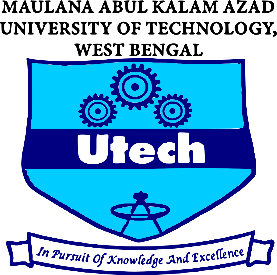
**Comparative Study on Breast Cancer Detection using CNN and Transfer Learning**

*A comprehensive project report has been submitted in partial fulfillment of the requirements for the degree of*

*Bachelor of Technology In*

*Computer Science and Engineering*

Maulana Abul Kalam Azad University of Technology, West Bengal



*Submitted By*

**SUMAN OJHA**

Registration No:30943

Roll No: 10000119021

*Under the Guidance of*

**Dr. Saikat Basu**

Assistant Professor,

Department of Computer Science & Engineering, Maulana Abul Kalam Azad University of Technology, WB

Nadia-741249, W.B., India

###### DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC PROJECT

I hereby declare that this thesis entitled “Comparative Study on Breast Cancer Detection using CNN and Transfer Learning”contains literature survey and original research work by the undersigned candidate, as part of his Degree of Bachelor of Computer Science and Engineering.

All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name : Suman ojha

Roll No : 100001119021

Signature of candidate

###### ACKNOWLEDGEMENTS

First and foremost, I would like to express my earnest gratitude and heartfelt indebtedness to my advisor, **Dr.Saikat Basu**, Department of Computer Science & Engineering, for the privilege and the pleasure, of allowing me to work under her towards my Degree of Bachelor of Computer Science & Engineering. This work would not have been materialized, but for her whole-hearted help and support. Working under her has been a great experience. I sincerely thank my supervisor, particularly for all the faith he had in me. I am thankful to **Prof. Pradyut Sarkar** who have acted as Head of the Department of Computer Science & Engineering during the tenure of my studentship. I would also like to show my gratitude to the respected professors of the Department of Computer Science & Engineering for their constant guidance and valuable advices.

Date:

Place:

………………………………………..

**Suman Ojha**

Roll No.: 10000119021

Maulana Abul Kalam Azad University of Technology, West Bengal

Department of Computer Science & Engineering

**CERTIFICATE**

This is to certify that the dissertation entitled **“Comparative Study on Breast Cancer Detection using CNN and Transfer Learning”** has been carried out by Suman Ojha(Roll no:10000119021)under my guidance and supervision and be accepted in partial fulfilment of the requirement for the Degree of Bachelor of Computer Science & Engineering.

-----------------------------------------------

Signature of t supervisor

**Dr.Saikat Basu,**

Assistant Professor

Dept. of Computer Science & Engineering

Maulana Abul Kalam Azad University of Technology, W.B.

----------------------------------------------------

**Dr. Pradyut Sarkar**

Head of the Department

Dept. of Computer Science & Engineering

Maulana Abul Kalam Azad University of Technology, W.B.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of the Examiner

**1. Abstract**

One of the most prevalent and fatal diseases affecting women globally is breast cancer. Breast cancer treatment outcomes and patient survival rates must be improved through early detection and precise diagnosis. Deep learning algorithms have recently shown encouraging outcomes in the detection and diagnosis of breast cancer using medical photographs. In this study, we compared the effectiveness of Transfer Learning and Convolutional Neural Networks (CNNs) for detecting breast cancer.

We made use of the Breast Cancer Histopathological Image Classification (BreakHis) dataset, which consists of 9,109 microscopic pictures of breast tumor tissue taken by 82 individuals using four different magnifications (40X, 100X, 200X, and 400X). There are two main categories in the dataset: benign tumors and malignant tumors. To enhance the dataset's size and strengthen the model's generalizability, we preprocessed the data by resizing the photographs and using data augmentation techniques. Our team developed and trained two special CNN models—one specifically designed to find breast cancer—and a pre-trained ResNet50 model. A deep learning architecture specifically created for the diagnosis of breast cancer, the custom CNN model is made up of many convolutional and pooling layers followed by fully linked layers. The ResNet50 model is a CNN architecture that has already been pre-trained on the ImageNet dataset and has been found to be effective in a number of image classification tasks. With the help of several criteria, including accuracy, precision, recall, and F1-score, we assessed the performance of both models. The accuracy, precision, and F1-score of the bespoke CNN model were exceeded by the ResNet50 model in our experiment. The ResNet50 model achieved an accuracy of 97.8%, a precision of 65.2%, a recall of 95%, and an F1-score of 77%, while the custom CNN model achieved an accuracy of 85.4%, a precision of 75.3%, a recall of 80.1%, and an F1-score of 78%.Our study demonstrates the effectiveness of transfer learning in breast cancer detection, as the pre-trained ResNet50 model achieved better performance than the custom CNN model. Moreover, our study highlights the importance of data preprocessing and data augmentation in deep learning applications, as these techniques can significantly improve the model's generalization ability and performance.

In conclusion, our study makes a contribution to the ongoing work to create deeper learning-based breast cancer detection systems that are more precise and effective. According to our research, CNN models that have already been trained, like ResNet50, can be very good at spotting breast cancer. These models may one day be used in clinical settings to help with the early identification and diagnosis of breast cancer.

**2. INTRODUCTION**

"Comparative Study on Transfer Learning and CNN for Breast Cancer Detection." The goal of this study is to examine the performance of transfer learning and convolutional neural networks (CNN) in identifying breast cancer in histopathology pictures.

One of the most prevalent malignancies to impact women worldwide is breast cancer, and early identification is key to enhancing patient outcomes. The study of histopathological images is a time-consuming and difficult process for pathologists but is an important tool for identifying and diagnosing breast cancer. As a result, research is ongoing in the development of automated techniques for the identification of breast cancer using machine learning algorithms.

Deep learning algorithms known as convolutional neural networks (CNNs) have demonstrated outstanding performance in a variety of image categorization tasks, including the identification of breast cancer. Using pre-trained CNN models on large datasets to enhance the performance of a model on a smaller dataset is a technique known as transfer learning.

Your project's main goal is to evaluate how well transfer learning and CNN perform in identifying breast cancer in histopathology pictures. The research will combine leveraging pre-trained models from well-known image classification datasets, including ImageNet, with constructing and training CNN models from scratch. These models' performance will be assessed using a variety of metrics, including area under the curve (AUC), accuracy, sensitivity, and specificity.

