

Explainable Breast Cancer Detection from Histopathology Images using Transfer Learning and XAI

by

Farhan Farabi
19101395

Farhan Hossen
19101447

Farhan Monsur
19101129

Mir Araf Hossain
19101105

Md. Mohibul Hasan
19101131

A thesis submitted to the Department of Computer Science and Engineering in
partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
January 2023

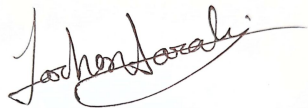
© 2023. BRAC University
All rights reserved.

Declaration

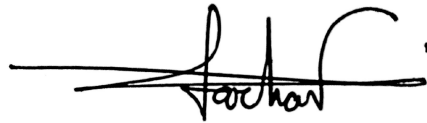
It is hereby declared that

1. The thesis submitted is our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



Farhan Farabi
19101395



Farhan Hossen
19101447



Farhan Monsur
19101129



Mir Araf Hossain
19101105



Md. Mohibul Hasan
19101131

Approval

The thesis titled “Explainable Breast Cancer Detection from Histopathology Images using Transfer Learning and XAI” submitted by

1. Farhan Farabi (19101395)
2. Farhan Hossen (19101447)
3. Farhan Monsur (19101129)
4. Mir Araf Hossain (19101105)
5. Md. Mohibul Hasan (19101131)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 23, 2023.

Examining Committee:

Supervisor:
(Member)



Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

Nowadays, research on many diseases such as cancer has been ongoing to determine how to reduce and minimise the effect. There are many characteristics of cancer that can be identified by their consistent cell proliferation and unique subgroups. Among cancer, breast cancer is responsible for many deaths each year and early detection increases the chance of survival. The proposed method employs three base models, VGG19, ResNet50V2 and MobileNetV2 which are trained on the BreakHis dataset, a public dataset of breast histopathological images. Furthermore, technology such as CNN and ML have become a tool for cancer researchers to identify cancer cells more efficiently. Feature extractors such as MobileNetV2, ResNet50V2 etc. models have been used for classification and detection. MobileNetV2 is a feature extractor for segmentation and object detection. Nearly all of the latest AI technology uses ResNet to build cutting-edge systems. A well-liked method for producing a class-specific heatmap using a trained CNN, a specific input image and a class of interest is called Grad-CAM. We trained our model using the transfer learning techniques using MobileNetV2, ResNet50V2, VGG19 as the base model and the weights of ImageNet. The model had an accuracy rate of 94.86%, 94.38%, 95.65% respectively. The features extracted from the last layer of the trained models are fused using concatenation and ensemble methods to improve the performance of the classifiers. Several linear classifiers including K-Nearest Neighbors (KNN), Stochastic Gradient Descent (SGD), AdaBoost, XGBoost, Decision Tree and Random Forest are used to classify the fused features. The results of the experiments show that the proposed method achieved high accuracy, with KNN classifier achieving the best result of 97.535% and Random Forest classifier achieving 97.455%. The proposed method is effective in breast cancer prediction and can assist pathologists in the diagnosis of breast cancer.

Keywords: AdaBoost, Decision Tree, Grad-CAM, MobileNetV2, Random Forest, ResNet50V2, VGG19, XGBoost.

Acknowledgement

First and foremost, thanks to the Great Allah, our thesis was completed without significant interruption.

Secondly, we want to thank our supervisor, Dr. Md. Golam Rabiul Alam, for his kind assistance and guidance. He was there for us anytime we needed him.

Thirdly, we would like to pay tribute to the late Dr. Hossain Arif sir, who served as our first supervisor and contributed so much for us, including helping us get our thesis off to a strong start and pointing us potential study possibilities.

And finally, thanks to our parents, without whom it could not have been possible. We are now just a few steps away from our graduation thanks to their heartfelt prayers and support.

Table of Contents

Declaration	i
Approval	ii
Abstract	iv
Acknowledgment	v
Table of Contents	vi
List of Figures	vii
List of Tables	viii
Nomenclature	ix
1 Introduction	1
1.1 Research Problem	2
1.2 Research Objectives	2
1.3 Thesis Organization	3
2 Literature Review	5
2.1 Related Works	5
2.2 Background Study	9
2.2.1 Deep Neural Network (DNN)	9
2.2.2 Convolutional Neural Network (CNN) Architectures for Feature Extraction	9
2.3 Explainable Artificial Intelligence (XAI)	11
2.4 Linear Classifiers	11
2.4.1 Support Vector Machine (SVM)	12
2.4.2 Adaptive Boosting (AdaBoost)	12
2.4.3 Extreme Gradient Boosting (XGBoost)	12
2.4.4 Logistic Regression	13
2.4.5 Random Forest	13
2.4.6 Decision Tree	13
2.4.7 Stochastic Gradient Descent Classifier (SGDC)	13
2.4.8 K-Nearest Neighbor (KNN)	14
3 Methodology	15
3.1 Dataset	17
3.2 Feature Extraction in Image Processing	18

3.2.1	Feature Extraction of MobileNetV2	18
3.2.2	Feature Extraction with VGG19	18
3.2.3	Feature Extraction with ResNet50V2	19
3.3	Breast Cancer Prediction	20
3.3.1	Feature Fusion	20
3.3.2	K-Nearest Neighbors (KNN)	21
3.3.3	Stochastic Gradient Descent (SGD)	21
3.3.4	Adaptive Boosting (AdaBoost)	22
3.3.5	Extreme Gradient Boosting (XGBoost)	23
3.3.6	Decision Tree	23
3.3.7	Random Forest	24
3.3.8	Logistic Regression	24
4	Implementation and Result	26
4.1	Implementation	26
4.1.1	Dataset Preparation	26
4.1.2	Fusion Model	27
4.1.3	GradCAM	28
4.2	Result	29
4.2.1	Base Model Performance	29
4.2.2	Fusion Model Performance	30
4.2.3	Performance Metrics	30
4.2.4	Classification Report	35
4.2.5	Discussion	37
5	Conclusion	38
	References	43

List of Figures

2.2.1	VGG Network Structure	10
3.0.1	Overview of Our CNN Classifier Network	16
3.1.1	Some Sample Images of the BreakHis Dataset [14] under Different Magnification	17
4.1.1	Data Augmentation	27
4.1.2	Grad-CAM Heatmap of VGG19, MobileNetV2, ResNet50V2 where the Red Region Indicates the Region with High Activations and Blue Means Low Activation	29
4.2.1	MobileNetV2	31
4.2.2	ResNet50V2	31
4.2.3	VGG19	32

4.2.4 Confusion Matrix of All the Proposed Classifier Used on the BreakHis Dataset	33
4.2.5 Comparisons of Accuracy, Sensitivity, Specificity and Matthews Correlation Coefficient of All the proposed Classifiers	34

List of Tables

3.2.1 The Summary of VGG19 Model	19
4.2.1 Existing Models used On the BreakHis Dataset	30
4.2.2 Our Models on the BreakHis Dataset	30
4.2.3 Existing Fusion Model used On the BreakHis Dataset	30
4.2.4 Our Fusion Model used On the BreakHis Dataset	30
4.2.5 Classification Report of the Classifiers	36

List of Equations

3.1 AdaBoost initialization Step 1	22
3.2 AdaBoost initialization Step 2	22
3.3 AdaBoost updating distribution equation	22
3.4 Final score equation 1	22
3.5 Final score equation 2	22
3.6 logit equation	25
3.7 sigmoid function	25
4.1 Sensitivity equation	31
4.2 Specificity equation	31
4.3 Precision equation	35
4.4 Recall equation	35
4.5 F1-score equation	35

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AdaBoost Adaptive Boosting

AI Artificial Intelligence

CNN Convolutional Neural Network

DNN Deep Neural Network

FFT Fast Fourier Transform

Grad – CAM Gradient-weighted Class Activation Mapping

KNN K-Nearest Neighbors

ML Machine Learning

ResNet Residual Network

SGD Stochastic Gradient Descent

SVM Support Vector Machines

VGG Visual Geometry Group

XAI Explainable Artificial Intelligence

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

Cancer refers to a disease where the body cell grows unrestrained and that can affect any body region. Cancer can also be referred to as aggressive tumors or neoplasms. The rapid growth of aberrant cells that can go beyond their borders is a defining trait of cancer. Furthermore, these cells infiltrate and spread to various organs in the human body. Detecting cancer can be done using image processing and deep learning, machine learning processes. Cancer will claim the lives of approximately 10 million people worldwide by the year 2020, making it one of the most preventable causes of death [51]. Furthermore, by 2030, the number of deaths from lung cancer worldwide is expected to reach 17 million [7]. Breast cancer (2.26 million) was the most frequent cancer among them and the other causes were lung cancer (2.21 million), colon cancer (1.93 million), prostatectomy cancer (1.41 million), skin cancer (1.20 million) and stomach cancer (1.09 million).

Breast cancer is the cause of rapid change or damage to the DNA that affects the breast tissue. All living things expand through cell division and the replacement of old ones. Through X-Ray imaging, this rapid and uncontrolled division of cells that cause breast tumors can be detected early. The most noticeable sign of breast cancer is the presence of a lump. As a result of a genetic change [12] (BRCA1 & BRCA2), oral contraceptives, hormone replacement therapy and menopause are the leading causes of breast cancer in women. Mammogram photos, histopathology images etc. provide information about tumors or malignant tissues after a biopsy using X-Ray images. Breast cancer detection has long been a challenge for pathologists and medical practitioners. Several imaging methods are employed to determine the best therapy option for each patient. To gather enough information, imaging techniques are frequently combined. Multiple deep learning and image processing approaches such as Support Vector Machines (SVM) [4], multistage classification, Convolutional Neural Network (CNN), Fast Fourier Transform (FFT) [1] and Two-Dimensional Principal Component Analysis (2DPCA), can be used to diagnose cancer. Deep learning is commonly used in medical imaging because it requires no prior knowledge of a related field. Chemotherapy is commonly used to eliminate cancer cells as a treatment for cancer. MobileNetV2 is a CNN model which is used to detect the cancer tissues. Chemotherapy works by inhibiting cancer cells' growth, division and reproduction. This is the mechanism through which chemotherapy kills cancer cells. However, the medications used to treat cancer cells are incredibly potent and can also destroy normal cells. Standard chemotherapy, classical chemotherapy and cytotoxic chemotherapy are names that used to describe the administration of these

powerful drugs.

In addition, physicians utilize chemotherapy in various ways and a variety of chemotherapy medications are employed. Neoadjuvant, adjuvant, leukaemia, lymphoma, alkylating agents and nitrosoureas. Nonetheless, many people also die due to medical blunders and over 250,000 people in the United States died as a result of medical blunders in 2012, as stated in a study by Johns Hopkins University [6] [58], as a result, misdiagnosis and mistreatment became the 3rd leading cause of death after heart disease and cancer. In the paper [58], they determined the condition of a cancerous tumour cell using various image processing techniques such as SVM, CNN, FFT etc.

1.1 Research Problem

One of the most common causes of death is cancer. However, accurate diagnosis and prognosis of this disease is difficult to achieve. We attempt to address this barrier by using machine learning and other computer-aided testing methods. With machine learning techniques, it has been possible to enhance the accuracy of cancer prediction outcomes by 15%–20% in recent years [6]. There have been multiple research reports that can aid with early cancer detection and prognosis based on various methodologies [8] [10] [11] [12].

In addition, these kinds of choices are typically made by medical professionals who have spent several years focusing on a subset of the subject. After that, they can apply their knowledge and skills to develop a diagnosis and treatment strategy tailored to the patient's needs. Despite this, many individuals die each year due to medical blunders. Medical errors also cost the United States \$20 billion per year. Therefore, minimizing medical errors is crucial [27]. In the field of breast cancer diagnosis, the integration of automated detection techniques has the potential to improve patient outcomes through earlier and more accurate identification of malignant tumors. However, current research utilizing publicly available datasets has not demonstrated a level of accuracy sufficient to fully replace physician involvement in the diagnostic process and fully realize the potential of computer-aided diagnosis (CAD) systems. Our study aims to address this limitation by developing a model that improves the accuracy and performance of breast cancer detection in order to facilitate the implementation of CAD systems in clinical practice [39].

