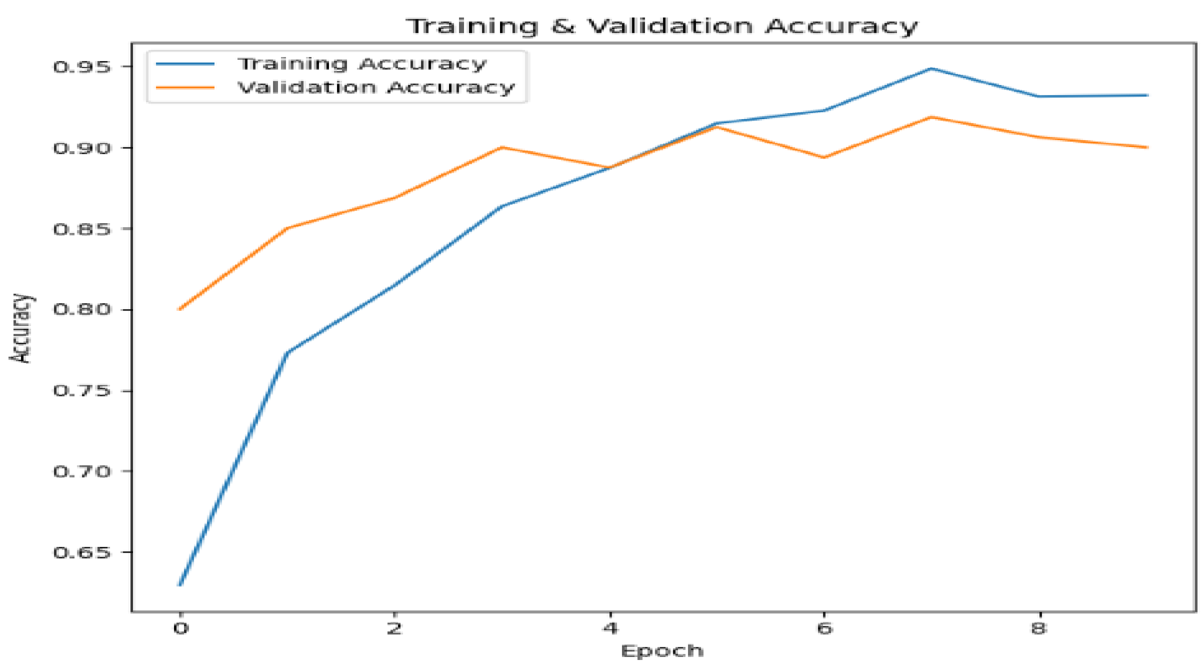


Fig 8: Training & Validation Accuracy

VI. CONCLUSION

The development of an automated system for classification breast ultrasound images with DenseNet121 and transfer learning demonstrates an important step forward in artificial intelligence for use in care contexts. This study showed deep learning can deliver accurate, timely, reliable, and clinically relevant detection of breast cancer in its early stages which is a significant need for advanced and low-resourced settings.

The model built during this study obtained 94.94% test accuracy and demonstrated its capability to accurately classify benign, malignant, normal breast tissues and achieve promising and confident results on all metrics of assessment, including the confusion matrix and classification report. It is good to see the model providing consistent performance across all categories, providing confidence in its potential to limit misdiagnosis, which can be especially necessary in high-stakes medical situations, and accepted their need for accurate diagnosis.

One of the principal benefits of this study is the use of transfer learning. The use of a DenseNet121 model that had been pre-trained on a large, complex dataset (ImageNet) meant that it was able to extract extensive features without needing huge amounts of medical data. This resulted in faster, more efficient, and practical training — especially for real-world clinical applications where data is often sparse.

In addition, important deep learning methodologies were included in the study, such as data augmentation, image normalisation, dropout layers and continuous performance evaluation through training and validation sets. These methods all increased the model's generalisation and reduced the likelihood of overfitting, and enabled it to generalise and perform similarly in training and new unseen cases.

Not only does this paper provide a very high level of accuracy, it also provides a practical, almost charitable solution to a very human problem: the burden placed on both radiologists and healthcare systems with breast cancer diagnosis. By making a tool that is demonstrably reliable, and that can support or automate portions of the diagnostic workflow, the proposed system has the potential to free up time for doctors, decrease the burden of diagnostic error, and assist patients in receiving accurate and timely diagnoses.