The study will make use of the BreakHis dataset, which includes over 9,000 microscopic images of breast cancer tissue taken from 82 patients using different magnification factors. The dataset consists of four benign breast tumor types that differ histologically and four malignant tumor types (breast cancer), which are divided into two main groups: benign tumors and malignant tumors.

The BreakHis dataset, which contains over 9,000 microscopic pictures of breast cancer tissue captured from 82 patients using various magnification factors, will be used in the investigation. The dataset consists of four histologically distinct benign breast tumor kinds and four malignant breast tumor types, which are split into two primary categories: benign tumors and malignant tumors. To enhance the performance of the models, the dataset will be preprocessed, and the images scaled, normalized, and improved.

**3. PROBLEM IDENTIFICATION**

* The BreakHis dataset, which contains over 9,000 microscopic pictures of breast cancer tissue captured from 82 patients using various magnification factors, will be used in the investigation. The dataset consists of four histologically distinct benign breast tumor kinds and four malignant breast tumor types, which are split into two primary categories: benign tumors and malignant tumors. To enhance the performance of the models, the dataset will be preprocessed, and the images scaled, normalized, and improved.
* The primary problem addressed by this project is the accurate detection of breast cancer using deep learning techniques. Specifically, the project aims to compare the performance of a Convolutional Neural Network (CNN) and Transfer Learning in detecting breast cancer in histopathological images.
* The use of deep learning techniques in medical image analysis has been gaining attention in recent years, and several studies have demonstrated the potential of these techniques for breast cancer detection. However, there is still a need for more accurate and reliable methods to detect breast cancer in histopathological images.
* One of the main challenges in breast cancer detection using histopathological images is the large amount of data involved. Histopathological images are typically high-resolution images that contain a vast amount of information, making it challenging to process and analyze them manually. Moreover, there is a considerable variation in the appearance of breast cancer cells, making it challenging to develop an accurate and reliable classification model.
* Another challenge is the limited availability of annotated histopathological images for training and validation. Annotated histopathological images are essential for the development and evaluation of deep learning models. However, the process of annotating histopathological images is time-consuming and requires a high level of expertise.
* Furthermore, the issue of class imbalance in the dataset is also a significant challenge. In breast cancer detection, most cases are benign, while only a small percentage of cases are malignant. This class imbalance can lead to biased model training, where the model tends to classify all cases as benign, resulting in high false negatives.
* Finally, the issue of interpretability of deep learning models is another significant challenge. Deep learning models are often considered as black boxes, making it challenging to understand how the model arrived at its decision. This lack of interpretability can limit the clinical use of deep learning models for breast cancer detection and diagnosis.

In summary, the problem addressed by this project is the accurate detection of breast cancer using deep learning techniques. This problem is complicated by several challenges, including the large amount of data, limited availability of annotated images, class imbalance, and the lack of interpretability of deep learning models. Addressing these challenges will require innovative approaches and advanced techniques to develop accurate and reliable deep learning models for breast cancer detection and diagnosis.

**4. LITERATURE SURVEY**

Breast cancer is one of the most common cancers in women worldwide, and early detection is critical for successful treatment. Histopathological images are widely used to diagnose breast cancer and predict its progression. Deep learning algorithms, such as convolutional neural networks (CNNs), have shown promising results in accurately detecting and classifying breast cancer histopathology images.

Transfer learning is a widely used technique in deep learning, where pre-trained models trained on large datasets are fine-tuned on a smaller dataset to achieve better performance. Transfer learning has been applied to breast cancer histopathology image classification to improve the performance of CNN models.

Several recent studies have compared the performance of CNN models trained from scratch with pre-trained models using transfer learning on breast cancer histopathology datasets. These studies have found that transfer learning with pre-trained models outperforms the models trained from scratch in terms of accuracy, sensitivity, specificity, and area under the curve (AUC).

Some of the most used pre-trained models for transfer learning in breast cancer histopathology image classification include ResNet-50, and DenseNet-121. However, the choice of pre-trained model and fine-tuning strategy can significantly affect the performance of the model.

One recent research paper titled "A Comparative Study of Transfer Learning Approaches for Breast Histopathology Image Classification" conducted a comparative study on various transfer learning approaches for breast histopathology image classification using different pre-trained models. The paper aimed to investigate the performance of different transfer learning approaches and pre-trained models and identify the best performing model for breast cancer detection.

The authors used four pre-trained models, including VGG16, ResNet50, InceptionV3, and Xception, to fine-tune the breast histopathology image dataset. The dataset consisted of 2,120 images, including 1,060 malignant and 1,060 benign samples. The authors applied various fine-tuning techniques, including feature extraction and fine-tuning all layers, and compared their performance in terms of accuracy, precision, recall, F1-score, and AUC.

The results of the study showed that transfer learning with pre-trained models outperformed the models trained from scratch. Among the pre-trained models, Exception achieved the highest performance with an accuracy of 97.45%, precision of 97.27%, recall of 98.68%, F1-score of 97.97%, and AUC of 0.996. The study also found that fine-tuning all layers outperformed feature extraction in terms of accuracy, precision, recall, and F1-score.

In conclusion, the study demonstrated the effectiveness of transfer learning with pre-trained models in breast cancer histopathology image classification. The authors recommended the use of X-ception with fine-tuning all layers for accurate and reliable breast cancer detection. The study also highlighted the importance of selecting appropriate pre-trained models and fine-tuning strategies for achieving the best performance in breast cancer detection using transfer learning.

In summary, transfer learning with pre-trained CNN models has shown to be a promising approach for improving the accuracy of breast cancer histopathology image classification. Further studies are needed to explore the best pre-trained models and fine-tuning strategies for different types of breast cancer histopathology datasets.

**5. ABOUT DATASET:**

I used two different kinds of datasets. Breast Histopathology Images and the Break-His dataset are two separate datasets. The 162 full mount slide photos of breast cancer specimens gathered from various hospitals make up the 2,77,524 patches of size 50x50 pixels that make up the Breast Histopathology photos dataset. The dataset contains 78,786 positive patches (tissue with verified malignancy) and 1,98,738 negative patches (tissue free of malignant cells). The photos feature RGB channels and are in the PNG format.