1.2 Research Objectives

The objective of our research is to develop a robust and accurate computer-aided diagnosis (CAD) system for breast cancer prediction. Despite the abundance of research on breast cancer detection using various publicly available datasets, there is still a lack of an effective CAD system that can minimize the need for physician participation and maximize the use of computer-aided design. Our proposed system aims to address this limitation and contribute to the field by utilizing advanced machine learning and deep learning techniques to improve the accuracy and efficiency of breast cancer diagnosis. The novelty of our approach lies in the integration of

multiple base models, feature fusion using ensemble techniques and the application of state-of-the-art linear classifiers to enhance the performance of the CAD system.

- We want to integrate three CNN models as feature extractors because using multiple CNN models as feature extractors in breast cancer prediction can increase accuracy, provide a more robust feature representation, reduce overfitting and improve generalization ability.
- Using the extracted features, we want to combine them into a single feature as it would improve the performance of the classifiers by combining multiple features extracted from different models, thus providing a more comprehensive representation of the data.
- Then using the fused feature we will use several classifiers to determine which fusion model gives a better accuracy and understanding of breast cancer classification.

Therefore, our research endeavors to contribute to the field of computer-aided diagnosis by providing a more comprehensive understanding of improved classification using CAD systems for breast cancer detection.

1.3 Thesis Organization

With the help of multiple CNN models, linear regression and XAI, this study has been separated into five sections to describe how it attempts to identify explainable breast cancer in images of breast histopathology. The research problem and the goals that the study will progressively achieve are introduced in Chapter 01 along with the research problem that needs to be solved. Our goal is to compare various feature extractor combinations used in CNN models and incorporate Explainable AI into the network to accurately distinguish between benign and malignant cells in breast histopathology images. Grad-CAM and an XAI technique will also be used to judge the model's performance outside the black box.

The works pertinent to this particular study are discussed in Chapter 02 along with the neural network models and methods necessary for the investigation. The outcomes of a few related publications that used the BreakHis dataset to predict or diagnose breast cancer are also included in this section. We have employed linear classification methods in addition to neural networking models to more accurately detect the presence of cancer cells.

The research methodology is covered in Chapter 03 which emphasizes the methodological steps taken to conduct the study gradually. The dataset at question is the Breast Histopathology dataset, also referred to as the BreakHis dataset. VGG19, ResNet50V2 and MobileNetV2 are three extractors used in feature extraction for image processing. This chapter expands on the Grad-CAM Explainable AI technique that we utilize for the network's cancer region. Additionally, the preparation as well as the specifics and characteristics of the dataset are also included in this section of the study.

The different approaches of CNN, DNN and Linear Classification are also discussed

in Chapter 03 in relation to their application. Models like ResNet50V2, VGG19 and MobileNetV2 are employed in convolution neural networks. The linear classification employs SVM, AdaBoost, XGBoost Classifier, Logistic Regression, Random Forest Classifier, SGD Classifier and K-Nearest Neighbors models. We've spoken about the steps we took to put the algorithms into practice so that we could effectively identify the existence of the malignant cell from the breast histopathology photos. In this case, the technique we used to implement the algorithms to get the necessary accuracy was explained.

The experimental assessment and the data analysis are highlighted in Chapter 04. The key variable which will be generated by the various models and a threshold value which will be used to compare the prediction value and determine whether or not the histopathological picture contains cancer, form the basis of our evaluation. In order to improve the dependability or credibility of our findings, we must check it once we have discovered it from sources outside the black box. Grad-CAM has been employed in this instance as an Explainable AI approach to extract a heatmap from the histopathology photos that would display the ROI region that the model has selected to provide the outcome.

By giving a summary of the complete research process, Chapter 5 concludes the investigation.

Chapter 2

Literature Review

2.1 Related Works

This part will look into the success rate of CNN models in identifying cancer and improving the accuracy. Also, we will study recent papers that have worked on identifying cancer cells from images. To that aim, we'll take a look at the challenges that various network systems have faced and the solutions that have been devised in response.

In [6] [57], we can see the predominance of neural network-based algorithms such as CNN and ANN in the automated detection and classification method. A file containing tiny biopsy photographs will be loaded into the application in this method and also the algorithms are used to read and segment images via image processing. Machine learning is used to train and test the photographs. It evaluates the pictures and gives a good or negative answer. This method saves time while accurately forecasting outcomes.

Breast cancer treatment is one area where these strategies may be put to use and the iterative development of the algorithm can be used to illustrate the discovery's value. Most of the most popular evolutionary algorithms, when correctly configured, showed comparable results in the first experiment. Methods for feature selection were used to improve accuracy in the second experiment. Last time it looked at this, it found out how to automatically create a supervised classifier using machine learning [30].

Nuclear characteristics acquired immediately from a digital scan of fine needle aspiration (FNA) slides have been used in another breast cancer research project. This methodology employs a computer-intensive search to find the attributes that allow the classification algorithm to best fit the data by uncovering diagnostic or prognostically important traits. These features are utilized sequentially with 90% of the data to build classifiers, with each classifier tested on the remaining 10% (cross-validation) [2].

In [17], a new approach for texture feature classification using neural networks is proposed. The paper's primary goal is to create distinct texture features by curving out the ROI of a mammography image. The texture features are then extracted using the ROI obtained. These characteristics are fed into a neural network that

determines whether the photos are malignant. For weight adjustment, the neural network is trained using the back-propagation technique.

The following are the primary phases of another proposed method from [29]:

- An image is sharpened.
- The environment and the item classify its texture.
- Its results are interpreted.

In [5], they detail both the background texture extraction and the object texture extraction processes. Information similar to this may be found in [5]. Aside from the sharpening phase, the object texture segmentation and background analysis classification steps make up the bulk of this approach.

Another technique called Median Filtered Image used for image enhancement outperformed when experimented [43], along with Gaussian, average filter. The median filter reduced noise and preserved edges in noisy mammography pictures, helping to improve image clarity so that the radiologist can discover abnormalities or possible damaged areas and distortions in intensity or pixel values. These fundamental capabilities improve the system's effectiveness and aid in detecting cancer tissue in the breast more efficiently [23] [35].

Parker et al. [58], in their work described a model based on DNN named C3R that recommends chemotherapy for colorectal cancer patients after surgery. This study's model used only one hospital patient dataset to limit generalization and ensure scalability. Furthermore, the colorectal cancer data dictionary harmonizes hospital data variables and formats so later researchers can add other hospitals' data. The C3R model accuracy was 70.5% Top-1 and 84.5% Top-2 NCCN (National Comprehensive Cancer Network). Korea's medical insurance arrangement prevents NCCN-recommended chemotherapy because they recommend chemotherapy for patients based on the HIRAS (Health Insurance Review and Assessment Service). For example, if patients choose a non-HIRA-compliant treatment, they won't be eligible for health insurance.

To detect breast cancer at an early stage, Ansar et al. [38] introduced a MobileNet-based architecture to differentiate between cancerous and noncancerous lumps. Compared to DDSM and CBIS-DDSM, it was faster to calculate and used less memory while still achieving the same level of accuracy (86.8%). As a result, they achieved better results than AlexNet, VGG16, GoogleNet and ResNet. They want to construct a CAD system that can run on its own, monitoring millions of routine imaging tests and sending out alerts to radiologists if they detect any tumors. They propose first training a complex CNN framework with fewer parameters on cropped picture patches (identified ROIs), then applying it to full mammography pictures. Due to the limited size of the dataset, CNN models overfit the data and had trouble keeping up with new data. When training occurrences were insufficient, data augmentation was used to boost the number of occurrences. Data flipping and rotation achieved this.

In the work by Saric et al. [34], they scanned H&E-stained histological slides from experienced pathologists and identified the cancerous spots. Then they extracted ROI and (256 x 256) patches with stride 196 that achieve a suitable level of patch, they used overlapping to create training samples from the ROIs and identified that if 75% of a patch's pixels are tagged as tumors then the whole patch is classified as a tumor. Ultimately, they measured the accuracy of VGG16 and ResNet50V2 and found that ResNet has better accuracy (75.2%) on the ImageNet dataset than VGG16 (70.5%) for Top-1 accuracy and 93% vs 91.2% for Top-5 accuracy. They found that the results deviate from the behavior that was expected. Although the results indicate that CNN-based classification has a promise for lung cancer diagnosis, the given method performs less well than existing methods for detecting other cancer types using entire slide images.

Lymph node sections stained with hematoxylin and eosin (H&E) were photographed and collected from two specialized hospitals in Vietnam to create the VBCan dataset which was introduced in a research by Vo-Le et al. [53]. There are 3,529 high-resolution (512 x 512) images in the collection. Then, a two-step process is provided for evaluating breast cancer detection effectiveness. This strategy combines feature extraction from a state-of-the-art CNN like VGG16, GoogLeNet or ResNet50V2 with a variety of conventional machine learning classifiers. The results of the optimization show that a 96.98% accuracy may be achieved on VBCan by using the ResNet50V2 model to extract features and Softmax as a classifier. When compared to other models, it is established that the VGG16 model's recall rate using its original classifier is the highest (97.76%) while the GoogLeNet model achieves the best precision (98.58%).

In their study, Manasa et al. [54] gathered a dataset from kaggle containing information about cancer and benign diseases. The collection has 1497 images of potentially cancerous moles and 1800 images of noncancerous moles. After adjusting their dimensions to (224 x 224 x 3), the images were read into a numpy array as RGB values. He employed ResNet50V2 and VGG16 in order to fine-tune the precision of his work. With VGG16 and ResNet50V2, we have a final accuracy of 86.6%. In addition, Ali khan et al. [40] used image processing and data augmentation techniques to a small dataset of 253 brain MRI pictures from brain MRI images dataset for brain tumor diagnosis, Kaggle, 2019. He utilized transfer learning to compare his scratched CNN model to others such as the VGG16, ResNet50V2 and InceptionV3, all of which had already been trained. ResNet50V2 which was trained with a 32-person batch across 15 epochs, achieved 92% accuracy on training data and 87% accuracy on validation data. However, InceptionV3 achieved 93% in training and 83% in validation. VGG16 achieved an accuracy of 90% on training data and 87% on validation data.

We can observe two efficient deep transfer learning-based models which are proposed in [45]. They used images from the ImageNet dataset together with a pre-trained deep convolutional neural network to enhance binary and multiclass categorization. To distinguish between malignant and benign sample tissues in binary and multiclass classification, they employed ResNet50V2 and DenseNet121 weights from ImageNet as starting weights and then enhanced these models using a deep classifier with data augmentation. Using improved hyperparameters, the proposed models were

tested in both magnification-based and non-magnification-based classification settings. This technique forecasts properly and rapidly. (In a system where several classes need to be categorized, the proposed method achieves a remarkable 98% accuracy.

CNN is used by Zhan Xiang and coworkers to classify histopathology photos into benign and malignant tumors using the Softmax function by extracting characteristics from the images. Using the BreakHis database, which has been open to academic research since 2014, we were able to demonstrate the algorithm's superior accuracy. Breast tumors, both benign and malignant are represented in BreakHis's 7909 microscopic histopathology images. The samples are histological slides of biopsied breast tissue stained with hematoxylin and eosin (H&E). Patients with breast cancer and clinical researchers at the P&D Laboratory worked together to collect and label images. For this study, the team of researchers used a customized version of the deep learning architecture InceptionV3. If you've seen InceptionV2, you'll be impressed with InceptionV3. Greater convolutions are the most unusual to factorize. (Achievery between 80% and 85%) [36].

The training samples were tiny and high-resolution, thus we used a patch-based approach to image classification [41]. Many mistaken patches plague patch-level datasets derived from whole slide images (WSIs), making patch-based classification difficult. The efficiency of patch-based categorization algorithms is not improved by considering how to deal with incorrectly tagged patches. Multi-layered features were retrieved from discriminative patches using a densely connected convolutional network (DenseNet), while mislabeled patches were filtered out using an unsupervised anomaly detection method using a generative adversarial network (AnoGAN). DenseNet121-AnoGAN excels in coarse-grained high-resolution pictures and works well at 40X and 100X magnifications. They put AnoGAN through its paces on the AlexNet, VGG16, VGG19 and ResNet50V2 networks. (At 99.13% and 99.38%, respectively) [41].

These authors used automated diagnostics based on machine learning to make correct and timely diagnosis. The nuclei of cells are separated using k-means clustering and DWT is then utilized to visualize the differences between the clusters. Discrete wavelets are used in this article, much like coiflets, but with scaling functions and vanishing moments instead. Energy, entropy and the logarithm of energy entropy are all provided via the LL, LH, HL and HH subbands, respectively. They used the BreakHis dataset in their study. Using the proposed method, the accuracy of linear SVM's was 93.3 percent, that of quadratic SVM's was 92.7 percent, and that of fine Gaussian SVM's was 91.3 percent [25].