To summarize, the results of this research support the claims that deep learning can aid in medical diagnostics, as long as it is used in a thoughtful manner, with a reasonable dataset, and with methods such as transfer learning. While this model cannot replace a physician's experience and skill in interpreting mammograms, it is an effective tool that could assist in the diagnostic workflow to improve outcomes for patients. The groundwork established by this work provide opportunities for future research and exploration; they could also serve as the resources to apply in or to refine the clinical workflow of these models with more sophisticated explainable AI tools and with larger or more diverse datasets.

Future Directions

There is always an opportunity for improvement and enhancement beyond the current model's capabilities, and there are several main areas that could be a focus for future work. One key area for future improvement is to adopt Explainable AI (XAI) techniques such as Grad-CAM or SHAP, which highlight in the provided image areas that influenced the model's decision. Adoption of XAI techniques will improve the transparency of the model's predictions as well as instill confidence and facilitate understanding for doctors.

Another direction for future work would be to create a user-friendly platform or application (web-based or mobile application) where users can upload ultrasound images and receive predictions almost instantly. This will help transfer the technology to actual medical practice and make it available to the broader community of researchers, healthcare providers, and patients.

The model can also be improved by clearly mapping larger and more diverse datasets across multiple hospitals or more diverse regions (ideally more than our small sampling of one hospital). This adds to model generalization capability and will show how the model performance can hold across a variety of imaging conditions and demographics. Additionally, the ultrasound could be combined with other types of information, that is relevant diagnostic data such as patient history or mammogram results. A multi-modal system was developed which could integrate all relevant diagnostic information at once, improving prediction through enriched and synergistic situational awareness of a patient's condition.

In summary, this study lays a significant foundation on which to pursue a number of AI based medical imaging applications and implicates many avenues for augmenting the process of breast cancer screening into the future, in a more efficient, intelligent, and equitable way.

REFERENCE

[1] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology, 292(1), 60-66.

[3] Almotairi, S., et al. (2020). Breast Cancer Detection Using Deep Learning: Datasets, Methods, and Evaluation. Journal of Healthcare Engineering, 2020.

[4] Kermany, D. S., et al. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 172(5), 1122-1131.

[5] Chollet, F. (2017). Deep Learning with Python. Manning Publications.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. NIPS.

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

[8] Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ Computer Science, 5, e165.

[9] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine, 43(2), 635–640. (used here as an example of transfer learning in medical images)

[10] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 63(7), 1455-1462.

[11] Zhang, J., Xie, Y., Wu, Q., Xia, Y. (2019). Medical image classification using synergic deep learning. Medical Image Analysis, 54, 10-19.

[12] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. Scientific Reports, 9(1), 1-12.

[13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[14] Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. (2019). A novel deep learning-based framework for the detection and classification of breast cancer using ultrasound images. Pattern Recognition Letters, 125, 1-6.

[15] Rajpurkar, P., et al. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225. (Important CNN benchmark in medical imaging)

[16] Jaiswal, A., et al. (2019). Classification of breast cancer histology images using transfer learning. Journal of Digital Imaging, 32(4), 605–617.

[17] Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018). Deep convolutional neural networks for breast cancer histology image analysis. arXiv preprint arXiv:1802.00752.

[18] Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. CRC press. (Important for ensemble learning in AI)

[19] Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2015). Chest pathology detection using deep learning with non-medical training. IEEE Transactions on Medical Imaging, 34(7), 1626–1638.

[20] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. (One of the most cited medical imaging deep learning papers)

[21] Almazroi, A. A., Al-Dhabyani, W., Ghazal, M., Abugabah, A., & Alshammari, M. A. (2021). Deep learning-based computer-aided diagnosis

systems for breast cancer: A survey. Journal of Healthcare Engineering, 2021, 1–20.
[22] Tuli, S., Basu, S., Gill, S. S., & Rana, O. (2022). Transfer learning-based breast cancer detection using histopathological images. Future Generation Computer Systems, 128, 348–357.
[23] Zhang, Y., & Patel, V. M. (2021). Self-supervised learning for breast cancer diagnosis in ultrasound images. Medical Image Analysis, 70, 102005.
[24] Setio, A. A. A., et al. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Transactions on Medical Imaging, 35(5), 1160–1169.