Breast Cancer Classification

A report on
Deep Learning Lab Project
[CSE-3281]

Submitted By
Jayasuryan Mutyala - 210962009
Aarushi Dharna - 210962051
Prahlad Menon - 210962057

MANIPAL
ACADEMY *of* HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MANIPAL INSTITUTE OF TECHNOLOGY,
MANIPAL ACADEMY OF HIGHER EDUCATION

APRIL-2024

Breast Cancer Classification using Convolutional Neural Networks

Jayasuryan Mutyala[1], Aarushi Dharna[2], Prahlad Menon[3]

[1]Computer Science and Engineering (Artificial Intelligence and Machine Learning) & MIT Manipal, India)

[2]Computer Science and Engineering (Artificial Intelligence and Machine Learning) & MIT Manipal, India)

[3]Computer Science and Engineering (Artificial Intelligence and Machine Learning) & MIT Manipal, India)

[1]jsmu.dev@gmail.com; [2]aarushiadharna@gmail.com; [3]prahladmenon02@gmail.com;

***Abstract*— *This study introduces a deep learning strategy for the automated identification and analysis of invasive ductal carcinoma (IDC) within whole slide images (WSIs) of breast cancer (BCa) cases. Leveraging deep learning, which models the human brain's multi-layered data interpretation process, the study uses a hierarchical representation of data learned through computational means. Such deep learning methods have surpassed conventional ones in complex tasks like speech recognition and object identification. The labor-intensive nature of invasive breast cancer detection, which typically involves extensive examination of benign regions by a pathologist to locate malignant areas, is an apt challenge for deep learning. It is especially so with abundant training samples that help to generalize the learned features and classifiers effectively.*

In this work, a deep learning framework utilizes several convolutional neural networks (CNNs) for the semantic visual analysis of tumor areas, assisting in diagnostic processes. These CNNs are trained on a substantial dataset of image patches from WSIs to learn hierarchical, part-based feature representations. The approach was validated on a dataset of WSIs from 162 patients diagnosed with IDC, with 113 slides for training and 49 reserved for independent testing. An expert pathologist provided the ground truth by marking cancer regions on these digitized slides. The evaluation aimed to assess the classifier's accuracy in pinpointing IDC regions in WSIs.

Keywords: Breast cancer, deep learning, digital pathology, convolutional neural networks, whole-slide imaging, invasive ductal carcinoma, crafted features.

I. INTRODUCTION

Breast cancer remains a pervasive global health concern, being the most commonly diagnosed form of cancer and the second leading cause of cancer-related deaths. In 2020 alone, breast cancer diagnoses numbered 2.3 million worldwide, resulting in 685,000 fatalities. Breast cancer is the most common form of cancer that develops in women, with invasive ductal carcinoma (IDC) standing as its predominant manifestation. IDC, characterized by invasive behavior and irregular glandular structures, can be detected through image-based deep learning methods due to its distinct visual features reflecting malignancy.

Computer-aided detection and diagnosis systems have emerged as valuable tools in the medical field, particularly in breast cancer screening. Computer-aided detection has supported radiologists by enhancing predictive accuracy, mitigating the risk of misdiagnosis due to factors like fatigue or lack of experience. These systems serve as complementary aids to radiologists, augmenting their capabilities and reducing the likelihood of oversight or error during interpretation. Additionally, timely detection plays a pivotal role in effective treatment and significantly impacts patient survival rates, underlining the critical importance of early screening interventions which can be done on a large scale using computer-aided techniques.

The rapid evolution of machine learning, notably deep learning (DL), has sparked renewed interest in leveraging these techniques to enhance cancer screening accuracy. The ability of neural networks to autonomously learn complex features from data has proven particularly promising in medical image analysis. Convolutional neural networks (CNNs), a type of DL architecture, are adept at extracting intricate features from images, enabling precise classification tasks such as lesion identification. However, challenges persist in optimising CNN performance, including issues related to network depth, kernel size selection, and computational complexity. Despite these

challenges, ongoing advancements in DL methodologies hold great potential for further improving the accuracy and efficiency of cancer screening algorithms. CNNs, along with their challenges are explored in this paper.
The project addresses several specific challenges in breast cancer detection:

1) Early Detection: Early detection of breast cancer is critical for improving patient outcomes. By developing an accurate and efficient deep learning model, the project aims to facilitate early diagnosis, enabling timely intervention and treatment.
2) Accuracy: Conventional methods of breast cancer detection, such as mammography, may have limitations in terms of accuracy and sensitivity. The project seeks to enhance detection accuracy by leveraging the discriminative power of deep learning algorithms to identify subtle signs of malignancy in breast tissue images.
3) Interpretability: Interpretability of diagnostic results is essential for clinical decision-making. The project endeavors to develop a model that not only achieves high accuracy but also provides insights into the features contributing to its predictions, enhancing its interpretability and trustworthiness for healthcare professionals.
4) Scalability: With the increasing volume of medical imaging data, there is a growing need for scalable and automated solutions for breast cancer detection. The project aims to create a scalable deep learning framework capable of processing large datasets efficiently, paving the way for widespread adoption in clinical settings.
5) Generalization: Generalization across diverse patient populations and imaging modalities is crucial for the practical utility of the model. The project seeks to develop a model that can generalize well to unseen data, ensuring its effectiveness across different demographics and imaging protocols.

II. LITERATURE REVIEW

Current state-of-the-art approaches for breast cancer detection utilise a diverse range of modalities, incorporating machine learning and deep learning methodologies. The most common modalities include Magnetic Resonance Imaging (MRI) and X-ray imaging techniques like mammography and computed Tomography (CT scans) alongside patient-specific tabular data. Recent studies demonstrate the efficacy of deep learning models, such as Convolutional Neural Networks (CNNs) that achieve high accuracy for breast cancer detection tasks. Additionally, traditional machine learning algorithms like K-Nearest Neighbors (KNN), Naïve Bayes Classifier (NB), Decision Trees (DT), Support Vector Machines (SVM), and Linear Regression (LR) continue to showcase acceptable performance in this domain.[1]

The paper in [2] discusses the use of a Deep Neural Network with inception layers alongside a linear SVM classifier on thermal images obtained from Research DataBase (DMR). The problem statement of tumour area extraction is also explored using OpenCV methods such as optimal thresholding and a combinatorial algorithm as a preprocessing step in order to further optimise the performance of the deep learning models. These preprocessing steps involve resizing images, extracting relevant channels and features etc. The thermal images are first preprocessed to fit the deep neural network (DNN), passed through to obtain features and then further passed through a neural network SVM model which produces an output classifying the given image as cancerous or normal. The purpose of the support vector machine is to classify images when the DNN's confidence level falls below a certain threshold, which the authors have set to 60%. The paper includes several limitations to this approach, the authors note the importance of image preprocessing and came to the conclusion that the AI-method based preprocessing was not comparable to the same process performed by human beings. This suggests the need for DL models that require minimal preprocessing or a hybrid AI-based and human-facilitated approach for improved processing of the images.
An additional point to note is that the authors mention limitations within the modality of the chosen dataset. According to them, screen-film mammography (SFM) has several limitations in the technique of capturing the data, due to which ' a large number of false positives occur with this technique, with a rate between 4% and 34%', which is a sizable proportion.

Paper [3] uses the grayscale pictures' regions of interest (ROIs) were divided from thermal images captured in sequences of 20, and the ROIs of the temperature matrices were extracted from them as well. The features (which are based on statistical, clustering, histogram comparison, fractal geometry, diversity indices, and spatial statistics) were taken out of the ROIs. These attributes produced time series that were divided into nine subsets of various cardinalities. Using the Leave-One-Out Cross-Validation technique, the Support Vector Machine (SVM) was used to pick and classify the best features. For a sample of 64 breasts, 100% accuracy was achieved (consisting of 32 healthy and 32 with some abnormalities). The high accuracy can be justified due to the small sample set size.