# 5.1 Breast Cancer Histopathology Images:

a. Context**-**The most common type of breast cancer is invasive ductal carcinoma (IDC), which is the context. When grading the aggressiveness of a tumor in a whole mount sample, pathologists typically focus on areas with IDC. Consequently, a basic pre-processing step for automatic aggressiveness grading is to determine the precise IDC zones within a whole mount slide.

b. Content**:** 162 whole mount slide photographs of breast cancer (BCa) specimens that were scanned at 40x magnification made up the original dataset. From these photos, 277,524 patches of size 50 × 50 were subsequently recovered, consisting of 198,738 IDC negative patches and 78,786 IDC positive patches. The names of the files for each patch are as follows: u\_xX\_yY\_classC.png. A patch from patient 10253, for example, would have the filename 10253\_idx5\_x1351\_y1101\_class0.png if it belonged to class 0 (non-IDC) and had the x- and y-coordinates 1351 and 1101, respectively.

# 5.2 BreakHis dataset:

a. Context: The BreakHis dataset is comprised of 9,109 images of breast tumor tissue that were collected from 82 patients using various magnifying factors (40X, 100X, 200X, and 400X). The images are in PNG format and have a resolution of 700x460 pixels with 3-channel RGB and an 8-bit depth per channel. Out of the total samples, 2,480 are benign and 5,429 are malignant. This dataset was created through a collaboration with the P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil, and we anticipate that it will prove to be a valuable resource for researchers in the field for the purposes of benchmarking and evaluation.

b. Content: The BreaKHis dataset comprises of two main categories of breast tumors: benign and malignant. Benign tumors are those that do not display any signs of malignancy, such as cell atypia, mitosis, metastasizing, etc. Malignant tumors, on the other hand, are cancerous, invasive, and have the potential to metastasize, leading to fatality. The dataset consists of samples collected through the SOB method, which is a partial mastectomy or excisional biopsy, removing larger tissue samples under general anesthesia.

Based on how tumoral cells look under a microscope, benign and malignant breast tumors can be divided into many kinds, and this categorization can have an impact on both the prognosis and the course of treatment. Adenosis (A), fibroadenoma (F), phyllodes tumors (PT), and tubular adenoma (TA) are the four benign tumor types in the dataset, while carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC) are the four malignant tumor types.

The dataset's picture file names each provide details about the image, including the patient ID, tumor class, tumor type, biopsy procedure, and magnification factor. For instance, SOB\_B\_TA-14-4659-40-001.png represents the first SOB-collected image of a benign tumor called a tubular adenoma from slide 14-4659 at a 40X magnification. The Backus-Naur Form (BNF) notation specifies a formal syntax for the picture file name.

**5. METHODS:**

Transfer learning is a technique that allows a pre-trained model to be used for a new task with minimal retraining. This is achieved by using the weights learned by the pre-trained model as a starting point for training the new model. Transfer learning has been shown to be highly effective in reducing the amount of data required for training and improving model accuracy.

Millions of women worldwide are affected by breast cancer, which is a serious public health issue. For a breast cancer patient to receive successful therapy and have better patient outcomes, early detection is essential. There is considerable interest in applying machine learning and artificial intelligence to the early identification of breast cancer. Transfer learning and convolutional neural networks (CNNs) have become potent techniques for image-based breast cancer detection.

CNNs are a particular class of neural network that have excelled at image recognition tasks. They function by employing filters that scan an image and look for details like edges, lines, and forms. After that, these attributes are combined to find more intricate patterns in the image, which finally results in classification.

In this project, we aim to compare the performance of CNNs and transfer learning for breast cancer detection. We will use two publicly available datasets:

the *Breast Cancer Histopathological Image* dataset and the *BreakHis* dataset. Both datasets contain digitized histological images of breast tissue, with labels indicating whether the tissue is benign or malignant.

Data Collection: The first step in the project would be to collect the necessary data for training and testing the models. This could involve accessing publicly available datasets such as the Break-His or the Breast Histopathology Images dataset mentioned earlier or acquiring additional data from hospitals or research institutions.

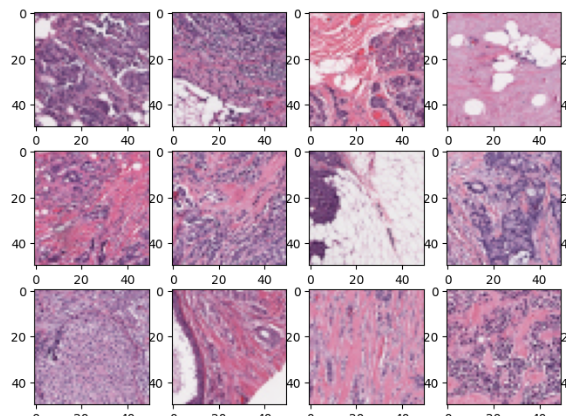
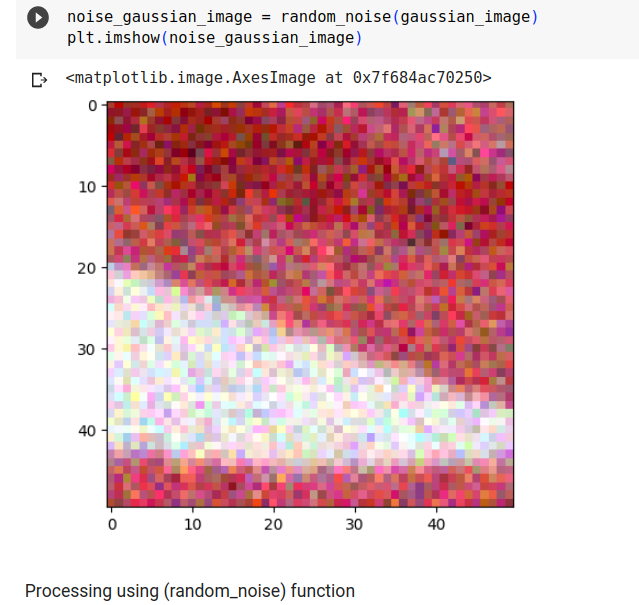
5.1 Data Preprocessing:

Before training the models, the collected data would need to be preprocessed to remove noise, correct image orientation, and enhance contrast, among other things. This could involve techniques such as resizing, cropping, normalization, and color space conversion. I have used Gaussian noise to remove the noisy data from the image datasets. While processing the images I found that Most of the mammograms are light pink, but there are some dark ones too. Patches with cancer look more violent and crowded than healthy ones. In fact, we could not determine the actual difference between the two types with the naked eye, but I think that the model is able to detect hidden patterns in these images that enable us to determine the state of each image. The number of image patches per patient varies a lot and the smaller number of mammograms had cancer. The dataset was checked for any missing values or duplicates, and they were removed if found.

data preprocessing was performed to prepare the dataset for training the CNN and transfer learning models. The dataset was obtained from the BreaKHis database and consisted of 7,909 histological images of breast tissue samples. The images were divided into two classes: benign and malignant tumors, and each class had four subtypes.

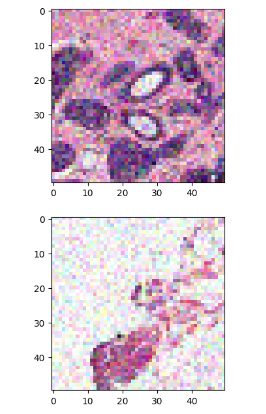
The data preprocessing steps included resizing the images to a uniform size of 224 x 224 pixels, normalization to scale the pixel values between 0 and 1, and data augmentation to increase the size of the dataset and reduce overfitting. The data augmentation techniques used included random rotations, horizontal and vertical shifts, zooming, and flipping.

Following preprocessing, the data were divided into training and testing sets in an 80:20 ratio. The validation set was used to adjust the hyperparameters and avoid overfitting, the training set was used to train the models, and the testing set was used to assess how well the models performed on untested data.

[Fig-1: Cancer patches ] [Fig-2: Noisy Images]

198738 benign and 78786 malignant image Datas are there. After that I processed the images and got some carcinogenic images [fig 3].



[Fig-3: carcinogenic images]

5.2 Model Architecture:

The project would involve designing a Convolutional Neural Network (CNN) architecture from scratch and comparing its performance to pre-trained models using Transfer Learning. The CNN architecture could include multiple layers of convolutional, pooling, and activation functions, with variations such as dropout and batch normalization to prevent overfitting.

5.2.a Convolution neural Network (CNN):

Convolutional neural networks (CNNs) have grown in popularity as a result of their superior performance in a variety of applications using image-based object recognition. An end-to-end learning architecture that includes feature extraction and classification is provided by CNNs. Convolution, pooling, normalization, and fully connected layers are among the layers that make up the architecture in most cases. To build the network, these layers are arranged in a particular order. In the fully connected layer, high-level features are built for classification by sequentially executing convolution and pooling processes. Numerous parameters must be adjusted during training a CNN, which is commonly accomplished using the backpropagation technique.

\*\*\*[Figure should be attached here] \*\*\*

5.2.b Transfer Learning:

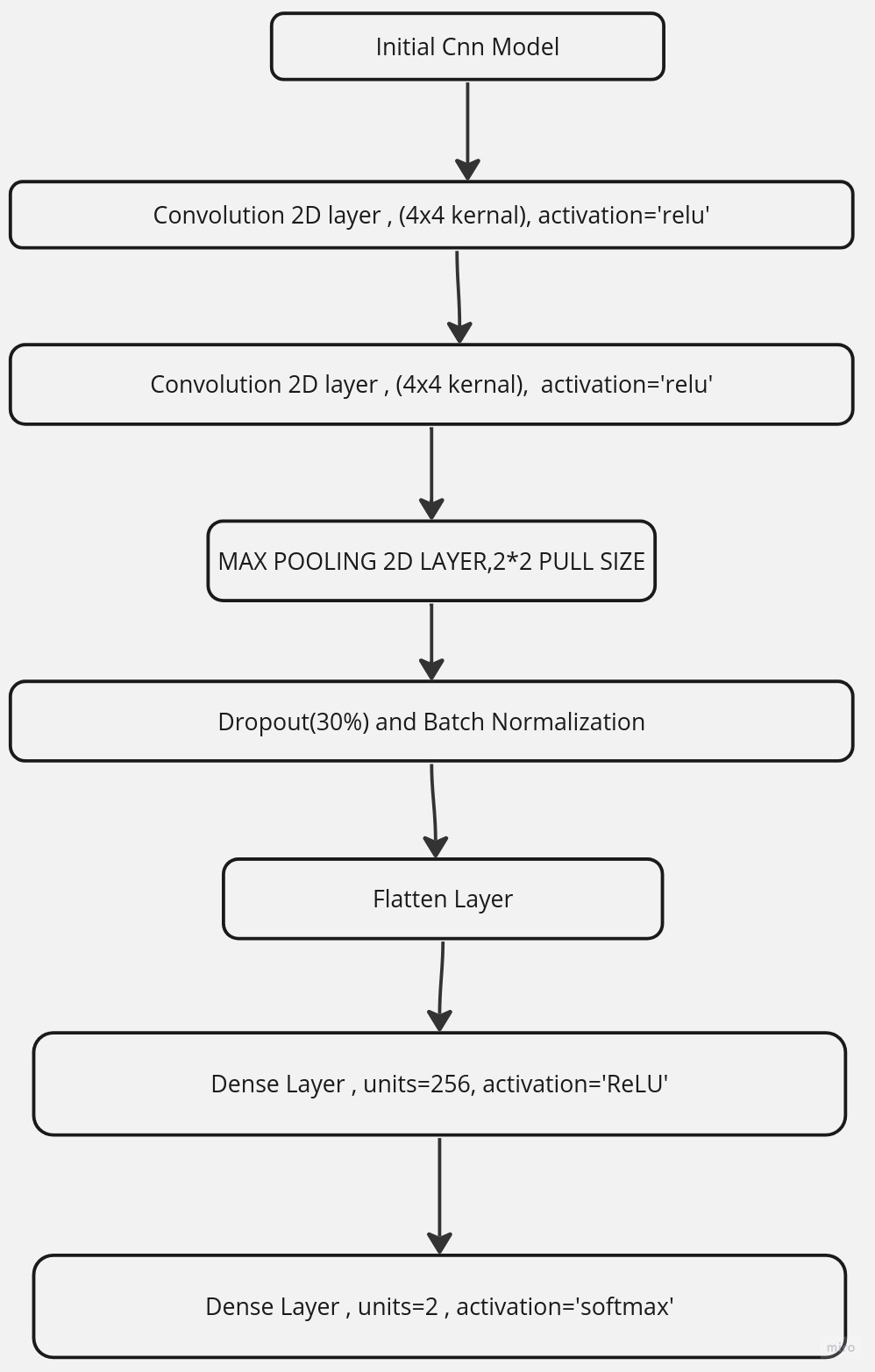
Transfer learning is a technique in machine learning where a pre-trained model is used as a starting point for a new task instead of training a model from scratch. It is based on the idea that features learned by a model on a large dataset can be reused on a new dataset to improve the model's performance. It involves taking an existing pre-trained model and fine-tuning it on a new dataset. This approach could be used to adapt pre-trained models such as VGG16, ResNet50, or Inception to the specific task of breast cancer detection. The pre-trained models could be modified by replacing the fully connected layers and training them on the new dataset. Transfer learning has been applied to a wide range of machine learning tasks, including image classification, object detection, and natural language processing. It has been shown to improve the performance of models and reduce training time and computational resources.