High-level features are extracted from the BreakHis benchmark histopathology image dataset using a transfer learning model based on Visual Geometry Group with 16-layer deep model architecture (VGG16) in paper [47]. In order to address various breast cancer scenarios, various machine learning models (classifiers) are applied. This study used 40X-magnified photos from the BreakHis dataset for its two primary investigations. The VGG16 is used to extract 4096 characteristics for each image. The dataset is built using these features and the class labels for the photos. As per [37], this dataset is split into 90% training data and 10% testing data. Then,

five different classifiers, RBF, SVM, LR, Poly SVM, KNN with $k=1$ and NN with 300 iterations were trained. The outcomes demonstrate that the suggested models can outperform contemporary classical machine learning methods.

Test accuracy, AUC, precision, recall and F1-score were used by Farjana Parvin and coworkers to rank five convolutional neural network architectures, including LeNet5, AlexNet, VGG16, ResNet50V2 and InceptionV1. In the BreakHis dataset, the InceptionV1 network achieved the highest scores at 40X, 100X, 200X and 400X magnification, respectively, at 89%, 92%, 94%, and 90%. Across all four magnification levels (40X, 100X, 200X and 400X), the InceptionV1 network performed best in terms of AUC, accuracy, recall and F1-score [42].

In order to classify breast cancer in histopathology photos, Boumaraf et al. [48] compared traditional ML techniques with deep learning. They used transfer learning on the VGG19 architecture to train deep learning methods and three extractors in addition to traditional classifiers trained on hand-crafted features. Binary classification (94.05% to 98.13%) and eight-class classification (76.77% to 88.95%) were both improved by the use of deep learning algorithms compared to traditional approaches. The decision-making process was shown with the help of Grad-CAM and model comprehension was improved as a result.

2.2 Background Study

2.2.1 Deep Neural Network (DNN)

Deep neural networks are difficult to understand. Hierarchies are made more accessible. Data is processed via a hidden layer. After each epoch, the input data error rate is reduced by changing node weights, back-propagating the network and repeating. Nodes in the input layer are infinite. The output layer of DNN with more nodes speeds up learning. It is possible to pick individual node outputs. Take into account the input and output nodes and the bias, learning rate, initial weights, hidden layers and hidden nodes. To avoid null network outcomes, this model employs bias. The model's default learning rate is 0.15. The network adjusts node weight based on the error rate during back-propagation. Hidden layers and nodes are defined by inputs and data size. Network termination is determined by the number of epochs or the learning model prediction. Model training time and resources are increased by using layers and nodes [26].

2.2.2 Convolutional Neural Network (CNN) Architectures for Feature Extraction

CNN examines image patterns. Patterns can be discovered by distorting an image. These patterns could be transmitted farther into our neural network to identify more complicated aspects. This aids CNN's photo identification.

Convolutional, pooling and fully connected CNN's contain three layers. First-layer neurons calculate the local output. Weights and regions are used to calculate each. Image input filter sizes commonly used include 3×3 , 5×5 or 8×8 . These filters

use a sliding window to scan the image for repeated patterns. The stride is located between two filters. Convolution overlaps when the stride is smaller than the filter dimension.

Convolutional layers highlight specific aspects of the image. Feature maps will alter if placements change. This can be fixed by down sampling each convolutional layer's output. A picture can be down-sampled by shifting the phase of convolutional layers. It is more customary to use a pooling layer. This method improves precision [46].

2.2.2.1 VGG19 Architecture

The well-known convolutional neural network model VGG19 is easy to use and exceptionally good at tasks like target identification and image categorization [13].

The following figure illustrates how the VGG network structure functions (Figure 2.2.1). VGG has a deeper network than a typical convolution neural network because it has numerous convolutional layers, multiple convolutional layers, nonlinear activation layers and more. The framework makes it simpler to extract visual features. VGG19 is the name of one of the 19-layer VGG networks. The VGG19 employs a 33 filter to gather features of the picture detail and consists of 5 stages of convolution layers, 5 pooling layers and 3 fully linked layers. Better feature vector extraction is made possible by the increase in the depth of the convolution kernel in the VGG19 network from 64 to 512. There will be a pooling layer following each tier of convolutional layers. Each pooling layer has the same dimensions and step sizes (22). The artificial neuron is a Rectified Linear Unit, but a pooling layer is applied to down-sample the pooling layer. There are 4096, 4096 and 1000 nodes in total throughout the three fully connected tiers. A Softmax Regression classifier which is the last layer of the convolutional neural network, classifies the input images by probability. The input image size is 224 x 224 [55].

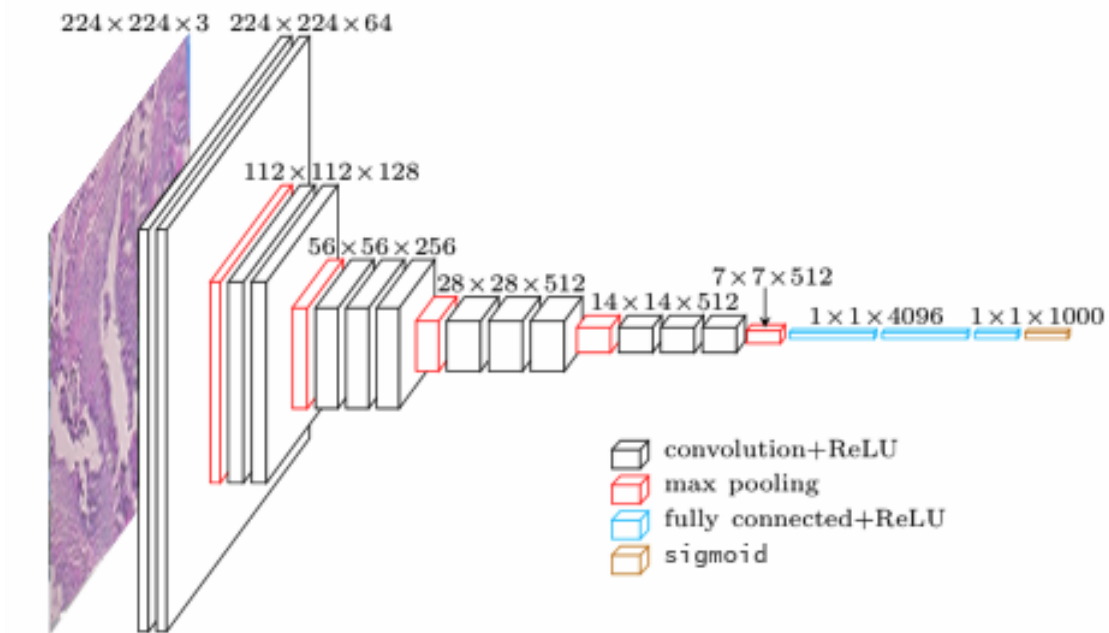


Figure 2.2.1: VGG Network Structure

2.2.2.2 ResNet50V2 Architecture

ResNet50V2 is also a convolution neural network with a structure of 50 layers. Kaim-ing He et al. proposed ResNet in 2015. The ResNet network offers a fresh concept. It facilitates the completion of challenging jobs and boosts the detection's precision. To fix the fully convolutional training process' saturation and accuracy loss problems, ResNet was developed. This work makes use of the ResNet50V2 architecture. Various sets of identical layers are present throughout the ResNet50V2 structure as represented by different. There are blocks known as identify blocks that are used to show how prior layers are utilized in subsequent layers. The degradation problem (accuracy initially saturates and then decreases) and counterfeit problem (vanishing or exploding gradients) in very deep network training are the fundamental differences in ResNet50V2 [56].

2.2.2.3 MobileNetV2 Architecture

A convolutional neural network design called MobileNetV2 intends to work well on mobile devices. The layers of the bottleneck are connected by residual connections in an inverted residual structure, which serves as its base. In order to filter features, lightweight depthwise convolutions are utilized in the intermediate expansion layer as a source of non-linearity. A total of 32 filters make up the first fully convolutional layer in MobileNetV2's architecture which is followed by 19 remaining bottleneck layers. MobileNetV2's primary organizational framework is based on that of MobileNetV1, its predecessor. The Depthwise Separable strategy is implemented by MobileNetV2 in order to address the problem of information loss that occurs in non-linear layers during the process of convolution blocks. The author also said that it also presented a brand new data structure that was referred to as inverted residuals [28].

2.3 Explainable Artificial Intelligence (XAI)

AI has the capacity to make judgments and such decisions can have both favourable and unfavourable effects on businesses. Understanding how AI makes judgments is crucial, just like when a business hires decision-makers. Many companies want to implement AI but many are hesitant to do so because they don't yet trust the model to make more crucial decisions because in most cases, the organisations don't have a clear understanding of how the model operates or the reasoning behind the AI's particular decisions. This is facilitated by explainability which sheds light on how models arrive at their conclusions. XAI is an approach to applying AI that explains the model's reasoning to humans, boosting their confidence in the model's ability to provide accurate results [59].

2.4 Linear Classifiers

An algorithm (classifier) for categorizing data called linear classification bases its judgments on a linear prediction function that combines a feature vector with a set of weights. A linear classifier determines a class's score as the weighted average of each of the three color channels' pixel values. Depending on the values we explicitly

defined for these weights, the function can either like or dislike particular colors at specific spots in the image. A few examples of convolutional neural networks are the well-known Support Vector Machine, AdaBoost, XGBoost, Logistic Regression, Random Forest etc. classifiers.

2.4.1 Support Vector Machine (SVM)

SVM was proposed by Vapnik based on the statistical learning theory which stands for Support Vector Machine. It was first created for binary classification but it can be effectively expanded for multiclass issues with applications in a variety of domains. Finding an optimum decision boundary that indicates the greatest separation between the classes is the main function of an SVM classifier. SVM's fundamental approach begins by resolving linear separable problems before expanding to include nonlinear issues [49].

2.4.2 Adaptive Boosting (AdaBoost)

In 1995, Yoav Freund and Robert Schapire created the statistical classification meta-algorithm known as AdaBoost or "Adaptive Boosting" using these weights. AdaBoost is also known as "Adaptive Boosting.". In order to enhance performance, AdaBoost can be used in combination with a wide variety of other learning methods. The result of other learning algorithms is combined into a weighted sum, which then serves as the representation of the final output of the boosted classifier. AdaBoost is initially presented as a method for binary classification. However, it has the potential to be extended to numerous classes as well as bounded intervals on the main diagonal [3].

2.4.3 Extreme Gradient Boosting (XGBoost)

The XGBoost machine learning framework is a robust, generalized gradient-boosted decision tree system. XGBoost is the finest machine learning software for regression, classification, and ranking tasks and enables parallel tree boosting. To appreciate how Extreme Gradient Boosting delivers concurrent tree boosting, it is essential to comprehend how classical supervised learning, decision trees, ensemble learning, and gradient boosting are generated. In supervised machine learning, a model is trained using algorithms to detect patterns within a dataset of features and labels, and the model is then used to predict the labels on the characteristics of a new dataset. A comparable ensemble learning strategy for decision trees to random forest. This dataset for classification and regression is known as a GBDT [19].

Gradient boosting refers to the concept of "boosting," or, in other words, enhancing a weak model by combining it with a number of additional weak models to get a model that is stronger as a whole. Gradient boosting is an extension of boosting in which the weak model generation process is expressed as a gradient descent method over an objective function. Gradient boosting sets desirable outcomes for the following model in an effort to minimise mistakes. Targeted outcomes for each scenario rely on the error gradient compared to the forecast (thus the name gradient boosting). The creation of Extreme Gradient Boost, the beginnings of "gradient boosting," was motivated by the need to improve the performance and computing speed of machine

learning models. It is a versatile and highly accurate gradient boosting technique that exceeds the processing limits of boosted tree algorithms. Extreme Gradient Boost produces trees continuously, whereas GBDT constructs them sequentially. It analyzes the quality of splits at each possible split in the training set using a level-wise technique that scans gradient values and partial sums [52].

2.4.4 Logistic Regression

Logistic Regression is a popular supervised machine learning method that is utilized often in classification and predictive analytics applications. Logistic Regression analyzes a dataset to calculate the probability of a particular occurrence, such as selecting one option over another or selecting neither option, given a set of independent factors and the data. Because the dependant variable might take on values between 0 and 1, the final outcome is a probability. A logit formula is used in the process of Logistic Regression. The odds, which are calculated by dividing the chance of success by the probability of failure, are then converted. In Logistic Regression, the data points are not organized in line rows as they are in traditional regression. There may be a lot here, a pile, with each pile designating a category and the same category name is present for each kind of data item. There may also be a pile. A method of optimization is utilized by the training classifier in order to determine which of the regression coefficients in the label produces the best results [31].