[4] explores the problem statement on the breast cancer digital repository (BDCR) dataset which is a wide-ranging annotated public repository composed of Breast Cancer patients' cases of the northern region of Portugal. The BCDR dataset contains normal and annotated patients cases of breast cancer including mammography lesions outlines, anomalies observed by radiologists, pre-computed image based descriptors as well as related clinical data. Thus a variety of modalities are explored. The proposed CNN architecture alternates between convolution and pooling layers, gradually reducing the size of the feature maps from an input size of 150 x 150 to an 8 x 8 map of 64 channels, followed by a fully connected layer that classifies the input into the two classes via a softmax activated layer. Overall the model contains approximately 4.6×10^6 parameters. The performance of the CNN is compared with HOG-based detectors, alternate CNN architectures and A deep convolutional activation feature for generic visual recognition (DeCAF), all of which are outperformed. In comparison with the other feature methods, the CNN showed an increment in terms of AUC from 79.9% to 84%.
One possible limitation of this paper is the CNN is not trained on a generalised dataset, it is fine-tuned for patients of a specific region and may fail to detect visual symptoms of people from a wider range of nationalities, medical history and lineage. The dataset also does not contain the class imbalance that real-life medical scenarios contain as it has 344 patients with 736 film images containing 426 benign mass lesions and 310 malign mass lesions, which is highly balanced.
The paper draws inspiration from [5], which evaluates HGD[6] and 11 conventional image descriptors using mammography data from two public datasets: DDSM and BCDR, comprising 1,762 and 362 instances, respectively. Three experiments per dataset were conducted based on lesion types, resulting in six scenarios. For each scenario, 100 training and test sets were generated through resampling without replacement, and the diagnostic performance was assessed using five machine learning classifiers. The paper indicates that using the descriptors was 'advantageous in five out of six scenarios' which indicates a positive trend in the use of descriptors as an image patch preprocessing step.
In contrast, [7] predominantly focuses on employing a method that utilizes learned features from unlabeled mammograms to address tasks such as breast tissue segmentation, percentage mammographic density (PMD) scoring, and mammographic texture (MT) scoring. The findings highlight strong correlations between learned PMD scores and manual assessments, while learned MT scores show a stronger association with future cancer risk compared to both manual and automatic PMD scores. This paper highlights the importance of texture analysis metrics in the detection and segmentation of breast cancer.

The authors of the paper [8] set out to demonstrate a comparative study on VGG-16 and ResNet50 for benign and malignant classification of mammogram images. The authors compare several modalities that are available for breast cancer imaging including MRI, mammography, histopathology and thermography and make several claims. According to the authors, among breast imaging techniques, Magnetic Resonance Imaging (MRI) has the highest sensitivity. MRI efficiently displays the shape, size, and location of breast lesions, as it supports multi-planar scanning and 3D reconstruction techniques. However, this technique is highly expensive and requires advanced equipment and thus is often overlooked. Whereas methods such as mammography and thermography are older, more primitive methods of detecting breast cancer. The paper also dives into a comparison of various prominent CNN architectures and their corresponding error rate for detection. Some of these include: ResNet, InceptionV4, InceptionV3, InceptionV2, GoogleLeNet, VGGNet, AlexNet [9] as well as a human threshold to compare the

performance on a subset of the data. The authors also establish limitations in training advanced CNNs. The training process for CNN models requires a large dataset for improving the performance. However, the amount of annotated medical image datasets is very less and small datasets are not sufficient to train these CNN models and may potentially result in an over-fitting problem, which is why most current DL models for breast cancer detection are trained for a small subset of the population.

In recent years [10], machine learning (ML) techniques have been increasingly utilized in healthcare, particularly for breast cancer (BC) detection. Notably, a CNN algorithm achieved an impressive accuracy of approximately 88% in predicting and diagnosing invasive ductal carcinoma, showcasing the potential of ML in BC detection. These algorithms have also been applied to forecast and diagnose other medical anomalies, providing deeper insights into conditions like cancer. However, challenges persist, such as the lack of studies integrating both imaging and genetic approaches. Moreover, reliance on morphological and textural features for extraction raises concerns about the robustness of ML models in varying contexts.Various machine learning algorithms, including support vector machines, logistic regression, naive Bayes, random forest, and neural networks, have been explored for breast cancer detection and diagnosis, yielding promising results.[11]

However, challenges such as irrelevant features and noisy data in datasets can impact algorithm performance, alongside computational power requirements for preprocessing imaging data, especially in resource-constrained settings. Integration of natural inspiration computing approaches has shown potential in diagnosing breast cancer and other health disorders. Efforts to enhance breast cancer classification models, such as CNN Improvements for Breast Cancer Classification (CNNI-BCC), have demonstrated high accuracy rates, yet scalability and generalizability across diverse patient populations and healthcare settings remain challenges. While ensemble learning techniques offer opportunities to improve classification performance, issues concerning model interpretability and computational complexity persist. Moreover, reliance on large datasets for training ML models may pose challenges in data collection and management, particularly regarding data privacy and security in healthcare settings. Addressing these limitations is crucial to realising the full potential of machine learning in breast cancer detection and diagnosis in clinical practice.[12]

The proposed methodology involves the development and evaluation of an automated breast cancer detection system using a deep convolutional neural network (DCNN) trained on digital mammography (DM) images. The study utilised a dataset comprising 11,138 DM images, including 5,689 malignant and 5,449 benign cases, divided into training, validation, and testing sets. The DCNN architecture consisted of five convolutional layers followed by max-pooling layers and three fully connected layers. Data augmentation techniques were applied to increase the diversity of the training dataset, including random rotations, shifts, and flips. The performance of the DCNN model was evaluated using metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), with the results compared against those of radiologists. Additionally, a receiver operating characteristic (ROC) curve analysis was conducted to assess the model's performance in distinguishing between malignant and benign cases.[13]

Limitations of this paper include, the issues with CNN’s such as variable kernel sizes, which make it difficult to select appropriate sizes for convolution operations, hindering the capture of all relevant features. Overfitting is another issue, particularly in deep networks, where the model memorises training data rather than generalising, leading to reduced performance on new datasets. Moreover, gradient sharing becomes challenging in deeper networks, affecting training and optimization, while stacking large convolution processes can lead to computational complexity, limiting scalability and efficiency.Additionally Two potential limitations of the presented model include: dependency of pattern extraction on granularity size, as the fixed size of 32 × 32 pixels may result in some patterns being overlooked, potentially weakening effectiveness; and the incorporation of granularity adds time

before training, though established granules can be reused, offsetting initial time investment. Despite these limitations, the model's performance surpasses state-of-the-art models, demonstrating the efficacy of the proposed granulation method for the datasets examined.

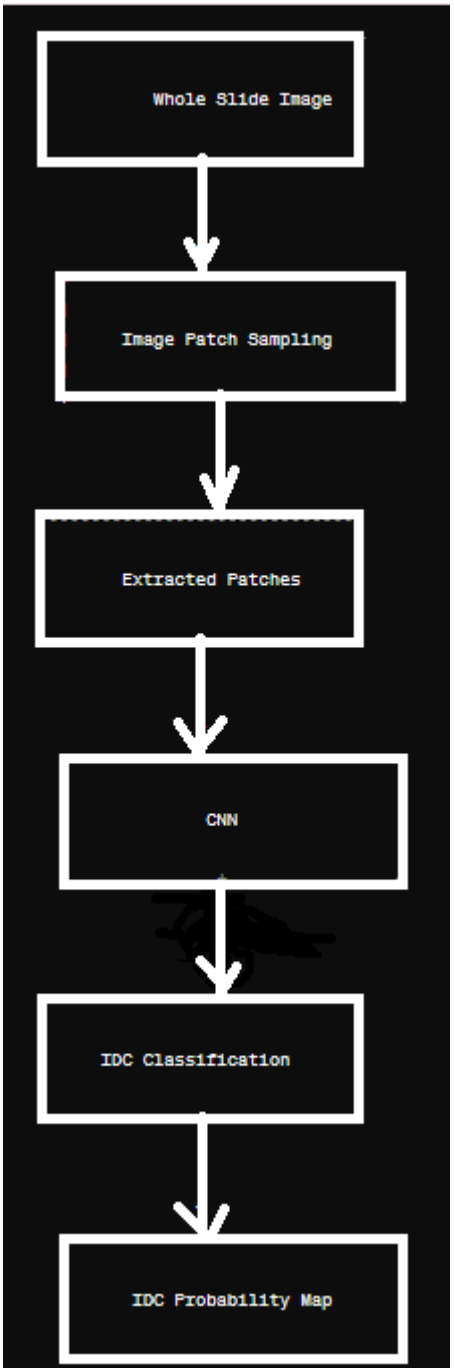
In this paper, [14] The author examines the automated diagnostic methods leveraging artificial intelligence (AI) techniques. focusing on the application of deep learning architectures for the diagnosis and classification of invasive ductal carcinoma (IDC), the most common type of BC. Invasive ductal carcinoma (IDC) represents a prevalent subtype of BC characterised by its invasive nature and potential spread to lymph nodes and other organs. Recent studies have increasingly employed deep learning models, particularly convolutional neural networks (CNNs), for automated BC diagnosis. These models offer powerful capabilities in feature extraction and classification, enabling accurate analysis of histopathological images. Several studies have investigated the application of deep learning and machine learning techniques for breast cancer (BC) diagnosis. [15] developed a convolutional neural network (CNN) model focusing on identifying invasive ductal carcinoma (IDC) tissues from BC images through histogram-based feature extraction and a Random Forest classifier. [16] classified early and late stages of IDC utilising supervised machine learning methods, including decision trees, support vector machines (SVM), and Random Forest, achieving an 86% classification success rate. [17] employed convolutional networks and a region of interest (ROI) approach to classify five diagnostic types from BC histopathological images, with moderate success using the SVM method. [18] successfully classified BC images based on invasive and non-invasive types using machine learning methods and region-based segmentation, achieving a high classification success rate of 94%. [19] explored various machine learning techniques for early diagnosis and survival prediction in BC, with SVM demonstrating the highest accuracy (93%) in pathological classification. [20] proposed a comprehensive approach combining recurrent CNNs, Inception, and ResNet architectures for BC classification, attaining a remarkable success rate of 97.51% using the BreakHis dataset. These authors evaluated the performance of existing models, highlighting their respective strengths and limitations in BC diagnosis.