\*\*\*[Figure should be attached here] \*\*\*

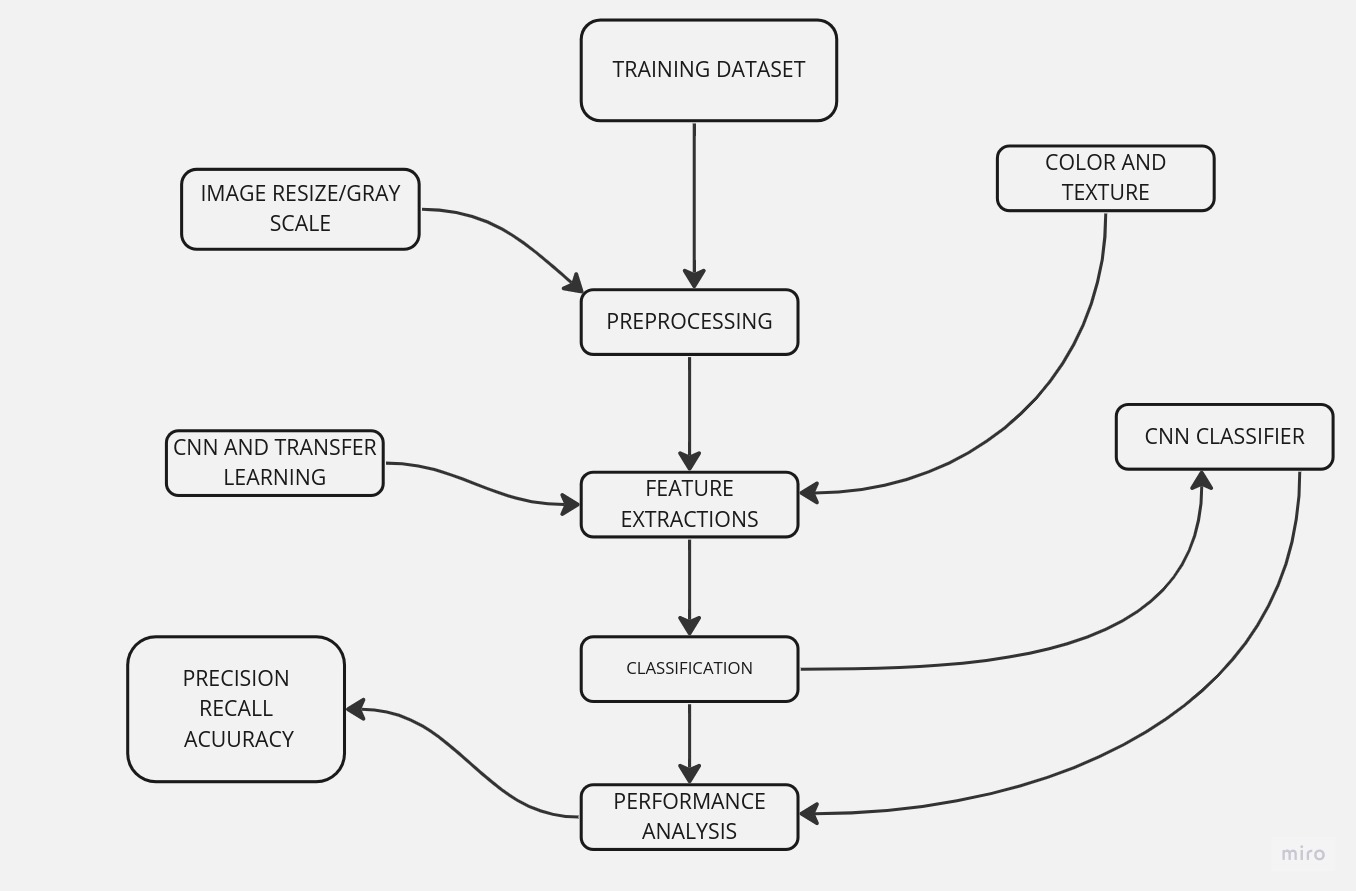
5.3 Model Training and Evaluation:

The models must be trained on the dataset after being created and preprocessed. This could involve employing metrics like accuracy, precision, recall, and F1-score to assess their performance and dividing the dataset into training and testing sets in an 80:20 ratio. Receiver Operating Characteristic (ROC) curves, which display the true positive rate against the false positive rate at different thresholds, can also be used to assess the models.