2.4.5 Random Forest

Random forest, developed by Leo Breiman and Adele Cutler, is a popular machine learning technique that combines the results of several decision trees into a single result. The random forest algorithm is an extension of the bagging method that generates a forest of decision trees that is independent of each other by randomly selecting features. Feature randomization, often called feature bagging or "the random subspace technique," ensures minimal correlation among decision trees by generating a random selection of features. The key difference between Random Forests and Decision Trees is this. While Decision Trees consider all possible feature splits, Random Forests only pick some of them. The Random Forest method may be used by computational biologists to classify gene expression data, identify biomarkers and annotate sequences. Thus, pharmacological responses to medications may be estimated by doctors [9].

2.4.6 Decision Tree

Decision Tree classifier is a supervised machine learning algorithm that is used for classification tasks. It is a simple and interpretable algorithm that creates a tree-like structure of decisions and their possible consequences. Decision tree works by recursively splitting the data based on the feature that provides the most information gain until a stopping criterion is reached [33].

2.4.7 Stochastic Gradient Descent Classifier (SGDC)

Stochastic Gradient Descent (SGD) classifier is a linear classifier that is used for classification tasks. It is a variant of gradient descent optimization algorithm, where

the weights are updated based on a random subset of the training instances at each iteration. This makes the algorithm more efficient and less prone to overfitting [44].

2.4.8 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a non-parametric, instance-based supervised learning algorithm. KNN is a type of lazy learning algorithm which means that it doesn't build a model until a new data point is encountered. In KNN, the idea is to find the k nearest points in the feature space to a new data point and predict the class based on the majority vote of the k nearest points. KNN is often used for classification tasks but can also be used for regression tasks [50].

Chapter 3

Methodology

Our research proposes a strategy for predicting breast cancer by fusing features and employing transfer learning. Firstly, the input images are Whole Slide Image (WSI) and the number of images are increased by augmenting the data such as slicing each WSI into a number of patches and preprocessing those patches by resizing the image into 224 x 224 x 3. The new dataset is then used into three pre-trained deep learning models: ResNet50v2, MobileNetV2 and VGG19 for training. From each of the trained models we extracted the feature from the last convolution layer and forwarded it to the Grad-CAM to display the heatmap by which the model defining the image is benign or malignant. As we can see in the below Figure 3.0.1, the Grad-CAM is explaining how the models are classifying each image by showing the heatmap of where the models are detecting the cancer cells. The last layer feature of each model was extracted and fused together using a concatenation process. The fused features were then used as input for linear classifiers such as KNN, SGD, AdaBoost, XGBoost, Decision Tree and Random Forest. The features from the different models were then fused together to create a more robust and informative representation of the image. A number of different metrics including accuracy, precision, recall and F1-score were used to assess the performance of the technique. In the following sections, we will offer a full description of the dataset that was used in the study as well as the pre-processing, base models, linear classifiers and explanation approach that was utilized.

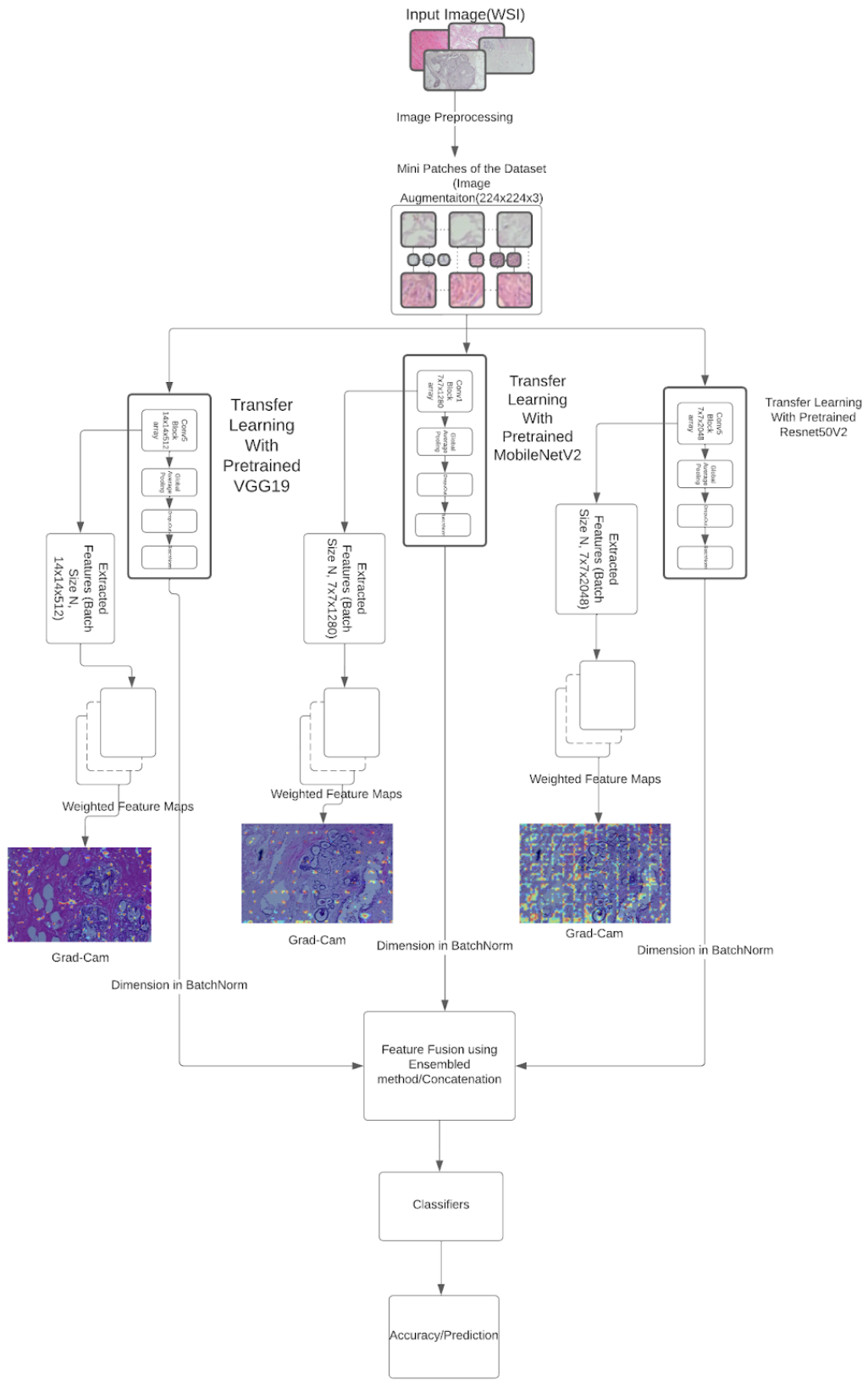


Figure 3.0.1: Overview of Our CNN Classifier Network

3.1 Dataset

Histopathological slides which contain tissue samples taken from a patient’s breast after a biopsy, are commonly used for the analysis of breast tumors. The 82 patients whose photos make up the 7909 in the BreakHis dataset were photographed by a Brazilian research institute between January and December of 2014. There are a total of 2480 photographs categorized as benign tumors and 5429 images categorized as malignant tumors. The microscopic appearance of the tumor is used to further classify each of these groups into one of four subgroups. There are numerous photos of each patient in the collection, all of which have been tagged with both the primary and secondary classes. Each picture is a unique combination of magnification (40X, 100X, 200X and 400X), pixel size (some are RGB and others are grayscale) and color scheme (some are black and white and others are color). Difficulties arise with this dataset because of the wide range of picture quality, magnification and tissue type included in the samples.

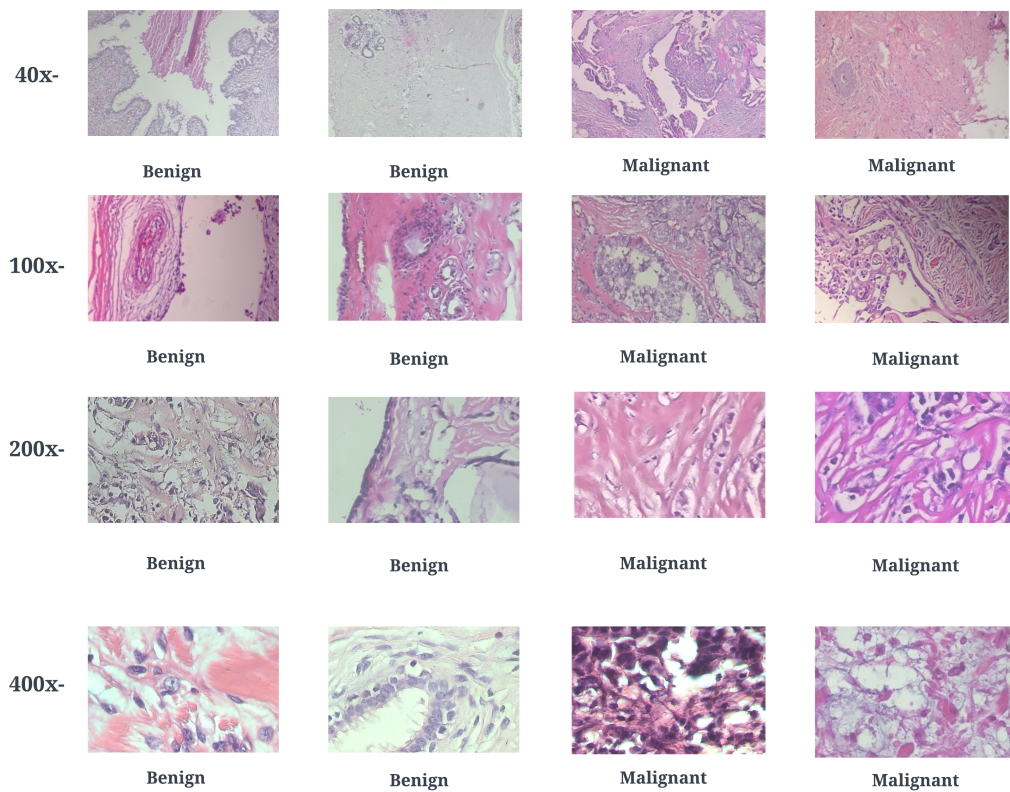


Figure 3.1.1: Some Sample Images of the BreakHis Dataset [14] under Different Magnification

3.2 Feature Extraction in Image Processing

The process of finding and extracting crucial and relevant information from a picture is referred to as "feature extraction" in the field of image processing. This allows the image to be represented in a form that is both more compact and more helpful. The goal of feature extraction is to reduce the amount of data in an image while preserving the most important information. One of the techniques we used was the Convolutional Neural Networks feature which essentially extracts the features from an image by training them on the BreakHis dataset. After the CNN has been trained, it may be used to extract features from fresh pictures and those features can then be employed.

3.2.1 Feature Extraction of MobileNetV2

A convolutional neural network, MobileNetV2 was designed specifically for use on mobile devices. The intermediate layer of this network's 19-layer structure is responsible for differentiating between types of attributes, while the final layer is responsible for classifying them. It has a 94.38% success rate in dividing 8,000 images into two groups. Parameters for N channels may be converted to those for M channels using the s step and the t expansion factor, as shown in [28]. This bottleneck uses a 1 folded layer in front of the deep folded layer and the linear activation function after the punctual folded layer to accomplish downsampling. The general layout of MobileNetV2's network is shown in Table [28], where conv2d denotes the standard convolution, avg pool denotes average pooling, c denotes the number of output channels and n is the number of iterations. Image classification in Breast Histopathology requires retraining MobileNetV2. The pictures are in sequence because they have to be 224 by 224 by 3. Data-specific layers have been added to the component that previously classified RGB pictures into two categories. The maximum number of epochs was set at 25, while the initial learn rate was set to 104. Images in the RGB color space are the input. Just 3,000 of the 8,000 photographs were chosen at random to serve as test data, while the remaining 6,400 were used for training.

3.2.2 Feature Extraction with VGG19

The object identification and classification method based on CNN architecture VGG19 had an accuracy rate of 95.65% while classifying 8000 images into two different groups (Benign and Malignant). The widely used and uncomplicated VGG19 method uses a transfer learning approach to determine the features of images. This network received a fixed-size (224 x 224) images as input, indicating that the matrix was shaped (224, 224, 3) with a learning rate of $1 * 10^{-4}$ [55]. In order to extract single-frame-based cues, we extract a fixed 512-dimension feature for each frame at a rate of 25 frames per second using a VGG19 network that has already been trained to identify whether benign or cancerous cells are present. Malignant cells are detected as one if they are present in the histopathological imaging whereas benign cells are detected as zero (negative).