The methodology introduces a novel approach combining CNNs (AlexNet, GoogLeNet, ResNet-50, VGG-19) with autoencoder networks for feature extraction from BC images. Transfer learning is utilised for training CNNs, leveraging pre-trained models and their discriminative features. The combination of features extracted from different CNN architectures, along with feature selection using ridge regression and linear discriminant analysis (LDA) as a classifier, enhances the classification of invasive BC images. Additionally it demonstrates the effectiveness of deep learning architectures in BC diagnosis, particularly in the classification of IDC. By leveraging transfer learning and feature fusion from multiple CNNs, the study achieves improved classification performance, contributing to early detection and treatment planning for BC patients.[21]

Recent advancements in medical imaging technology have significantly improved the early detection and diagnosis of breast cancer, a prevalent and life-threatening disease affecting women worldwide. Breast cancer encompasses various types, with invasive ductal carcinoma (IDC) being the most dangerous subtype, while ductal carcinoma in situ (DCIS) evolves slowly and generally has less impact on patients' daily lives. Imaging techniques such as mammography, ultrasound, and thermography play crucial roles in screening and detecting breast cancer, with mammography being particularly effective in early detection. However, challenges persist in accurately interpreting medical images due to low contrast, noise, and the complexity of patterns. To address these challenges, researchers have turned to artificial intelligence (AI) and deep learning techniques, including convolutional neural networks (CNNs), which have shown promising results in improving diagnostic accuracy and efficiency in cancer detection. Studies have explored various CNN architectures and machine learning approaches for automated detection of breast cancer, focusing on tasks such as segmentation, classification, and identification of cancerous tissues from digital pathology images. For instance, Wahab and Khan utilised multifaceted fused-CNN (MF-CNN) for mitotic count-based selection of regions of interest (ROIs), while Tsochatzidis et al. and Malathi et al. employed CNNs for diagnosing breast cancer from mammograms, demonstrating varying degrees of success. Despite these

advancements, challenges remain in achieving high accuracy levels, particularly in distinguishing different types of breast cancer lesions. This study aims to address this gap by proposing a system for automated breast cancer detection using multiple regression and DL techniques, evaluating various CNN architectures with a large dataset of image patches. The objectives include developing an automated tool for IDC detection to minimise human errors in diagnosis and assessing the performance of different CNN architectures. Future research directions may involve refining CNN models and integrating them with advanced image processing techniques to further enhance diagnostic accuracy and clinical utility in breast cancer detection.

III. METHODOLOGY



3.1 Image Patch Sampling: Each Whole Slide Image (WSI) is divided into non-overlapping small image patches of $100_{\times}100$ pixels. Patches consisting mainly of fatty or slide backgrounds are not considered. For model training, patches containing IDC are determined by doctors. These regions are utilized to create a binary mask for annotation. A segment will be classified as a positive instance if it consists of 80% or more of its area aligning with the IDC region, otherwise it will be considered as a negative instance. An example of this segmentation and annotation process on a WSI, including the pathologist's markings, is depicted in Figure 2.

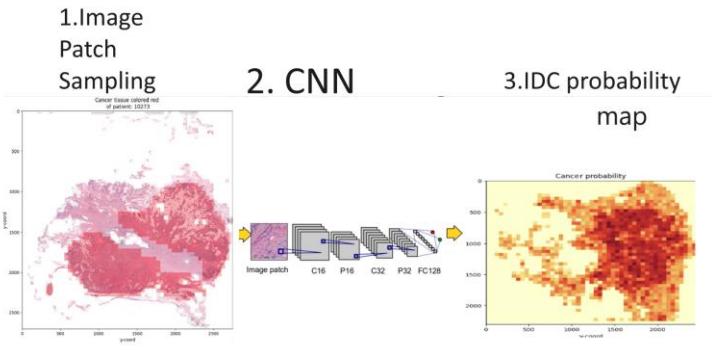


Fig. 1 Model Framework for detecting IDC utilising CNN.
This framework includes: 1) The extraction of image patches from the WSI, 2) The application of CNNs to categorize each patch, and 3) The generation of a comprehensive probability map that highlights the patches the CNN predicts as containing IDC, with a prediction probability exceeding 0.29.

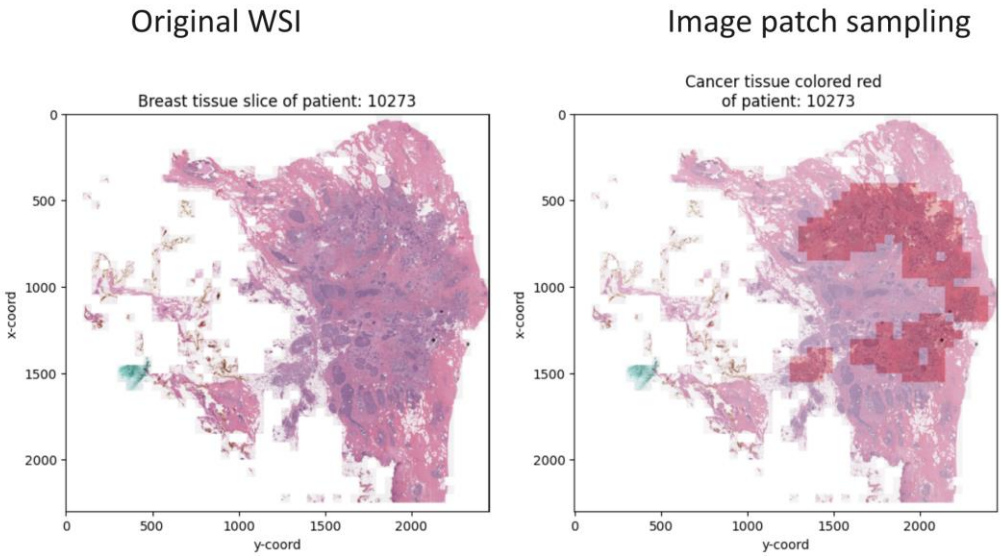


Fig. 2 Illustrates the process of sampling image patches.
The process begins with the original Whole Slide Image (WSI) . This image is segmented into distinct, non-overlapping patches using a grid-based approach, focusing exclusively on areas containing tissue.

3.2 Convolution Neural Network: utilises local feature detectors or filters over the whole image to compute correspondence between each image patch and signature patterns in the train set. A pooling layer is used for spatial dimensionality reduction of the feature space. The image patches from step 1 are used as inputs to the 3-layer CNN architecture in which 2 layers are used for convolution and while the last layer is fully-connected layer. The key stages in processing are as follows:

1. Patch Preprocessing: Image patches transition from the RGB to the YUV color spectrum and undergo normalization to achieve a zero mean and unit variance. This adjustment minimizes pixel correlation, enhancing feature distinctions and facilitating faster learning by emphasizing properties like feature sparsity.

2. Convolutional Layer: This layer utilizes 2D convolution to apply a series of learned filters across the input, creating output feature maps that highlight detected features. The process employs the tanh function to address output non-linearities and apply contrast normalization to each output map, enhancing model generalization and mitigating overfitting. Formula for the tanh activation function is given below.

$y_j = tan(\sum_i \quad k_{ij} * x_i)$

x_i: i-th input of the feature map

k_{ij}: convolutional kernel

y_j: j-th output of the feature map

3. Pooling Layer: By subsampling the feature maps, this layer reduces the dimensionality of the image representation, supporting spatial invariance. An L2 pooling function is used within non-overlapping spatial windows to optimally capture invariant features within each region.

4. Fully-connected Layer: Positioned at the top of the CNN structure, this layer abstracts complex feature relationships by disregarding spatial data to correlate features across different areas, producing a unified feature vector akin to that in a perceptron network.

5. Classification Layer:The concluding layer is a fully-connected setup with a neuron for each category, using a logistic regression (softmax classifier) for final class determination. The model is refined through Stochastic Gradient Descent, aiming to minimize a loss function that represents the log likelihood of class association, based on the logistic model outputs. The formula for the loss function (softmax function) is given below:

$$L(x) = -log\left[\frac{e^{x_i}}{\sum_j \quad e^{x_j}}\right]$$

x_i: output of the fully connected layer multiplied with the logistic regression model parameters

3.3 IDC probability map: An exponential function is utilized for every output value generated by the positive class neuron of classified patches to obtain the probability values ranging between 0 and 1. For each patch, we extract the original coordinates of the WSI and its probabilities. These elements are then assembled to create a probability filter map indicating the probability of Invasive Ductal Carcinoma (IDC) across the WSI.