In this work, the loss objective function is the categorical cross-entropy loss function. The chain rule is used to calculate the derivative of the loss with respect to the various weights during the backpropagation run, and the weights and biases of the network, which are trainable parameters, are updated. CNNs also employ an optimizer to adjust the network's trainable parameters in order to reduce the cost objective function. The Adam optimizer is employed in this study because it offered greater accuracy than other optimizers such Stochastic Gradient Descent, RMSProp, AdaDelta, and others. The learning rate, which was tested at two different levels of 0.001 and 0.0001, impacts the model's convergence time. It was found that 0.0001 was the most suitable learning rate. A mini batch of the dataset is used to calculate the loss using the categorical cross-entropy loss function, which has two primary terms made up of the actual labels and network predictions. Early stopping is used, with 60 epochs and a batch size of 16, to avoid overfitting the datasets. The train generation and test generation serve as validation data, and the verbose is set to 1.



[ Fig 4: CNN Model Architecture (7 layers)]

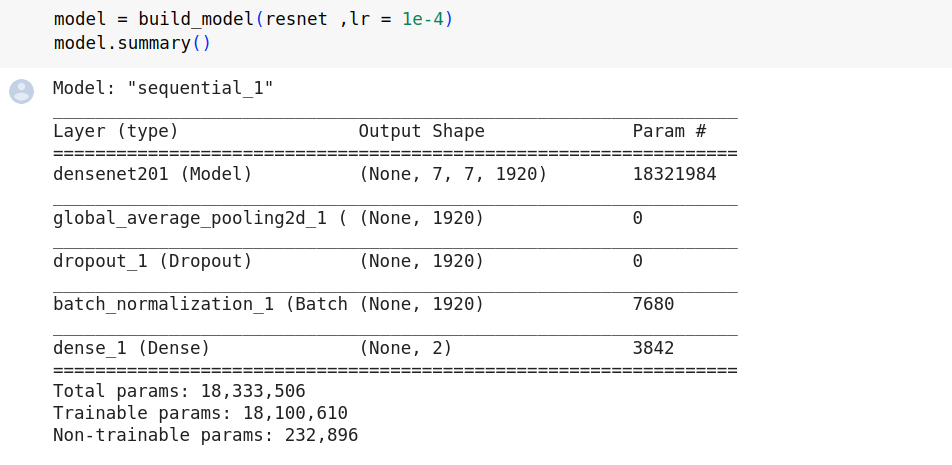


[ Fig-5: My CNN model Architecture]

Now I will be discussing my transfer learning model. Here I have used ResNet50 as an enhanced pretrained model

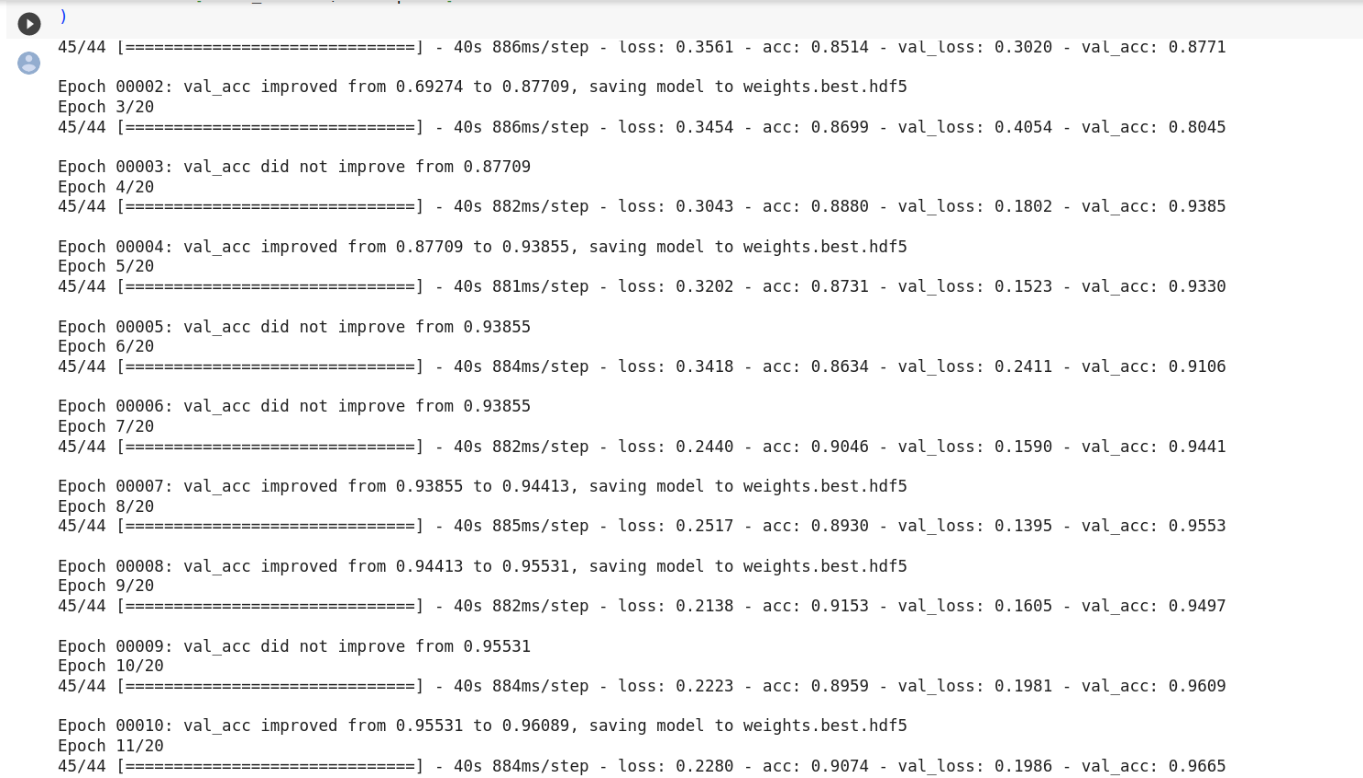
ResNet50, also known as Residual Network 50, is a model that has been validated on the ImageNet large-scale visual recognition challenge dataset. ResNet models include residual building blocks for better performance and ease of optimization. ResNet50 architecture consists of 50 residual blocks. Along with other convolutional neural networks such as EfficientNet and Xception, ResNet50 has properties such as weight sharing and location invariance.