In order to categorize breast histopathology photos and develop the presence of cancer cells using vision sensors, the VGG19 neural network model will be deployed. It functions essentially as a CNN feature extractor that is run across completely linked

layers.

Table 3.2.1: The Summary of VGG19 Model

Layer (type)	Output Shape	Param #
input_4 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 28, 28, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 512)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_average_pooling2d_2 (GlobalAveragePooling2D)	(None, 512)	0
dropout_2 (Dropout)	(None, 512)	0
batch_normalization_2 (Batch_normalization)	(None, 512)	2048
dense_2 (Dense)	(None, 1)	513
Total params	20,026,945	
Trainable params	20,025,921	
Non-trainable params	1,024	

The VGG19 extracts features from the frames with the greatest detection score, providing us with both frame-level and object-level data.

3.2.3 Feature Extraction with ResNet50V2

The ResNet50V2 basic feature extractor is another basic feature extractor that was compared in this research. It was started using parameters that were pre-trained on ImageNet [16]. Analyses are performed using relative nets, which have a maximum depth level of 152 layers and are eight times deeper than VGG nets. However, relative nets are easier to train on the ImageNet dataset. While dividing 8000 photos into two distinct categories, ResNet50V2 achieved an accuracy percentage of 98.56%. (Benign and Malignant). Once the ImageNet classification models have started the model, the BreakHis dataset is utilized to further improve and perfect the model. When training networks using ResNet50V2 or a residual learning architecture, it is simple to train networks that are far bigger than those that were previously utilized. The feature map $A \in R^{K*U*V}$ was derived with the assistance of the ResNet50V2 that was employed. To put it another way, the feature map has K channels, each

of which has a height value of U and a width value of V . This feature map may be used to generate a D -dimensional feature vector which can be represented as $x \in \mathbb{R}^D$ after it has been flattened by a thick layer. The bottlenecks in this model need demodulation to be performed. The pre-trained ResNet50V2 model anticipates receiving brief batches of normalized input photos with such a three-channel RGB dimension ($3 * H * W$), where the values 224 are used to represent both H and W . Once the images have been imported between $[0$ and $1]$, they need to be normalized by applying mean values of $[0.485, 0.456, 0.406]$ and standard values of $[0.229, 0.224, 0.225]$. First, the weights are set up according to [15] and then the models are trained using learning rates beginning at $1 * 10^{-4}$. Once the error has plateaued, the learning rates are divided by 10 and then the weights are reset. The architecture is described in the following paper [32].

3.3 Breast Cancer Prediction

In the proposed breast cancer prediction method, we fused concatenation and ensemble method in order to fuse the extracted features and use those features in several linear classifiers. A linear classifier is a form of supervised machine learning algorithm that generates a prediction by using a linear combination of the input data. Linear classifiers are one of the most common types of machine learning algorithms. Some of the classifiers used in this method include:

3.3.1 Feature Fusion

Concatenation and ensemble methods are two commonly used techniques for feature fusion in image processing.

Concatenation is a simple method for combining features from multiple sources by concatenating them into a single feature vector. This is done by appending the features from one source to the features from another source to form a new feature vector. As a result, the new feature vector will have more features than the original feature vectors, which may boost the classifier's accuracy.

Ensemble method, on the other hand, is a method for combining multiple models to improve the performance of the classifier. It is based on the idea that multiple models trained on the same data can produce different results and by combining the results of these models, the final prediction can be more accurate. Two broad classes of ensemble techniques exist: averaging and boosting. Boosting methods combine the predictions of many models by providing more weight to the models that perform better, whereas averaging methods take the mean or median of the forecasts.

Both concatenation and ensemble methods can be used to improve the performance of classifiers by combining the features or predictions of multiple models. Nevertheless, the problem at hand and the facts at hand will dictate the solution ultimately chosen. Concatenation can be useful when the features from multiple sources are independent and have similar scales, while ensemble methods can be more effective when the features from multiple sources are correlated and have different scales.

3.3.2 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is a non-parametric and instance-based machine learning method utilized for both classification and regression tasks. The core concept of KNN involves identifying the k-closest training examples to a new sample and determining its classification or value based on the majority vote or mean of its k-nearest neighbors. The implementation of this method is straightforward and it makes no presumptions about the data's underlying probability distribution in order to function properly. The only necessary parameter is the value of k which defines the number of closest training samples to take into account for prediction.

KNN algorithm is composed of two steps: distance calculation. When a distance metric, such as Euclidean, Manhattan, or Minkowski, is used to determine how far apart a new sample is from the training examples and K-Nearest Neighbors selects the k-closest training instances to utilize for prediction. The prediction in classification is done by majority voting and in regression, by averaging the values of the k-nearest neighbors. It is possible to boost the algorithm's efficiency by The performance of the algorithm can be improved by selecting the most relevant features and normalizing them to the same scale. However, KNN algorithm is computationally expensive for large datasets and may not perform well in high dimensional data as the distance metric becomes less reliable in such spaces.

3.3.3 Stochastic Gradient Descent (SGD)

Most large-scale classification issues are tackled with Stochastic Gradient Descent (SGD), a linear classifier technique. An algorithm for maximizing efficiency that is used to find the optimal set of parameters for a given model. Standard Gradient Descent (SGD) is based on the premise that the model parameters should be updated in an iterative fashion in a direction that is opposite to the negative gradient of the loss function relative to the existing parameter values. The algorithm is called "stochastic" because it uses a random sample of the data to estimate the gradient of the loss function in each iteration.

The basic steps of the SGD algorithm are:

1. Initialize the model parameters with random values.
2. Shuffle the training data.
3. For each training sample, compute the gradient of the loss function with respect to the model parameters.
4. Update the model parameters in the direction of the negative gradient.
5. Iterate through steps 2-4 until the model converges or for a predetermined number of times.

The main advantage of SGD is its scalability, it can handle large datasets with millions of examples and features and it can be easily parallelized. However, it is sensitive to the choice of the learning rate and initial values of the parameters and it may require multiple runs with different settings to get the best results. Additionally,

it may not converge to the global minimum of the loss function and it can get stuck in a local minimum, specially in non-convex loss function.

3.3.4 Adaptive Boosting (AdaBoost)

AdaBoost (Adaptive Boosting) is an ensemble method that iteratively improves the performance of a weak classifier by giving more weight to the misclassified samples. Each iteration adds a new weak classifier to the mix, which is trained on the weighted samples and then used in conjunction with the predictions of the prior iterations to arrive at a final prediction. The weights of the samples are adjusted in each iteration to give more importance to the misclassified samples. AdaBoost is a fast and effective algorithm that has been widely used in various applications, including image processing and computer vision.

AdaBoost is an iterative machine learning method. The weights of incorrectly categorized samples rise with each iteration while the weights of successfully classified samples fall. This type of weight updating is crucial since it allows AdaBoost to concentrate on classifying challenging data in order to reduce training error. The following is an algorithm of the generalized AdaBoost:

Algorithm :

Initialization:

1. Given training data from the instance space

$$S = \{(x_1, y_1), \dots, (x_m, y_m)\} \quad (3.1)$$

where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y} = \{-1, +1\}$

2. Initialize the distribution

$$D_1(i) = \frac{1}{m} \quad (3.2)$$

Algorithm:

for $t = 1, \dots, T$: **do**

Train a weak learner $h_t : \mathcal{X} \rightarrow \mathbb{R}$ using distribution D_t

Determine weight α_t of h_t

Update the distribution over the training set:

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_t y_t h_t(x_i)}}{Z_t} \quad (3.3)$$

where Z_t is a normalization factor chosen so that D_{t+1} will be a distribution.

end for

Final score:

$$f(x) = \sum_{t=0}^T \alpha_t h_t(x) \quad (3.4)$$

$$H(x) = \text{sign}(f(x)) \quad (3.5)$$

3.3.5 Extreme Gradient Boosting (XGBoost)

In this thesis, we employed XGBoost as one of the linear classifiers to classify the fused features obtained from our proposed feature fusion method. XGBoost, short for Extreme Gradient Boosting, is an implementation of gradient boosting algorithm that is widely used in machine learning and data science competitions due to its high performance and efficiency. It is a form of ensemble learning that produces a robust model by combining the predictions of several other, less accurate models.

The basic idea behind XGBoost is to iteratively add weak models to the ensemble, with each new model aiming to correct the mistakes of the previous models. This is achieved by training the new model on the residual errors of the previous models. The final ensemble is generated by adding all of the unreliable models together and weighting each one, where the weights are learned by minimizing a loss function.

XGBoost uses decision trees as the weak models, and it has several techniques to improve the performance of the decision trees such as regularization, sparsity control, and tree pruning. It also has a built-in mechanism for handling missing values and categorical variables, which makes it suitable for handling large datasets with a mix of categorical and numerical variables.

In our proposed method, we used XGBoost as a classifier to classify the fused features obtained from the feature fusion method. The fused features were created by concatenating the features extracted from the last layer of the trained CNN models. We found that XGBoost achieved an accuracy of 96.8% in classifying the fused features, which is a good performance when compared to other linear classifiers.

However, it should be noted that XGBoost is a powerful algorithm but it can be computationally expensive and may require fine-tuning to achieve optimal performance. Another limitation is that it is possible for it to be sensitive to noise in the data as well as outliers. Despite these limitations, in our proposed method, we found that XGBoost performed well and contributed to the overall high accuracy of our breast cancer prediction system.

3.3.6 Decision Tree

As part of the breast cancer prediction work that we did for our thesis, we employed the usage of a Decision Tree as one of the linear classifiers. The tasks of classification and regression are frequent applications for decision trees, which are a form of supervised learning algorithm that are utilized extensively. They do this by recursively splitting the feature space into smaller subsets, often known as leaves, which include observations that are analogous to one another. At each stage of the partitioning process, the algorithm chooses the characteristic to focus on as well as the threshold value that will produce the greatest amount of information gain. The information gain is a measurement of how successfully the feature divides the observations into the various target classes or values. The process of partitioning continues until one of the stopping criteria such as a maximum tree depth or a minimum number of samples per leaf, is satisfied. This might be the case.

Once the Decision Tree is trained, the tree structure can be used to make predictions

by traversing the tree from the root to a leaf. At each internal node, the algorithm compares the value of the selected feature to the threshold value, and chooses the next node to visit based on the comparison. Once a leaf is reached, the algorithm assigns the class or the target value associated with that leaf to the input instance.

In our thesis, we used the Decision Tree classifier to classify the fused features extracted from the last layer of the trained models. We compared its performance with other linear classifiers such as K-Nearest Neighbors, Stochastic Gradient Descent, AdaBoost and Random Forest.

3.3.7 Random Forest

To categorize the fused features we derived from the trained base models, we included the Random Forest classifier in our thesis as one of the linear classifiers. A method for improving classification accuracy by combining many decision trees, Random Forest is an ensemble method.

When training, the algorithm generates numerous decision trees, each of which is trained on a different sample of data and with a different collection of characteristics. This practice known as bootstrapping or bagging, is intended to increase accuracy. All of the decision trees' outputs are averaged to arrive at a single forecast. By generating a diversified group of trees that are less susceptible to slight variations in the data, this method helps to mitigate the overfitting problem prevalent in decision trees.

Random Forest also performs a random subset selection of features at each split point in a decision tree, this technique is known as random subspace method, this helps to decorrelate the trees and make them less dependent on any one feature.

In our thesis, we used the Random Forest classifier to classify the fused features obtained from the last layer of the trained base models.

3.3.8 Logistic Regression

Logistic Regression is a type of linear classifier that is used to predict the probability of a binary outcome, such as a positive or negative class label, based on one or more input features. Logistic Regression is used to predict the probability of a binary outcome, such as a positive or negative class label. Using a logistic function, which is sometimes referred to as the sigmoid function, the algorithm creates a model of the relationship that exists between the input characteristics and the binary output.