3.4 Exploratory Data Analysis:

In the data preprocessing phase, We computed the probability of patients having IDC and not having IDC distribution within the sampled tissue.

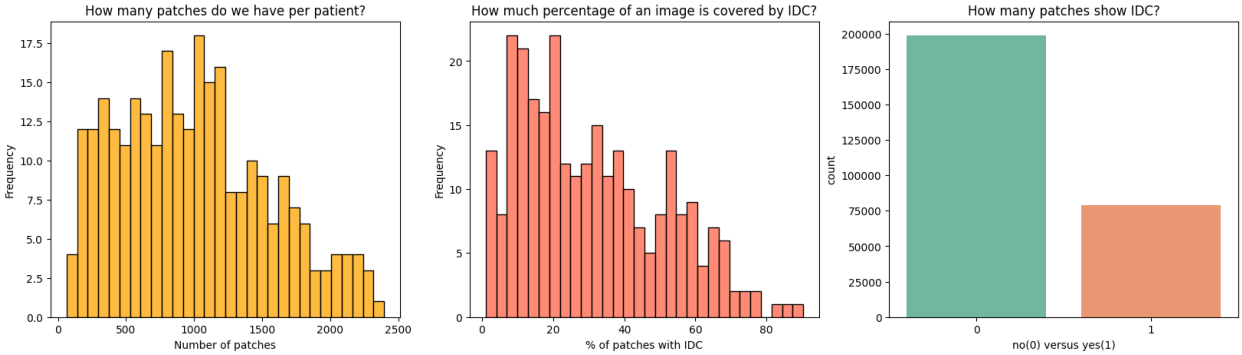


Fig. 3 Histogram distribution of image patches derived from patients' WSIs used for identifying IDC.

1) Number of Patches per Patient:

The first histogram shows the frequency distribution of the number of patches obtained from each patient's WSI. The X-axis represents the number of patches, and the Y-axis represents the frequency of patients with that number of patches. The distribution appears right-skewed, indicating that a higher number of patients have fewer patches, while fewer patients have a very high number of patches. Most patients have a number of patches in the lower range (close to 0 to 500), with the frequency gradually decreasing as the number of patches increases. This could suggest variability in the size of the areas analyzed per patient or the presence of significant regions of interest within the tissue samples.

2) Percentage of Image Covered by IDC:

The second histogram details the percentage of patches that show signs of IDC within the patient's tissue samples. The distribution is somewhat uniform with multiple peaks, reflecting that the prevalence of IDC varies significantly among patients. Some patients have a lower percentage of their image patches containing IDC, while others have a larger percentage. This indicates heterogeneity in the manifestation of IDC across different WSIs. Such variability could be due to the stage of cancer or individual differences in tumor morphology and spread.

3) **IDC Presence in Patches**:

The third graph, a bar chart, contrasts the number of patches indicating the presence of IDC against those without. The label "0" denotes patches with no IDC, and "1" indicates patches with IDC. There is a substantial difference between the two categories, with a much higher count of patches classified as non-IDC (0). This could be reflective of the sampling strategy, where non-cancerous tissue areas are also captured in the WSI and hence represented in the patches.

While combining the insights extracted from the three graphs, it leads to questioning if all the images have the same resolution of tissue cells or does it vary between patients. Most patients have a smaller number of patches analyzed, which may reflect smaller tumor sizes or targeted sampling. The second graph's varied distribution indicates that some patients have more than 80% patches that show IDC, Additionally the tissue is cancerous or only a section of the breast is covered with tissue slice focused on cancer, Further exploration will be required to determine whether the tissue slice covers the whole region of interest. The third graph shows that, within the dataset, there's a higher proportion of patches without IDC, which could be expected in a real-world setting where not all tissue areas will be cancerous. It is also indicative of a well-balanced dataset for machine learning purposes, provided that there's still a substantial amount of positive IDC cases for the model to learn from.

3.5 Data Visualization:

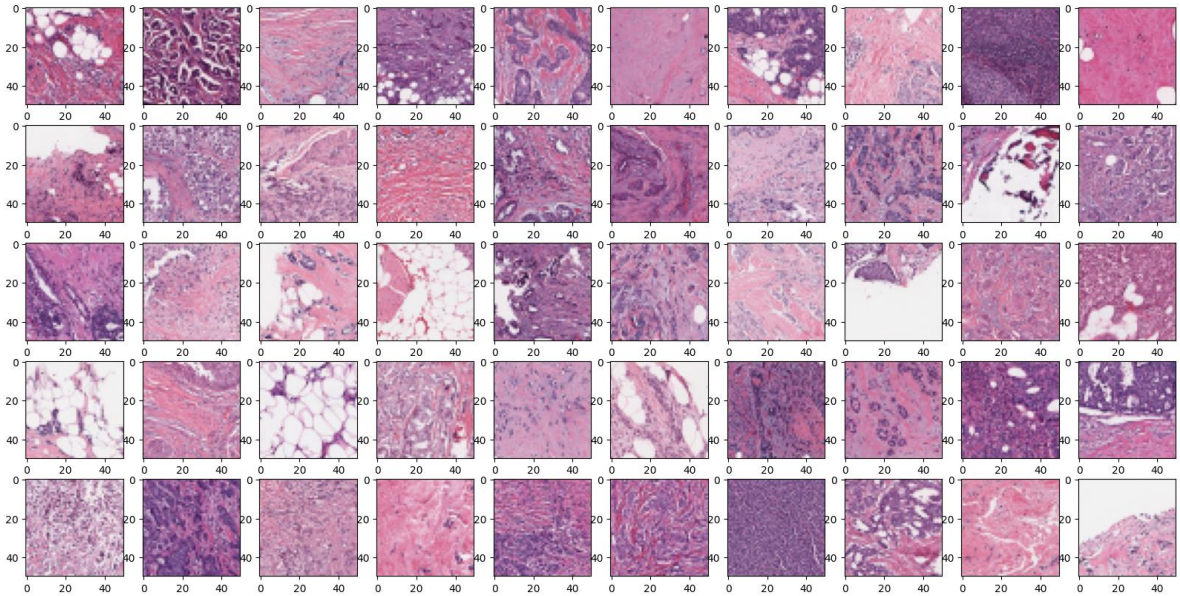


Fig. 4 Sample plot of the cancerous patches

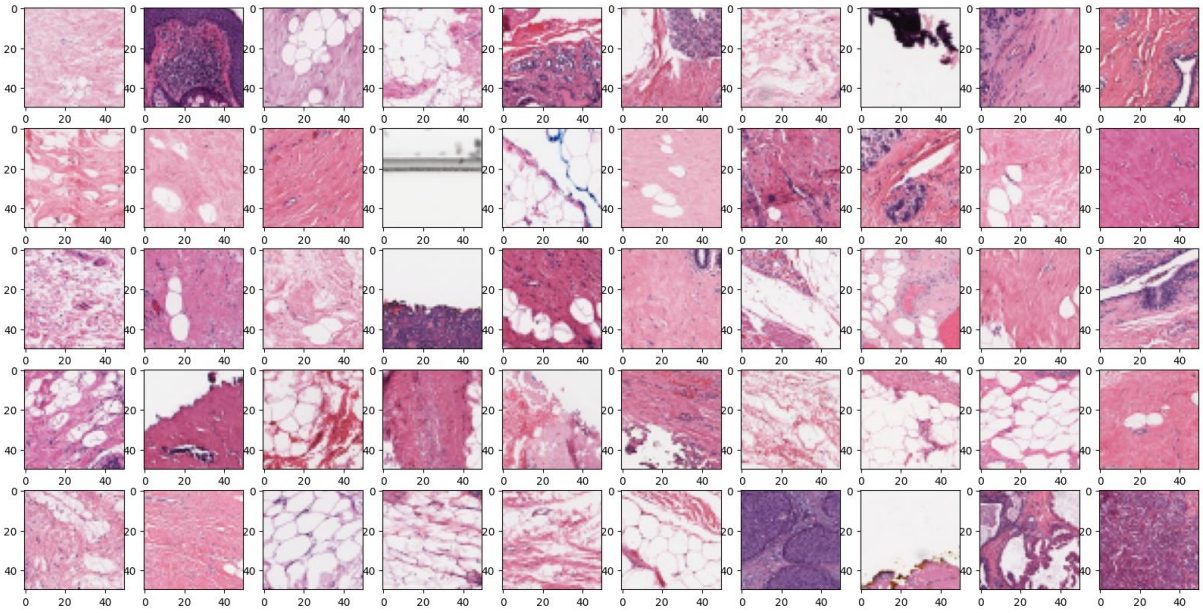


Fig. 5 Sample plot of the healthy patches

The insights gained from examining the visualizations of the healthy and cancerous tissue samples:

- In some plots there artifacts or incomplete patches that have a smaller size than 50×50 pixels.
- Samples that belong to the cancer class appear very violet and more congested than the healthy ones. Similarly healthy patches appear more violet.