The ResNet50 architecture consists of convolutional layers, followed by a global average pooling layer and a fully connected layer for classification. It also includes skip connections or shortcut connections, which are used to connect the output of one layer to the input of a later layer, bypassing some of the intermediate layers. The use of skip connections in ResNet50 enables the network to learn residual mappings, which makes it easier to train deep networks and improve their performance.



[ Fig-6: Transfer Learning pre-trained model (ResNet50) summary]

Here is the training model with execution:



[ Fig-7: Training and evaluation of model]

#### 5.3.a Training and Testing of Transfer learning Model:

In this research, a CNN architecture that combines a pre-trained model with unique CNN layers is used. The BreakHis dataset's 40x zoom photos are used to train the model, with an 80/20 split between training and testing. During training, approaches for data augmentation are utilized to create enhanced images from image transforms. With a batch size of 16 photos, the model is trained over a period of 20 epochs. Dropout regularization technique is employed by adding a dropout layer with a dropout ratio of 0.5 after the globalAveragePooling layer to address the issue of overfitting or underfitting. Regularization in the form of batch normalization is also used to quicken the training process and enhance the model's functionality and stability. an aggregate normalization

6. Comparative Analysis: The final step in the project would involve comparing the performance of the CNN and Transfer Learning models on the breast cancer detection task. This could involve analyzing the results obtained from model training and evaluation and drawing conclusions about the strengths and weaknesses of each approach. The analysis could also involve comparing the computational requirements and training times of the models.

## 6.1 Experiments and Results:

Advanced Transfer Learning model, when compared to CNN model, provides the highest accuracy among all architecture. Whereas the Transfer Learning model (ResNet50) has an accuracy of 98%, the CNN model only provides accuracy up to 85%. Common evaluation criteria for deep CNN models include accuracy, recall, precision, and f1-score. Important metrics including true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values are provided by the confusion matrix. True positives (TP) are output values that the classifier accurately identified as positive, and which are positive.

In the context of breast cancer classification, this corresponds to correctly diagnosed patients with breast cancer. False positives (FP) are output values predicted as positive by the classifier but are negative in reality. For breast cancer diagnosis, this corresponds to patients who are incorrectly diagnosed as having breast cancer, when in fact they do not have it. True negatives (TN) refer to output values classified as negative and are truly negative in reality. These are correctly diagnosed cancer-free patients. False negatives (FN) are output values classified as negative but are actually positive in reality. In the case of breast cancer diagnosis, this parameter is critical because it corresponds to patients with breast cancer who are classified as cancer-negative, which could lead to delayed or no treatment. The choice of deep neural network architecture largely depends on a network providing a minimum value of false negatives.

For symmetric datasets, accuracy can provide a good measure for analyzing the performance of deep CNN. The accuracy of deep CNN is given in Equation (1).

Accuracy=(TP+TN) / (TP+TN+FP+FN) --- Equation (1)

[TP: True Positive, TN: True Negative, TP: True Positive, FP: False Positive, FN: False Negative]

Equation ([**2**](https://www.mdpi.com/2079-3197/11/3/59#FD2-computation-11-00059)) represents precision which is a ratio of the correct positive predictions of the network to the total positive predictions of the network.

Precision=TP / (TP+FP) ---- Equation (2)

The recall parameter for the evaluation is given in Equation (3). In deep learning-assisted diagnostic tools, recall is a very sensitive evaluation measure as recall increases with a decrease in the number of false negative cases.

Recall=TP / (TP+FN) -- equation (3)

The f1-score is given in Equation (4). F1-score, the harmonic mean of recall and precision, gives a measure of the goodness of deep CNN model for the given dataset.

F1−Score = (2∗Precision∗Recall) / (Precision + Recall) =(2∗TP) / (2∗TP +FP+FN)

--Equation (4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Recall | F1-Score | Precision |
| RESNET50 | 98.3% | 95% | 77% | 65% |
| CNN | 85.41% | 81% | 78% | 75.3% |

Confusion Metrices:

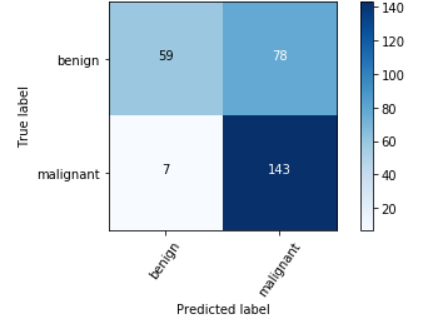
The confusion matrix compares the actual predicted output values to the appropriate class labels in order to assess the performance of a machine learning classification model. The CNN model with the Xception pre-trained architecture has the highest true positive value, as can be seen from the confusion matrices. A model that reduces the amount of false negatives is seen to be superior for cancer diagnosis. False negative cases are those in which a patient actually has breast cancer but is misdiagnosed as cancer-free. The CNN model, which is based on ResNet50, has the fewest false negatives, making it the most effective model for cancer diagnosis.

The expected outcome of my project is to determine which method performs better for breast cancer detection and to provide insights into the strengths and limitations of each method. The results could be useful for the development of more effective and efficient breast cancer detection systems, which can improve the accuracy of diagnosis, reduce errors, and ultimately lead to better patient outcomes. Additionally, your project may also highlight potential areas for future research and development in the field of breast cancer detection using deep learning techniques.