The first step in the Logistic Regression technique is to define a linear function of the input characteristics. This function is also known as the logit and it is then fed into the sigmoid function in order to generate a probability value that is between 0 and 1. After that, a threshold may be applied to the likelihood to provide a yes or no forecast. The following is a definition of the logit:

$$\text{logit} = w_0 + w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n \quad (3.6)$$

where w_0, w_1, \dots, w_n are the model weights and x_1, x_2, \dots, x_n are the input features. Here is how we characterize the sigmoid function:

$$p = \frac{1}{1 + e^{-\text{logit}}} \quad (3.7)$$

The goal of the Logistic Regression algorithm is to find the best values for the weights w_0, w_1, \dots, w_n that maximize the likelihood of the observed data. Most of the time, this is done with maximum likelihood estimation, which is an optimization technique that finds the weights that make the likelihood function as high as possible.

Once the model has been trained, it can be used to predict the likelihood of a binary outcome for new input features. The value of the probability can then be used as a threshold to make a yes or no prediction. For example, if the probability is more than 0.5, the sample is considered positive and if it is less than 0.5, the sample is considered negative.

The Logistic Regression algorithm is a simple, easy-to-use algorithm that is used in many fields such as medicine, natural language processing and image classification.

These classifiers were chosen because they are good at classifying images and can be used in different ways with the fused feature space. Metrics like accuracy, precision, recall and F1-score were used to judge how well the classifiers worked. Then, the results of each classifier were compared to see which one did the best job.

Chapter 4

Implementation and Result

4.1 Implementation

In this thesis, three base models (ResNet50V2, VGG19 and MobileNetV2) were utilized to construct a fusion model. Machine learning methods such as SVM, AdaBoost and XGBoost were employed for predictions. Additionally, the Grad-CAM technique was employed to visualize the model's prediction process. The implementation process involved preprocessing input data, extracting features from the final layers of the base models and aggregating these features to generate the fusion model. In this part, the results of the feature extraction and prediction implementation are discussed.

4.1.1 Dataset Preparation

The Breast Cancer Histopathological Image Classification dataset, also known as BreakHis dataset, consists of 9,109 microscopic images of breast tumor tissue obtained from 82 patients using various magnifying factors (40X, 100X, 200X and 400X). These images are divided into two main categories: benign tumors and malignant tumors. The dataset includes 2,480 benign samples and 5,429 malignant samples. Each image has a resolution of 700 X 460 pixels, with 3 channels in RGB format and 8-bit depth per channel and is in PNG format.

4.1.1.1 Slicing

From the BreakHis dataset, we have taken the 40X zoomed slices from an image where each image's width is 700 and height is 460. The size of each slice is 50 x 50. After that 50,000 random Benign and 50,000 random malignant image slices were selected. Among the 50,000 random class images, 10000 were randomly selected for testing.

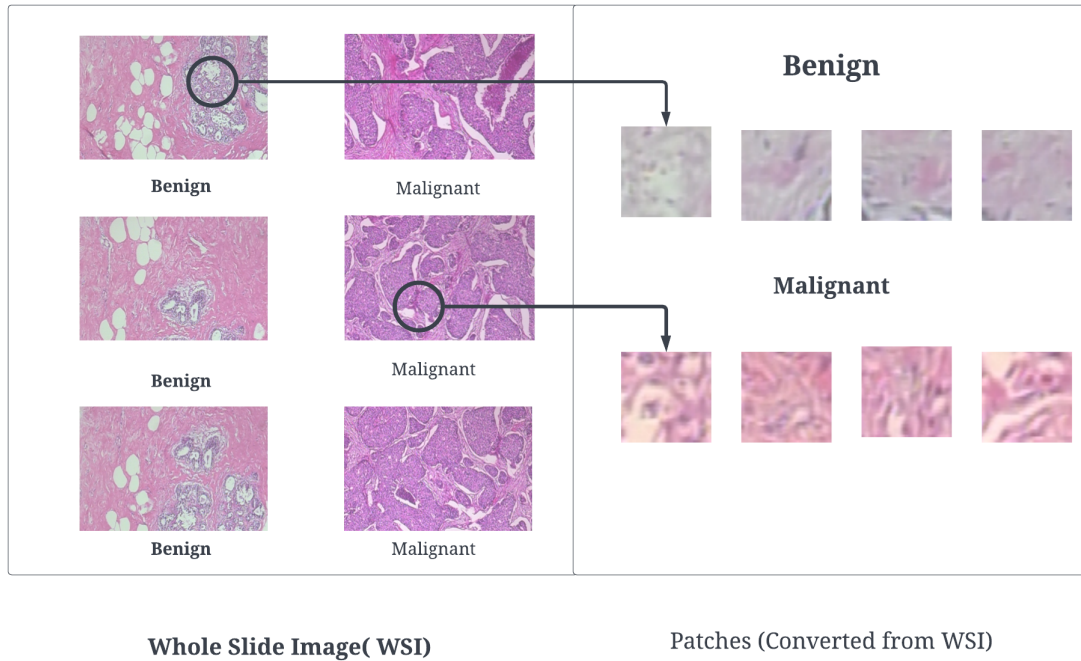


Figure 4.1.1: Data Augmentation

4.1.1.2 Data Preprocessing and Classification

To preprocess the BreakHis dataset, we resize the frames into 224 x 224 x 3 dimensions. The data extracted is then annotated and classified into Benign and Malignant classes.

4.1.1.3 Dataset Splitting

The dataset is divided into training and testing sets in an 80:20 proportion.

4.1.2 Fusion Model

4.1.2.1 Model Build Function

A separate function was defined that takes a backbone as the base model and aggregates the mentioned layers with it to start training. We have trained 3 models individually with the same dataset but with different base models as the backbone namely ResNet50V2, VGG19 and MobileNetV2. In each model, we have used additional layers such as Global Average Pooling, Dropout, Batch Normalization and Dense layer. For the Dense layer, 'sigmoid' was used as the activation function and 'imagenet' was used as the weight parameter. For model compilation, 'binary cross entropy' was used for loss and 'Adam' as the optimizer.

4.1.2.2 Feature Extraction

After training the ResNet50V2, VGG19 and MobileNetV2 models, we have extracted the features of those models. To extract the feature, we have loaded the .h5 files

and used `model.layers[-2].output` to eliminate the last Dense layer from each model and proceed with individual training and test dataset to generate predictions. After that, all of them were concatenated as train and test features. The extracted features were saved in `.npy` formats. Labels were also created consisting of 40,000 0s and 1s for training and 10,000 0s and 1s for tests.

4.1.2.3 Feature Fusion

In our paper, we used the concatenation technique to fuse the three extracted features from the base model. Feature fusion using concatenation is a technique used to combine multiple sets of features into a single feature representation. This technique is often used in image classification tasks, where multiple feature extractors are applied to an image in order to extract features from it. These are then concatenated together to form a single, more comprehensive feature representation. The advantage of using feature fusion is that it allows for the combination of multiple types of information from the image, which can lead to a more robust and accurate classification. Additionally, it allows for the use of multiple feature extractors, which can be trained with different architectures or on different subsets of the data, leading to a more diverse set of features that can be used for classification. It also helps to mitigate the problem of overfitting. The method is simple and efficient, it just requires a simple concatenation of different feature vectors along a certain axis.

4.1.2.4 Fusion Model Creation

In our paper, we used the concatenation technique to fuse the three extracted features from the base model. Feature fusion using concatenation is a technique used to combine multiple sets of features into a single feature representation. This technique is often used in image classification tasks, where multiple feature extractors are applied to an image in order to extract features from it. These are then concatenated together to form a single, more comprehensive feature representation. The advantage of using feature fusion is that it allows for the combination of multiple types of information from the image, which can lead to a more robust and accurate classification. Additionally, it allows for the use of multiple feature extractors, which can be trained with different architectures or on different subsets of the data, leading to a more diverse set of features that can be used for classification. It also helps to mitigate the problem of overfitting. The method is simple and efficient, it just requires a simple concatenation of different feature vectors along a certain axis.

4.1.3 GradCAM

Grad-CAM (Gradient-weighted Class Activation Mapping) is a method for producing a class-specific heatmap of an image which highlights the regions of the image that are most important for the classification of given class. The goal of Grad-CAM is to identify the region of interest (ROI) from an image. This is a crucial part for making a prediction, so that the model can focus on these regions when making the prediction.

The main objective of Grad-CAM is to utilize the gradients taken from the output of the final convolutional layer of a CNN with respect to the input image to compute

the importance of each feature map in the final layer. These gradients are then used to weight the feature maps, producing a heatmap that highlights the regions of the image that are most important for the classification of a given class.

For our pretrained model (VGG19, Resnet50V2, MobileNetV2) we gave WSI image as a whole and generated patches to detect and predict individual patches and then join the ROI into one single heatmap.

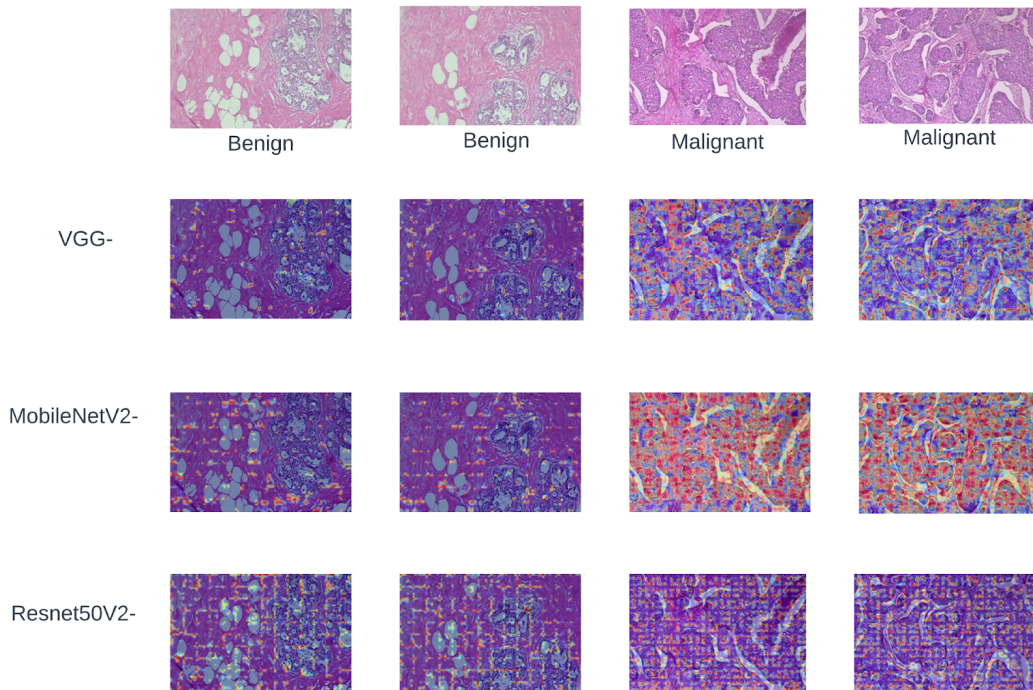


Figure 4.1.2: Grad-CAM Heatmap of VGG19, MobileNetV2, ResNet50V2 where the Red Region Indicates the Region with High Activations and Blue Means Low Activation

4.2 Result

4.2.1 Base Model Performance

In our proposed system, we employed three base models, VGG19, ResNet50V2 and MobileNetV2, which were used for training on the BreakHis dataset. The accuracy of these models were 95.65%, 94.38%, 94.86% respectively. These results compare favourably with other recent works in the field. For example, the proposed system used Image Level Accuracy (ILA) as the metric and using Whole Slide Image (WSI). The comparison study's findings are summarised in Table 4.2.1 where we see [24] had an accuracy of 90.96% using AlexNet as the base model and ImageNet as the weights. Moreover, using InceptionV3 [21] got an accuracy of 85.9%. These results demonstrate the advantage of our proposed method displayed in Table 4.2.2 in accurately classifying breast cancer in histopathological images.

Table 4.2.1: Existing Models used On the BreakHis Dataset

Work	Preprocessing	Patch/Slide	Feature Extractor	Transfer Learning	Training/Test	Metric	Result%
[18]	-Res(350x230) -SMI -DA(Rot, Flip)	WSI	NDCNN	None	70%/30%	ILA	77.5
[21]	-SMI	WSI	InceptionV3	ImageNet	70%/30%	ILA	85.9
[24]	-Res(227x227)	WSI	AlexNet	ImageNet	Not specified	ILA	90.96

Table 4.2.2: Our Models on the BreakHis Dataset

Preprocessing	Patch/Slide	Feature Extractor	Transfer Learning	Training/Test	Metric	Result%
-Res(224x224x3) -Slicing	Patches(Converted from WSI)	VGG19	ImageNet	80%/20%	ILA	95.65
		ResNet50V2				94.38
		MobileNetV2				94.86

4.2.2 Fusion Model Performance

After we fused the extracted feature from our three base models we found that the two best fused models (extracted feature and classifiers) were KNN classifier(K-Nearest Neighbour) and RFC (Random Forest Classifier) which are highlighted in Table 4.2.3. In the tables below, we compare our fused model with some of the best performing fused models by other authors.