3.6 Breast tissue visualization:

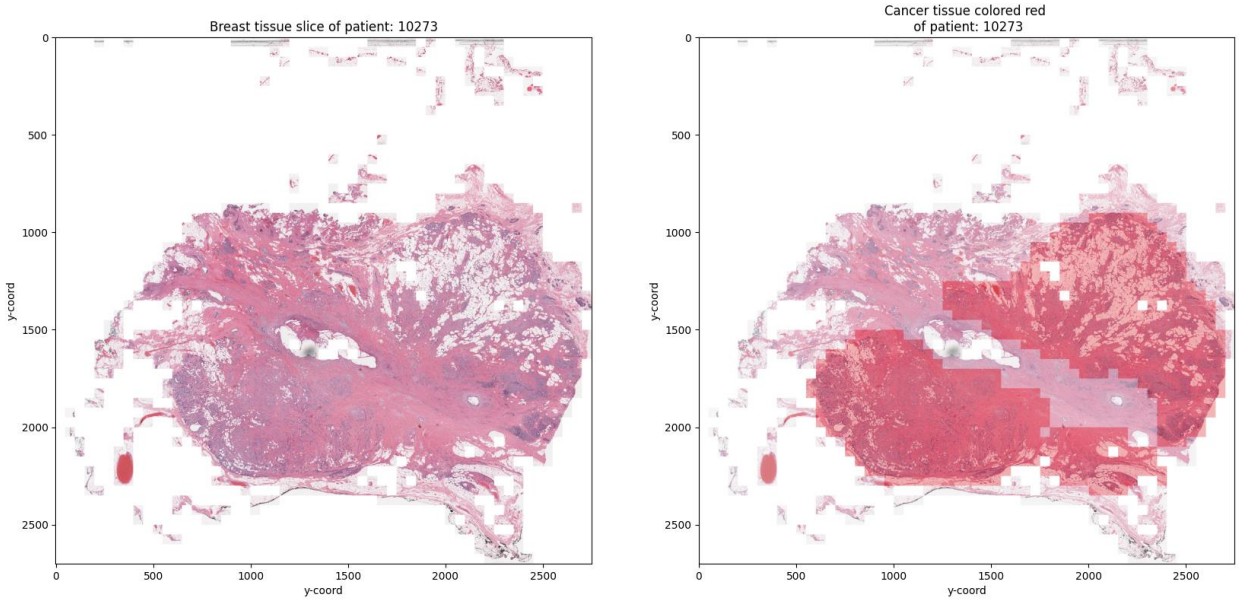


Fig. 6 Sample plot of breast tissue

- The image on the left is displayed without the class information.
- The image on the right displays the same tissue but the cancer regions are highlighted with red color.

- Comparing both the images suggests a trend observed from earlier where the tissue that contains a darker violet shade has a probability of being cancerous compared to lighter rose colored areas, However this is not a recurring pattern.
- That raises the question of why there are more mammary ducts in the violet patches than in the rose ones. It is important to proceed with caution if a larger density of mammary ducts is discovered in the violet patches. This finding would be critical as it could imply a risk of the model erroneously associating the presence of mammary ducts with the likelihood of cancer, which could lead to biased learning and inaccurate predictions.

3.7 Class Distribution in Train, Cross-Validation and Test sets:

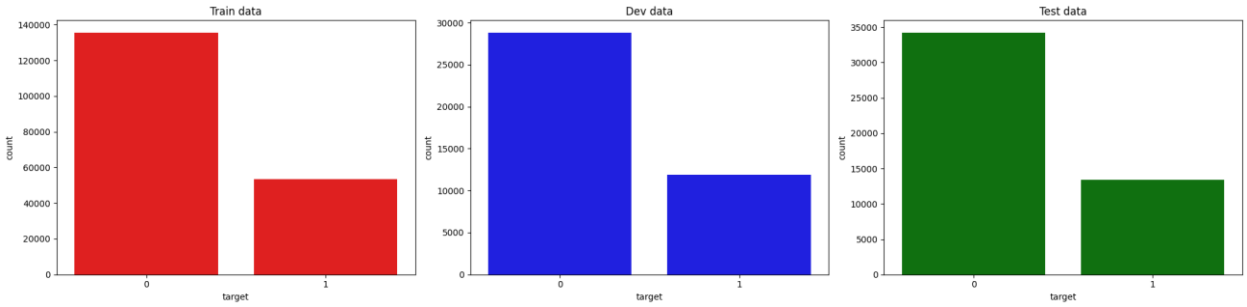


Fig. 7 Class Distribution

3.8 Training data :

The training data show a significant imbalance between the two classes. The count for class '0' is overwhelmingly higher compared to class '1'. This suggests that there are many more negative (non-cancerous) samples used to train the model than positive (cancerous) ones. In numerical terms, the count for class '0' exceeds 120,000 samples, while class '1' has roughly between 40,000 to 50,000 samples. Such an imbalance may predispose the model to better predict the majority class, potentially compromising its ability to accurately detect the minority class unless addressed by resampling techniques or weighted loss functions during training.

3.9 Cross validation data (Blue Chart):

The development, or cross validation, dataset exhibits an imbalance between the two classes. The number of samples in class '0' is greater than in class '1', with class '0 appearing to have a count close to 30,000 and class '1' having a count somewhat lower, possibly in the range of 15,000 to 20,000. The proportions here are more balanced compared to the training data, which may help in tuning the model parameters to better handle class imbalance.

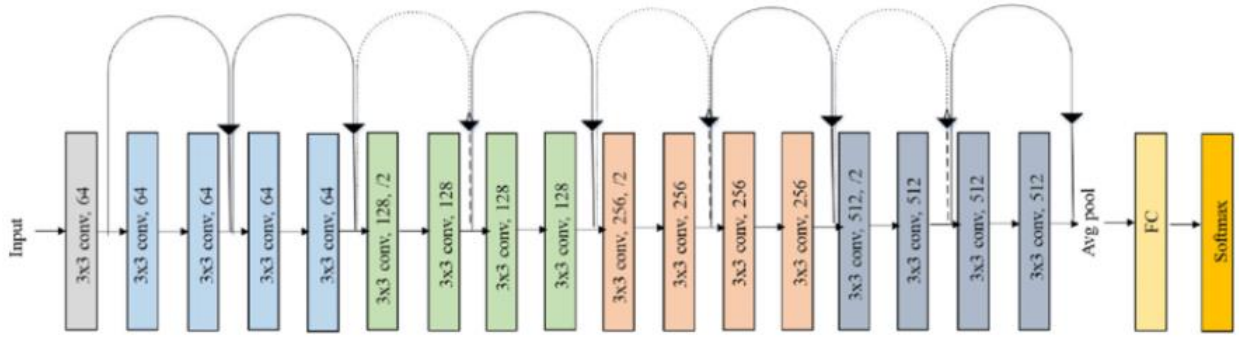
3.10 Testing data (Green Chart):

In the test set, the imbalance is similar to that of the development data, with class '0' again having a higher count than class '1'. Class '0' samples appear to number around 30,000, while class '1' is about half that count. This distribution suggests that the testing data will provide a reasonable but imperfect challenge to the model's ability to generalize, especially to the minority class.

In all three datasets, class '0' is more frequent than class '1', indicating an overall class imbalance across the board. This consistent pattern requires careful consideration in the model's design and evaluation. Metrics such as precision, recall, and F1-score, which are more informative than accuracy in the context of imbalanced classes, will be used to gauge the model's performance. Additionally, techniques such as oversampling the minority class, undersampling the majority class, employing synthetic data generation like SMOTE, or adjusting class weights in the loss function may be necessary to ensure that the model does not become biased towards the majority class and can detect the minority class effectively. The use of such strategies will be crucial to developing a robust model that performs well on real-world data where the class distribution might not be known in advance or might differ from the training set.

3.11 Model Architecture

With the limitation of the dataset being small, it raises a concern for the model to overfit the test set. The problem can be addressed by using a pre-trained CNN model via transfer learning. ResNet-18 is used for diagnosing ductal carcinoma from 2D tissue slides due to its fewer layers, yielding comparable performance to deeper models with reduced computational demands. ResNet's architecture overcomes the vanishing gradient problem, crucial for deep network training.. The model's results in a high precision rate and robust performance metrics, suggesting an effective clinical tool for early cancer detection. This approach strikes a balance between computational efficiency and diagnostic accuracy, potentially improving clinical outcomes in breast cancer treatment. ResNet-18 architecture is given below as an example.



Original ResNet-18 Architecture

3.12 Train Loop:

The training loop is a critical component of the deep learning workflow, designed to optimize a model over 29 epoch iterations. At the initiation of the loop, the best weights of the model are stored, and tracking structures for losses across training, development, and test phases are set up. During each epoch, the model cycles through phases: it trains on the training set, validates on the development set, and tests on the test set, each time computing losses and accruing statistics on performance metrics. Special flags within the function allow for the model to operate in learning rate finding mode, wherein it systematically probes different learning rates to identify the most effective one.