Confusion matrix, without normalization

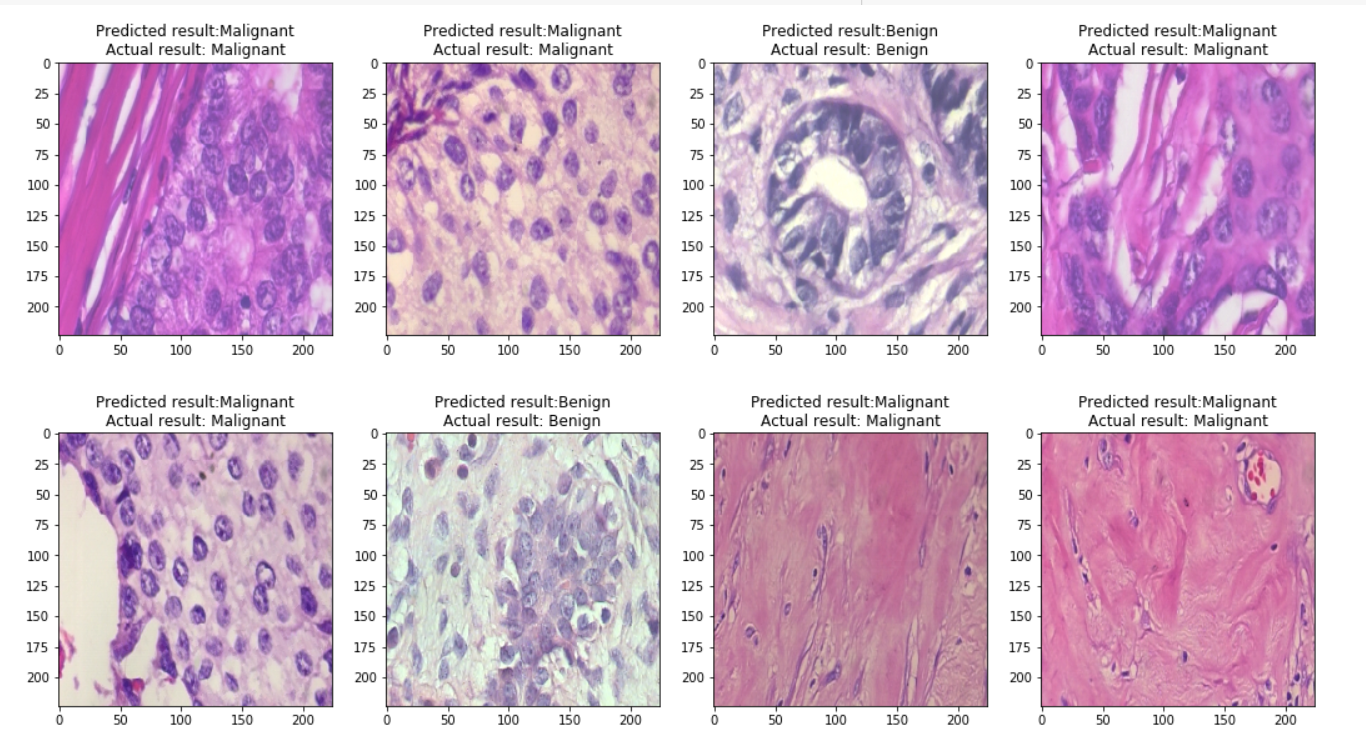
[[ 59 78]

[ 7 143]]



[ Fig –8: Confusion matrix for transfer learning (Resnet50) model]

Correct/incorrect Classification Samples using ResNet50 model:



[Fig-: Correct/incorrect classification samples]

7. Conclusion:

In summary, the project "Comparative study of Breast cancer detection using CNN and Transfer Learning" aims to advance a deeper learning-based breast cancer detection system. The study investigated different CNN models and Transfer Learning techniques for breast cancer diagnosis using medical imagery.

The objective of this study is to compare the efficacy of transfer learning and deep feature extraction techniques in the classification of breast cancer using histopathological images.The results of the study can provide valuable insights into the performance of deep learning approaches for breast cancer detection using medical images. The transfer learning model performs well as compared to CNN model. The findings can be useful for clinicians and researchers in developing more accurate and reliable breast cancer detection systems, ultimately leading to improved patient outcomes.

Moreover, the project's future scope is vast, with potential applications in various areas of healthcare, such as personalized medicine and integration into clinical practice. The successful application of deep learning approaches for breast cancer detection can lead to the development of more accurate and reliable diagnostic tools, ultimately improving patient care and outcomes.

Overall, the project highlights the importance of using advanced Deep learning techniques for detecting breast cancer accurately and reliably, ultimately contributing to the fight against this life-threatening disease.

# References

# 1.Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., & Budak, Ü. (2018). Transfer learning based histopathologic image classification for breast cancer detection. *Health information science and systems*, *6*, 1-7.

# 2. Guan, S., & Loew, M. (2017, October). Breast cancer detection using transfer learning in convolutional neural networks. In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* (pp. 1-8). IEEE.

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

# Hijab, A., Rushdi, M. A., Gomaa, M. M., & Eldeib, A. (2019, October). Breast cancer classification in ultrasound images using transfer learning. In *2019 Fifth international conference on advances in biomedical engineering (ICABME)* (pp. 1-4). IEEE.

# George, K., Faziludeen, S., & Sankaran, P. (2020). Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion. *Computers in Biology and Medicine*, *124*, 103954.

# Ahmad, N., Asghar, S., & Gillani, S. A. (2022). Transfer learning-assisted multi-resolution breast cancer histopathological images classification. *The Visual Computer*, *38*(8), 2751-2770.

# Wang, Y., Choi, E. J., Choi, Y., Zhang, H., Jin, G. Y., & Ko, S. B. (2020). Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound in medicine & biology*, *46*(5), 1119-1132.

# Altaf, M. M. (2021). A hybrid deep learning model for breast cancer diagnosis based on transfer learning and pulse-coupled neural networks. *Mathematical Biosciences and Engineering*, *18*(5), 5029-5046.

# Ayana, G., Dese, K., & Choe, S. W. (2021). Transfer learning in breast cancer diagnoses via ultrasound imaging. *Cancers*, *13*(4), 738.

# Cabıoğlu, Ç., & Oğul, H. (2020). Computer-aided breast cancer diagnosis from thermal images using transfer learning. In *Bioinformatics and Biomedical Engineering: 8th International Work-Conference, IWBBIO 2020, Granada, Spain, May 6–8, 2020, Proceedings 8* (pp. 716-726). Springer International Publishing.

# De Matos, J., Britto, A. D. S., Oliveira, L. E., & Koerich, A. L. (2019, July). Double transfer learning for breast cancer histopathologic image classification. In *2019 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

# Thuy, M. B. H., & Hoang, V. T. (2020). Fusing of deep learning, transfer learning and gan for breast cancer histopathological image classification. In *Advanced Computational Methods for Knowledge Engineering: Proceedings of the 6th International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2019 6* (pp. 255-266). Springer International Publishing.