Table 4.2.3: Existing Fusion Model used On the BreakHis Dataset

Work	Preprocessing	Patch/Slide	Feature Extractor	Classifier	Transfer Learning	Training/Test	Metric	Result%
[22]	-Res(224 224) -GSC	WSI	VGG	NN	ImageNet	75%/25%	ILA	84.0
[39]	-Res(224x224) -DA(CROP, ANP, FLIP)	WSI	ResNet-152, GoogleNet	SVM	ImageNet	80%/20%	ILA	81.2 ± 2.5
[20]	None	WSI	LPQ	SVM	None	70%/30%	ILA	91.1

Table 4.2.4: Our Fusion Model used On the BreakHis Dataset

Preprocessing	Patch/Slide	Feature Extractor	Classifier	Transfer Learning	Training/Test	Metric	Result%
-Res(224x224x3) -Slicing	Patches(Converted from WSI)	VGG19, ResNet50V2, MobileNetV2	KNN	ImageNet	80%/20%	ILA	97.535
			RFC				97.455

According to research done by Badejo et al. [20] , Local Phase Quantization (LPQ) fused with SVM gave an accuracy of 91.1% . Our models gives a better reading of the cancer localization in the BreakHis Histopathology Images.

4.2.3 Performance Metrics

4.2.3.1 Confusion Matrix

Confusion matrix is a tool by which the performance of a classification model can be assessed. It displays a table showing the predicted class and actual class for each test sample. The matrix is separated into four parts: true positives, true negatives, false positives and false negatives. These counts can then be employed to calculate various evaluation metrics such as accuracy, precision, recall and F1-score. This way we can have a clear understanding of how well your model is performing and identify where it may need improvement. Confusion matrix is calculated to validate our model which also provides Sensitivity and Specificity score.

Sensitivity and Specificity are two measures used to evaluate the performance of a classification model, particularly in the context of medical diagnosis or binary classification problems.

Assuming, True Positives = T_P , False Negatives = F_N , True Negatives = T_N , False Positives = F_P .

$$\text{Sensitivity (also known as true positive rate or recall)} = \frac{T_P}{T_P + F_N} \quad (4.1)$$

It tells us what proportion of actual positive cases are correctly identified by the model.

$$\text{Specificity (also known as true negative rate)} = \frac{T_N}{T_N + F_P} \quad (4.2)$$

It indicates the proportion of actual negative cases that are correctly identified by the model.

Both Sensitivity and Specificity provide different insights into how well the model is able to identify positive and negative cases, respectively. A high sensitivity indicates that the model is effective at identifying positive cases, whereas high specificity implies that the model is efficient at identifying negative cases.

The confusion matrix of the models used are as follows:

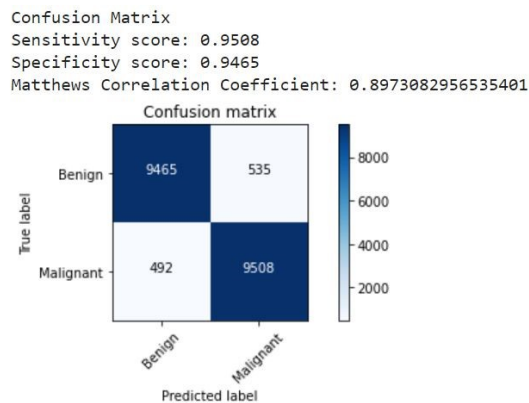


Figure 4.2.1: MobileNetV2

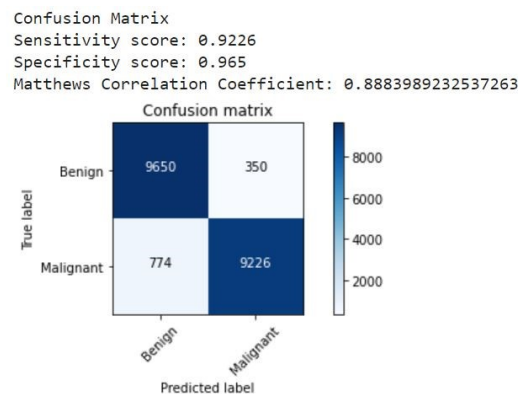


Figure 4.2.2: ResNet50V2

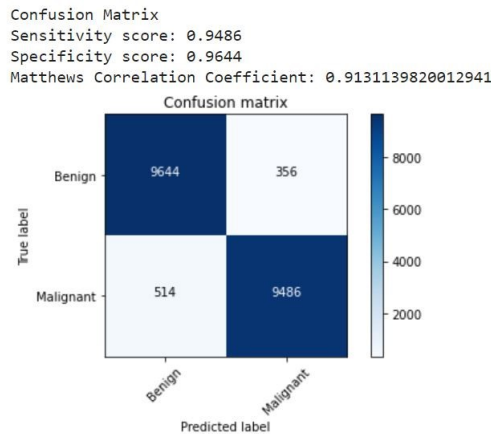


Figure 4.2.3: VGG19

The confusion matrix of the classifiers used are as follows:

SVM: Here from the confusion matrix, we found the model successfully detected benign images 9719 times but predicted it wrong for 281 images. On the other hand, it detected malignant type for 9697 images correctly and predicted wrong for 303 images.

AdaBoost: Here from the confusion matrix, we found the model successfully detected benign images 9714 times but predicted it wrong for 286 images. On the other hand, it detected malignant type for 9664 images correctly and predicted wrong for 336 images.

XGBoost: Here from the confusion matrix, we found the model successfully detected benign images 9753 times but predicted it wrong for 247 images. On the other hand, it detected malignant type for 9719 images correctly and predicted wrong for 281 images.

Logistic Regression: Here from the confusion matrix, we found the model successfully detected benign images 9738 times but predicted it wrong for 262 images. On the other hand, it detected malignant type for 9716 images correctly and predicted wrong for 284 images.

Random Forest: Here from the confusion matrix, we found the model successfully detected benign images 9776 times but predicted it wrong for 224 images. On the other hand, it detected malignant type for 9715 images correctly and predicted wrong for 285 images.

Decision Tree: Here from the confusion matrix, we found the model successfully detected benign images 9627 times but predicted it wrong for 373 images. On the other hand, it detected malignant type for 9607 images correctly and predicted wrong for 393 images.

SDG: Here from the confusion matrix, we found the model successfully detected benign images 9706 times but predicted it wrong for 294 images. On the other hand, it detected malignant type for 9731 images correctly and predicted wrong for 269

images.

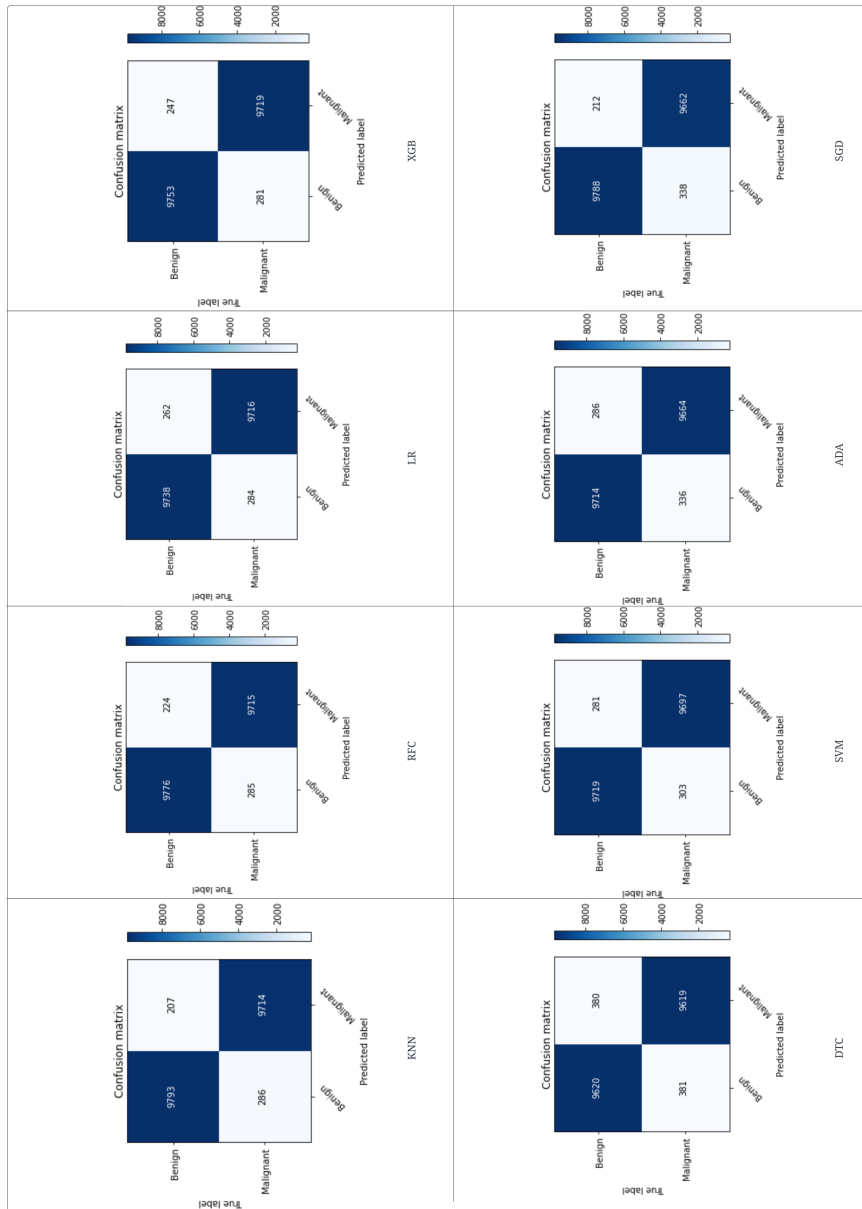


Figure 4.2.4: Confusion Matrix of All the Proposed Classifier Used on the BreakHis Dataset

KNN: Here from the confusion matrix, we found the model successfully detected benign images 9793 times but predicted it wrong for 207 images. On the other hand, it detected malignant type for 9714 images correctly and predicted wrong for 286 images.

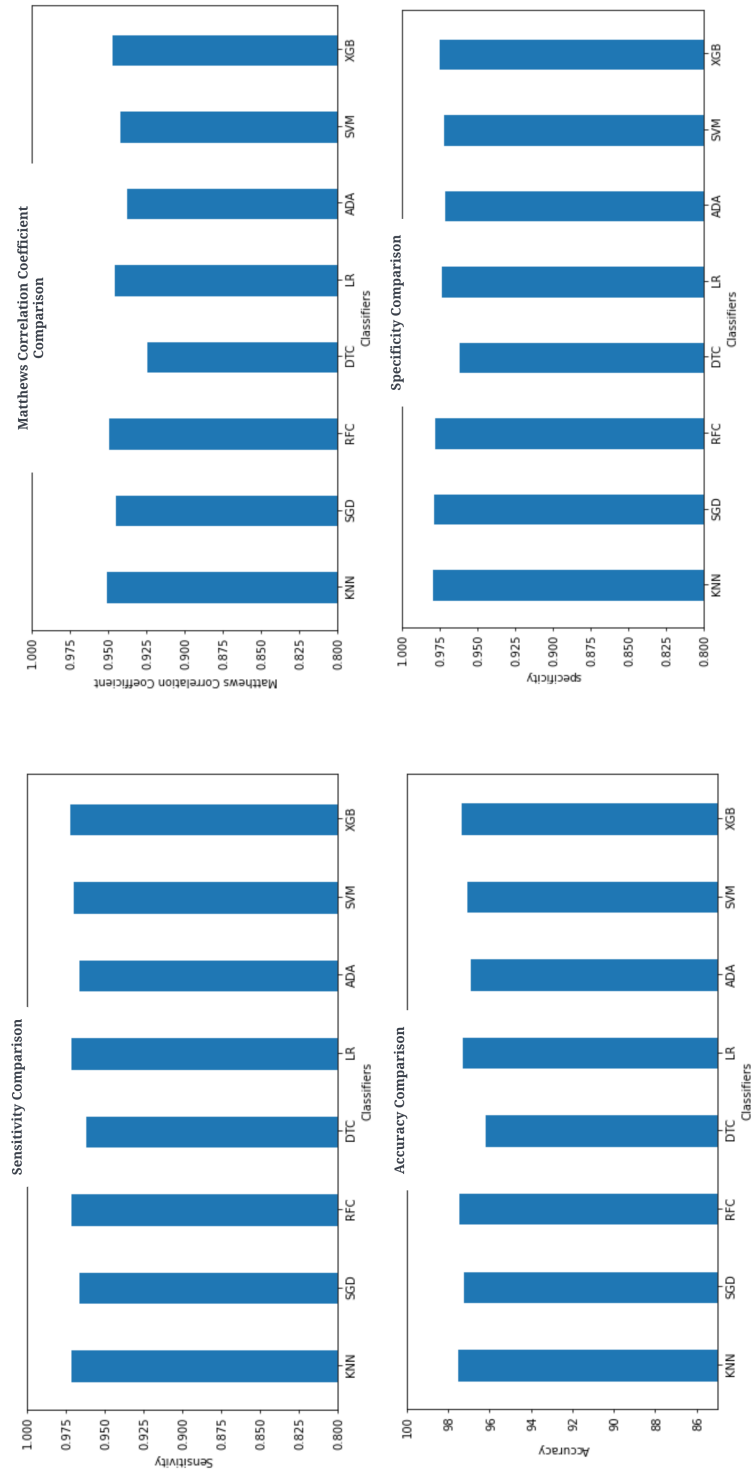


Figure 4.2.5: Comparisons of Accuracy, Sensitivity, Specificity and Matthews Correlation Coefficient of All the proposed Classifiers

Figure 4.2.5 demonstrates about the comparison of 4 matrices, accuracy, sensitivity, specificity and Matthews correlation coefficient, among the 8 classifiers we used in our research.