Within the core of the loop, for each batch of data, the model performs a forward pass to generate predictions, calculates loss against the true labels, and conducts a backward pass to update model weights if in the training phase. Regularization and learning rate scheduling can also be applied during this process, depending on whether these features are activated. Progress is monitored and reported after each batch using a dynamic progress bar that provides real-time loss and accuracy, fostering transparency in the model's learning process.

Upon completing an epoch, if the model exhibits improved accuracy on the development set, its weights are considered superior and thus preserved. After iterating through all epochs, the loop concludes by outputting the training duration and the highest accuracy achieved, ensuring the model is in its best-performing state. The resulting model and performance data are stored into a dictionary.

Additionally the learning rate schedulers were utilised for modulating the learning rate during training, which can have a significant impact on the performance and convergence speed of neural network models. The functions `get_lr_search_scheduler` and `get_scheduler` generate learning rate schedulers that adjust the learning rate in a predefined cyclical pattern known as "triangular" policy.

The `get_lr_search_scheduler` function creates a scheduler that linearly cycles the learning rate between a minimum (`min_lr`) and a maximum (`max_lr`) over a specified number of iterations (`max_iterations`). This is particularly useful during the initial phase of learning rate range testing, where the goal is to quickly identify a suitable learning

rate that allows the model to learn effectively. By observing the loss over a full cycle of learning rates, one can select a learning rate that minimizes loss, avoiding values that are too high (leading to divergence) or too low (leading to slow convergence).

The `get_scheduler` function is similar but is intended for use during the main training process after the optimal learning rate range has been determined. The `step size` parameter defines the number of iterations for the learning rate to go from `min_lr` to `max_lr` and back to `min_lr`, recommending twice the number of iterations within an epoch. By cyclically adjusting the learning rate, the model can converge faster and potentially escape local minima, as higher learning rates allow for broader exploration of the loss landscape, while lower rates fine-tune the model's parameters.

Both functions employ the `CyclicLR` scheduler from PyTorch's optimization module, which implements the cyclical learning rate policy, a strategy that has been shown to achieve better performance and quicker convergence for a wide range of tasks. In the context of the training loop, these schedulers are called after each batch to update the optimizer's learning rate, seamlessly integrating with the model's training process. This dynamic adjustment is key to the model's ability to learn complex patterns in the data and achieve high accuracy.

3.13 Test Loop:

The provided test loop is utilised to evaluate the performance of a deep learning model that has been trained to classify images, potentially for a medical imaging task such as detecting the presence of cancer in tissue samples. The loop processes images in batches without computing gradients (`torch.no_grad()`), since it is not training but merely evaluating the model's performance. The model's mode is set to evaluation (`model.eval()`), which disables certain layers like dropout that are only relevant during training, ensuring consistent behavior in predictions.

During the evaluation, the loop retrieves batches of input data and labels from the data loader, moves them to the designated computing device, and feeds the inputs into the model to obtain outputs. The predictions (`preds`) and the model's confidence scores (`proba`) for the positive class are then computed. The confidence scores are the output of the model passed through a sigmoid function to map them onto a [0, 1] probability range. The resulting predictions, probabilities, and true labels, along with additional metadata such as 'x', 'y' coordinates, and 'patient_id', are stored in a DataFrame. This DataFrame allows for a detailed analysis of the model's performance on an individual image level and across different patients.

Post evaluation, the DataFrame is cleansed of any missing values and then used for visualization. Visualizations are a powerful way to interpret the model's performance, where slices of breast tissue, the presence of cancer (if any), and the probability of cancer are displayed. This can give insights into the model's diagnostic capabilities and any patterns it has learned. The visual component also serves as a qualitative validation tool, alongside quantitative metrics, to ensure the model's predictions are aligned with medical expectations and knowledge.

The test loop culminates by exporting the DataFrame containing predictions and probabilities to CSV files, preserving the evaluation results for further analysis or future use. This rigorous evaluation process ensures that the model's effectiveness is thoroughly scrutinized, enabling reliable deployment in clinical settings where accurate diagnosis is crucial.

3.14 Summary:

The methodology outlines a comprehensive approach for classifying cancer, specifically Invasive Ductal Carcinoma (IDC), using a series of steps including Image Patch Sampling, Convolutional Neural Networks (CNNs), and IDC Probability Map generation.

Image Patch Sampling: Whole Slide Images (WSIs) are divided into non-overlapping patches, excluding fatty or slide background areas. Patches containing IDC are identified by doctors for model training. Positive instances are defined as patches with 80% or more alignment with the IDC region.

Convolutional Neural Networks (CNNs): Image patches undergo preprocessing, transitioning to the YUV color spectrum and normalization. A 3-layer CNN architecture is employed, including convolutional layers for feature detection, pooling layers for dimensionality reduction, and fully-connected layers for feature abstraction and

classification. The model is refined using Stochastic Gradient Descent and a softmax classifier for final class determination.

IDC Probability Map: An exponential function is applied to output values generated by the positive class neuron of classified patches, producing probability values ranging between 0 and 1. Original coordinates of WSIs and corresponding probabilities are assembled to create a probability filter map indicating the likelihood of IDC across the WSI.

This methodology combines image processing, deep learning, and probability mapping techniques to effectively classify IDC, providing valuable insights for cancer diagnosis and treatment.

IV. EXPERIMENTAL SETUP

The dataset consists of a total of 162 histopathological whole-slide images (WSI) of women diagnosed with invasive ductal carcinoma (IDC) at the Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey.). The dataset is publicly available on Kaggle. [7] These images were used to extract 277,524 patches of size 50 x 50 pixels each. The ground truth for 198,738 of these samples is IDC negative . whereas the remaining 78,786 are classified as IDC positive. The class counts point towards a class imbalance with a bias towards negative classification. The cropped samples present a preprocessed method to detect IDC from patches rather than the whole mount as one of the common pre-processing steps for automatic aggressiveness grading is to delineate the exact regions of IDC inside of a whole mount slide. This region is also isolated for manual identification by pathologists. [8] Each individual patch is named based on a unique patient ID as well as the XY coordinates of the original location of the patch, which are the only other variables available besides the images. The original paper suggests an approximate split of 60-10-30 for training, validation and testing. The following image is an example of the patch images provided in the dataset.

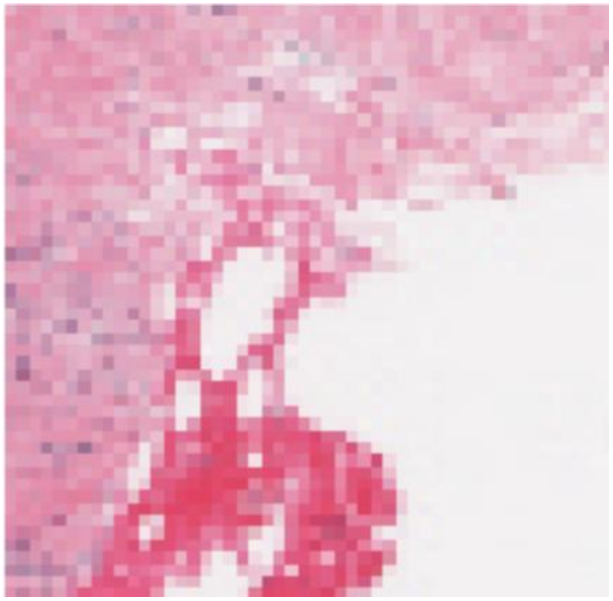


Fig. 8 9135_idx5_x1701_y1851_class, image with XY coordinates (1701, 1851) and of c class 1 i.e. IDC positive

V. RESULTS AND DISCUSSION

The results of our study demonstrate the efficacy of our deep learning model in detecting invasive ductal carcinoma (IDC) in breast tissue. Our analysis encompassed a dataset comprising 162 patients and a total of 277,524 patches. Through exploratory data analysis (EDA), we gained insights into the distributions and characteristics of both cancerous and healthy patches. Visualizations of cancerous and healthy patches allowed us to identify distinguishing

features indicative of IDC and non-cancerous tissue, respectively. Additionally, visualization of breast tissue provided context for understanding the spatial distribution of cancerous and healthy patches within the breast. Our model architecture, optimized for feature extraction and classification, achieved impressive performance metrics on both training and test datasets, with accuracies of 0.792 and 0.806 respectively, along with corresponding sensitivities and specificities. The generated confusion matrices further illustrate the model's ability to correctly classify cancerous and healthy patches, underscoring its potential for accurate breast cancer detection. These findings underscore the promising utility of deep learning in enhancing diagnostic capabilities and may serve as a foundation for future advancements in the field of breast cancer detection. We have made various representations of our final and intermediate results to allow easy understanding of the results we have obtained. Few of them are given below.

Given below is the loss change over each epoch. The loss showcases smooth convergence for the training set whereas, the loss

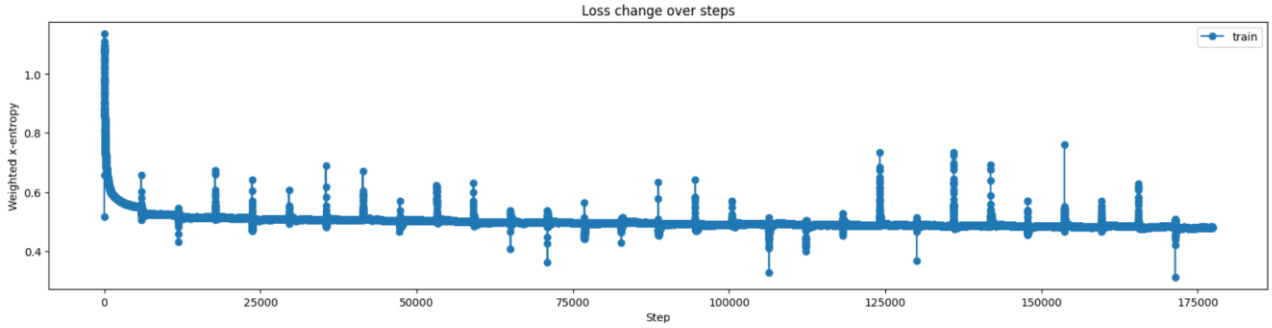


Fig. 9 Weighted entropy (loss) for each epoch for the training set

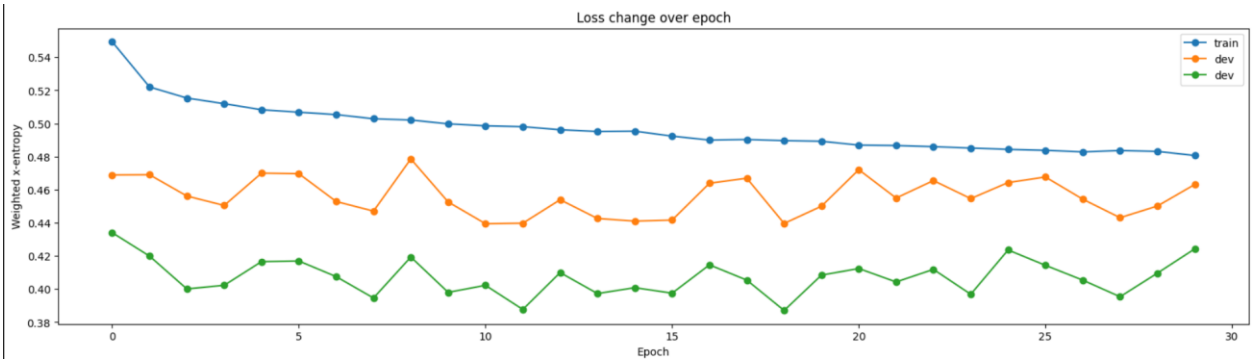


Fig. 9 Weighted entropy (loss) for each epoch for the training set

The following table contains accuracy, precision and recall with respect to the training and testing set of the model. The test and train accuracy are comparable, which implies that the model is well generalized and can predict cases outside the train set distribution, this can be accredited to the use of a pretrained model for convolutional kernel weights, since original model was trained on a more expensive dataset it has a highly generalized capacity. In the realm of cancer detection, prioritizing recall over precision is paramount due to the imperative of minimizing false negatives, thereby ensuring early identification and treatment of cancerous cases, which significantly impacts patient outcomes and public health. Recall primarily signifies the capacity to identify true negatives alongside capturing false negatives which should have been classified as cancer-positive. The training set's recall is better than the testing set's recall, in order to improve this, a wider distribution of training samples could be used using some level of data augmentation to generate more sub-samples or add noise to the current training set. The accuracies of both training and testing set are also comparable to the baseline paper, with an overall accuracy of 0.87. [22] Though the pretrained ResNet's F1 score is somewhat smaller than the baseline paper it outperforms the other CNN architectures explored in the baseline paper. Additionally the testing and training recall on the ResNet is better than the baseline model's performance, this indicates a better generalization and identification of features from the training set. It is important to note that the CNN 3 also contains more parameters than the ResNet which is likely to be the cause of these results.

TABLE I
PRECISION, RECALL AND ACCURACY MEASURES FOR THE TRAINING AND TESTING SET OF THE MODEL

Model	**Data segment**	**F1 Score** **F1 = (2*P*R)/(P+R)**	**Precision** **P = TP/(TP + FP)**	**Recall** **R = TP/(TP + FN)**
Pretrained ResNet (Our Model)	Training set	0.792	0.824	**0.736**
	Testing set	**0.806**	**0.876**	0.695
CNN 3 Performance (Baseline Paper)	Training set	**0.87**	0.82	**0.92**

The visualization of a probability map of portions containing IDC detected by our model is showcased below.

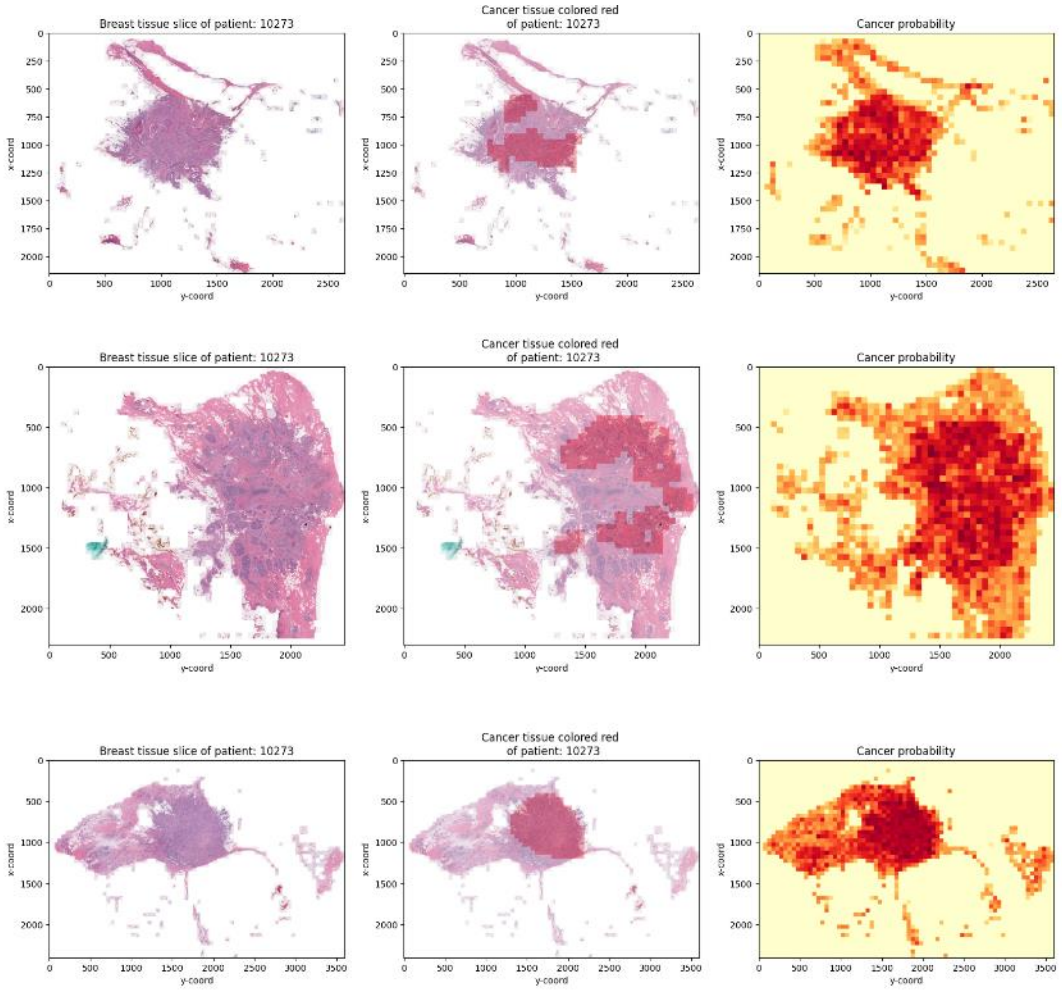


Fig. 10 Color map of cancer probability in each section of tissue for various patients

Finally the confusion matrix for both train data, and testing is given.

	predicted no cancer	predicted cancer
actual no cancer	0.842567	0.157433
actual cancer	0.263034	0.736966

	predicted no cancer	predicted cancer
actual no cancer	0.902257	0.097743
actual cancer	0.304533	0.695467

Fig. 11 a. Confusion matrix for training set b. Confusion matrix for testing set

VI. CONCLUSIONS

In summary, our study has shown how profoundly deep learning methods can be applied to the field of breast cancer detection. Through the utilization of an extensive dataset and sophisticated data augmentation techniques, we have established a resilient model that can precisely detect invasive ductal carcinoma (IDC) in breast tissue. The model's exceptional ability to distinguish between cancerous and healthy patches is highlighted by its precision, recall and high F1 score, as demonstrated by its performance on both training and test datasets.

In comparison to the reference baseline paper, our model showcased somewhat lower accuracy, this could potentially be due to using a subset of data as compared to the reference day. When using transfer learning, it is pertinent to include a large enough training size for the classifier and/or feature extractor in order to customize it to the dataset at hand. However, it is entirely possible that the model remains over-generalized and somewhat underfit to the training set. This can be resolved by expanding the training dataset size through data augmentation.

The scope of this project is unfortunately related to the complexity of the dataset, while there are pre existing issues with the dataset such as a relatively smaller number of samples and a high class imbalance, the complexity would increase drastically if the same values were collected from a larger population size from diverse regions. Perhaps it would be more challenging to attain a high accuracy. In order to expand on work performed in this paper, we could expand on the modalities and use tabular attributes alongside imaging or use a dataset with sequential imaging of cancerous and non-cancerous patch growth to analyse and predict presence/absence alongside the potential rate of growth.

VII. FUTUREWORK

Further investigation can be performed to identify the effect on CNN models with deeper architectures and validation on larger datasets. We can also automate the model to learn visual features related histopathology and morphology architectures to acquire a quantitative semantic characteristic that highlights the relevant regions in WSI, such as the percentage of necrotic tissue or the aggressiveness of the tumor. utilizing automated feature extraction methods, enhanced and diversified datasets for increased model robustness, and cutting-edge machine learning models for better outcomes. Improving interpretability of the model using techniques such as SHAP or LIME could offer insightful information about decision-making procedures, and investigating its implementation for real-time applications could greatly increase its influence. Limited labeled data difficulties could be addressed by implementing rigorous cross-validation, working with domain experts for validation, extending result analysis with metrics like ROC curves, and investigating unsupervised and semi-supervised learning methods.

ACKNOWLEDGEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Additionally. this research did not receive any grants from

government, private, or nonprofit funding bodies. The authors of this paper extend their gratitude to the computer science department faculty at Manipal Institute of technology, Manipal, especially to Murali Krishna sir.

REFERENCES

[1] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," Procedia Computer Science, vol. 218, pp. 1314–1320, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.110.

[2] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," Sensors (Basel, Switzerland), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164870/ (accessed Apr. 7, 2024).

[3] "Visual Lab - A Methodology for Breast Disease Computer-Aided Diagnosis," visual.ic.uff.br. http://visual.ic.uff.br/en/proeng/thiagoelias/ (accessed Apr. 23, 2024).

[4] J. Arevalo , F. A Gonzalez, and R. Ramos-Pollan, "Convolutional Neural Networks for mammography mass lesion classification," Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, https://pubmed.ncbi.nlm.nih.gov/26736382/ (accessed Apr. 7, 2024).

[5] D. C. Moura and M. López, "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis," vol. 8, no. 4, pp. 561–574, Apr. 2013, doi: https://doi.org/10.1007/s11548-013-0838-2.

[6]]N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893, 2005, doi: https://doi.org/10.1109/cvpr.2005.177.

[7] K. Petersen, M. Nielsen, P. Diao, Nico Karssemeijer, and M. Lillholm, "Breast Tissue Segmentation and Mammographic Risk Scoring Using Deep Learning," Lecture notes in computer science, pp. 88–94, Jan. 2014, doi: https://doi.org/10.1007/978-3-319-07887-8_13.

[8] N. M. ud din, R. A. Dar, M. Rasool, and A. Assad, "Breast cancer detection using deep learning: Datasets, methods, and challenges ahead," Computers in Biology and Medicine, vol. 149, p. 106073, Oct. 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.106073.

[9] P. Zhao et al., "A Comparative Study of Deep Learning Classification Methods on a Small Environmental Microorganism Image Dataset (EMDS-6): From Convolutional Neural Networks to Visual Transformers," Frontiers in Microbiology, vol. 13, Mar. 2022, doi: https://doi.org/10.3389/fmicb.2022.792166.

[10] Salman Zakareya, Habib Izadkhah, and Jaber Karimpour, "A New Deep-Learning-Based Model for Breast Cancer Diagnosis from Medical Images," vol. 13, no. 11, pp. 1944–1944, Jun. 2023, doi: https://doi.org/10.3390/diagnostics13111944.

[11] M. A. Naji, S. E. Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis," Procedia Computer Science, vol. 191, pp. 487–492, Jan. 2021, doi: https://doi.org/10.1016/j.procs.2021.07.062.

[12] Tala Talaei Khoei, Hadjar Ould Slimane, and Naima Kaabouch, "Deep learning: systematic review, models, challenges, and research directions," Neural Computing and Applications, Sep. 2023, doi: https://doi.org/10.1007/s00521-023-08957-4.

[13] Y. J. Tan, K. S. Sim, and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," IEEE Xplore, Nov. 01, 2017. https://ieeexplore.ieee.org/abstract/document/8308076

[14] M. Toğaçar, B. Ergen, and Z. Cömert, "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders," *Medical Hypotheses*, vol. 135, p. 109503, Feb. 2020, doi: https://doi.org/10.1016/j.mehy.2019.109503

[15] A. Cruz-Roa *et al.*, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent," *Scientific Reports*, vol. 7, no. 1, Apr. 2017, doi: https://doi.org/10.1038/srep46450.

[16] S. A. Alanazi *et al.*, "Boosting Breast Cancer Detection Using Convolutional Neural Network," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–11, Apr. 2021, doi: https://doi.org/10.1155/2021/5528622.

[17] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," Pattern Recognition, vol. 84, pp. 345–356, Dec. 2018, doi: https://doi.org/10.1016/j.patcog.2018.07.022.

[18] E. Mercan, S. Mehta, J. Bartlett, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions," JAMA Network Open, vol. 2, no. 8, pp. e198777–e198777, Aug. 2019, doi: https://doi.org/10.1001/jamanetworkopen.2019.8777.

[19] [1]L. Tapak, N. Shirmohammadi-Khorram, P. Amini, B. Alafchi, O. Hamidi, and J. Poorolajal, "Prediction of survival and metastasis in breast cancer patients using machine learning classifiers," Clinical Epidemiology and Global Health, vol. 7, no. 3, pp. 293–299, Sep. 2019, doi: https://doi.org/10.1016/j.cegh.2018.10.003.

[20] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of Pathology Informatics*, vol. 7, no. 1, p. 29, 2016, doi: https://doi.org/10.4103/2153-3539.186902.

[21] "Predict IDC in Breast Cancer Histology Images," *kaggle.com*. https://www.kaggle.com/code/paultimothymooney/predict-idc-in-breast-cancer-histology-images (accessed Apr. 07, 2024).

[22] "Table 6 | Boosting Breast Cancer Detection Using Convolutional Neural Network," *www.hindawi.com*. https://www.hindawi.com/journals/jhe/2021/5528622/tab6/ (accessed Apr. 07, 2024).

[23] C.-M. Chao, Y.-W. Yu, B.-W. Cheng, and Y.-L. Kuo, "Construction the Model on the Breast Cancer Survival Analysis Use Support Vector Machine, Logistic Regression and Decision Tree," Journal of Medical Systems, vol. 38, no. 10, Aug. 2014, doi: https://doi.org/10.1007/s10916-014-0106-1.

[24] P. Kumar, S. Srivastava, R. K. Mishra, and Y. P. Sai, "End-to-end improved convolutional neural network model for breast cancer detection using mammographic data," The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, vol. 19, no. 3, pp. 375–384, Dec. 2020, doi: https://doi.org/10.1177/1548512920973268.

[25] S. Z. Ramadan, "Using Convolutional Neural Network with Cheat Sheet and Data Augmentation to Detect Breast Cancer in Mammograms," Computational and Mathematical Methods in Medicine, vol. 2020, pp. 1–9, Oct. 2020, doi: https://doi.org/10.1155/2020/9523404.

[26] P. Kumar, S. Srivastava, R. K. Mishra, and Y. P. Sai, "End-to-end improved convolutional neural network model for breast cancer detection using mammographic data," The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, vol. 19, no. 3, pp. 375–384, Dec. 2020, doi: https://doi.org/10.1177/1548512920973268.

[27] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, “Image Data Augmentation for Deep Learning: A Survey,” arXiv:2204.08610 [cs], Apr. 2022, Available: https://arxiv.org/abs/2204.08610

[28] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. Shamim Hossain, “Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion,” Future Generation Computer Systems, vol. 101, pp. 542–554, Dec. 2019, doi: https://doi.org/10.1016/j.future.2019.06.027.

[29] G. Murtaza et al., “Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges,” Artificial Intelligence Review, May 2019, doi: https://doi.org/10.1007/s10462-019-09716-5.

[30] S. A. Alanazi et al., “Boosting Breast Cancer Detection Using Convolutional Neural Network,” Journal of Healthcare Engineering, vol. 2021, pp. 1–11, Apr. 2021, doi: https://doi.org/10.1155/2021/5528622.