4.2.4 Classification Report

A classification report is a method to evaluate the performance of a classification model by showcasing various evaluation metrics like precision, recall and F1-score.

Precision is a metric that measures the proportion of instances that were correctly predicted as positive by the model out of all the instances that were predicted as positive by the model. It is a measure of the accuracy of the model in identifying positive cases, and it is determined using :

Assuming, True Positives = T_P , False Negatives = F_N , True Negatives = T_N , False Positives = F_P .

$$Precision, Pr = \frac{T_P}{T_P + F_P} \quad (4.3)$$

Recall is a metric that measures the proportion of actual positive instances that were correctly predicted as positive by the model. It is a measure of the effectiveness of the model in identifying all actual positive cases and it is calculated as the proportion of true positive predictions made by the model out of all the actual positive instances. It is determined using:

$$Recall, Rc = \frac{T_P}{T_P + F_N} \quad (4.4)$$

F1-score is a measure of the balance between precision and recall. It is calculated as the harmonic mean of precision and recall and is a way of combining the two metrics into a single value. F1-score is calculated as:

$$F1_{score} = 2 * \left[\frac{Pr * Rc}{Pr + Rc} \right] \quad (4.5)$$

In breast cancer classification, precision and recall are important metrics to consider because a high precision means that a high proportion of the positive predictions detected by the model are actually true positives, which is essential for reducing false alarms. Recall is also important because if a model outputs a high recall, then it indicates it can identify a high proportion of actual positive instances, which is essential for reducing missed detections.

In breast cancer classification, F1-score is important because it helps in balancing the trade-off between Precision and Recall. It ensures that the model has a good balance of identifying a high proportion of true positive instances while also keeping the number of false alarms low.

Table 4.2.5: Classification Report of the Classifiers

Classifier		KNN	DTC	LR	RFC	SGD	AdaBoost	SVM	XGBoost
Precision	Benign	97	96	97	97	97	97	97	97
	Malignant	98	96	97	98	98	97	97	98
Recall	Benign	98	96	97	98	98	97	97	98
	Malignant	97	96	97	97	97	97	97	97
F1-Score	Benign	98	96	97	97	97	97	97	97
	Malignant	98	96	97	97	97	97	97	97
Macro avg accuracy		98	96	97	97	97	97	97	97
Weighted avg accuracy		98	96	97	97	97	97	97	97
Accuracy%		97.535	96.195	97.27	97.455	97.25	96.89	97.08	97.36

4.2.5 Discussion

The K-Nearest-Neighbors (KNN) technique is both easy to implement and very reliable, making it useful in a wide variety of contexts. The algorithm makes predictions based on how closely the input features of the test cases match those of the training examples. When there is a lot of noise in the data and the border between classes is unclear, KNN can be an effective method. In addition, KNN is less vulnerable to the impacts of outliers and other types of noise since it makes no assumptions about the underlying distribution of the data. It's also possible that KNN performed better than alternative algorithms because of the complicated decision boundaries and high noise levels in our dataset.

While XGBoost is a robust method, it may be underperforming KNN since our dataset is too small. Only about 100,000 images are used in our training and testing collection. In most cases, XGBoost's best results can only be achieved with a sizable amount of training data. It may not have learned the intricate connections between the characteristics and the target variable as well as KNN if given a smaller dataset. Furthermore, a simpler technique, such as KNN, could have been more appropriate given the data and the nature of the problem at hand.

Chapter 5

Conclusion

In conclusion, the incidence of breast cancer among women is a significant concern that requires prompt attention. For patients to have a better chance of surviving breast cancer, early diagnosis is essential. With the advancement of technology, there is a growing need for automated methods for breast cancer diagnosis to improve the speed and accuracy of diagnosis. Our proposed method, which utilizes deep learning techniques, aims to address this need.

We employed a transfer learning approach using the publicly available BreakHis dataset of breast histopathological images, fine-tuning three base models, VGG19, ResNet50V2 and MobileNetV2 with the imagenet weights. Our experimental results revealed that the suggested approach achieved a high level of accuracy, with the KNN classifier achieving the best result of 97.535% and the Random Forest Classifier achieving 97.455%. These results demonstrate that our suggested method performs well in breast cancer prediction and may support pathologists in making a breast cancer diagnosis.

Grad-CAM was also used to enhance the clinical interpretability of the suggested models and offer a visual description of the the method for making decisions. As a result, pathologists may be more inclined to view computerized deep learning techniques as reliable and legitimate diagnostic aids for breast cancer.

Our research serves as a meaningful step forward in the rapidly expanding field of artificial intelligence in healthcare, specifically in the area of breast cancer detection. We have a strong conviction that our suggested approach will be an invaluable tool in the early diagnosis of breast cancer and eventually help to save many lives.

References

- [1] G. Carter, C. Knapp, and A. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast fourier transform processing," *IEEE transactions on audio and electroacoustics*, vol. 21, no. 4, pp. 337–344, 1973.
- [2] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Image analysis and machine learning applied to breast cancer diagnosis and prognosis," *Analytical and Quantitative cytology and histology*, vol. 17, no. 2, pp. 77–87, 1995.
- [3] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [4] O. Edfors, M. Sandell, J.-J. Van de Beek, S. K. Wilson, and P. O. Borjesson, "Ofdm channel estimation by singular value decomposition," *IEEE Transactions on communications*, vol. 46, no. 7, pp. 931–939, 1998.
- [5] D. Cascio, F. Fauci, R. Magro, *et al.*, "Mammogram segmentation by contour searching and mass lesions classification with neural network," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2827–2833, 2006.
- [6] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer informatics*, vol. 2, p. 117693510600200030, 2006.
- [7] J. J. Dignam, L. Huang, L. Ries, M. Reichman, A. Mariotto, and E. Feuer, "Estimating breast cancer-specific and other-cause mortality in clinical trial and population-based cancer registry cohorts," *Cancer*, vol. 115, no. 22, pp. 5272–5283, 2009.
- [8] H. M. Heneghan, N. Miller, and M. J. Kerin, "Mirnas as biomarkers and therapeutic targets in cancer," *Current opinion in pharmacology*, vol. 10, no. 5, pp. 543–550, 2010.
- [9] B. Xu, Y. Ye, and L. Nie, "An improved random forest classifier for image classification," in *2012 IEEE International Conference on Information and Automation*, IEEE, 2012, pp. 795–800.
- [10] K. Zen and C.-Y. Zhang, "Circulating micrnas: A novel class of biomarkers to diagnose and monitor human cancers," *Medicinal research reviews*, vol. 32, no. 2, pp. 326–348, 2012.
- [11] D. Madhavan, K. Cuk, B. Burwinkel, and R. Yang, "Cancer diagnosis and prognosis decoded by blood-based circulating microrna signatures," *Frontiers in genetics*, vol. 4, p. 116, 2013.

- [12] O. Fortunato, M. Boeri, C. Verri, *et al.*, “Assessment of circulating micrnas in plasma of lung cancer patients,” *Molecules*, vol. 19, no. 3, pp. 3038–3054, 2014.
- [13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [14] M. Veta, J. P. Pluim, P. J. Van Diest, and M. A. Viergever, “Breast cancer histopathology image analysis: A review,” *IEEE transactions on biomedical engineering*, vol. 61, no. 5, pp. 1400–1411, 2014.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [17] P. Giri and K. Saravanakumar, “Breast cancer detection using image processing techniques,” *Oriental journal of computer science and technology*, vol. 10, no. 2, pp. 391–399, 2017.
- [18] E. M. Nejad, L. S. Affendey, R. B. Latip, and I. Bin Ishak, “Classification of histopathology images of breast into benign and malignant using a single-layer convolutional neural network,” in *Proceedings of the International Conference on Imaging, Signal Processing and Communication*, 2017, pp. 50–53.
- [19] X. Ren, H. Guo, S. Li, S. Wang, and J. Li, “A novel image classification method with cnn-xgboost model,” in *International Workshop on Digital Watermarking*, Springer, 2017, pp. 378–390.
- [20] J. A. Badejo, E. Adetiba, A. Akinrinmade, and M. B. Akanle, “Medical image classification with hand-designed or machine-designed texture descriptors: A performance evaluation,” in *international conference on bioinformatics and biomedical engineering*, Springer, 2018, pp. 266–275.
- [21] Y. Benhammou, S. Tabik, B. Achchab, and F. Herrera, “A first study exploring the performance of the state-of-the art cnn model in the problem of breast cancer,” in *proceedings of the international conference on learning and optimization algorithms: theory and applications*, 2018, pp. 1–6.
- [22] S. Cascianelli, R. Bello-Cerezo, F. Bianconi, *et al.*, “Dimensionality reduction strategies for cnn-based classification of histopathological images,” in *International conference on intelligent interactive multimedia systems and services*, Springer, 2018, pp. 21–30.
- [23] R. Dass, “Speckle noise reduction of ultrasound images using bfo cascaded with wiener filter and discrete wavelet transform in homomorphic region,” *Procedia computer science*, vol. 132, pp. 1543–1551, 2018.
- [24] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, “Transfer learning based histopathologic image classification for breast cancer detection,” *Health information science and systems*, vol. 6, no. 1, pp. 1–7, 2018.

- [25] R. Karthiga and K. Narasimhan, “Automated diagnosis of breast cancer using wavelet based entropy features,” in *2018 Second international conference on electronics, communication and aerospace technology (ICECA)*, IEEE, 2018, pp. 274–279.
- [26] S. Karthik, R. Srinivasa Perumal, and P. Chandra Mouli, “Breast cancer classification using deep neural networks,” in *Knowledge computing and its applications*, Springer, 2018, pp. 227–241.
- [27] T. L. Rodziewicz, B. Houseman, and J. E. Hipskind, “Medical error prevention,” 2018.
- [28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [29] C. Varma and O. Sawant, “An alternative approach to detect breast cancer using digital image processing techniques,” in *2018 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2018, pp. 0134–0137.
- [30] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, “Automated breast cancer diagnosis based on machine learning algorithms,” *Journal of healthcare engineering*, vol. 2019, 2019.
- [31] Q. Dong, X. Zhu, and S. Gong, “Single-label multi-class image classification by deep logistic regression,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 2019, pp. 3486–3493.
- [32] M. Hassan, “Resnet (34, 50, 101): Residual cnns for image classification tasks,” *Neurohive. io*, 2019.
- [33] R. Jankovic, “Classifying cultural heritage images by using decision tree classifiers in weka,” in *Proceedings of the 1st International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding Co-Located with 15th Italian Research Conference on Digital Libraries (IRCDL 2019), Pisa, Italy*, 2019, pp. 119–127.
- [34] M. Šarić, M. Russo, M. Stella, and M. Sikora, “Cnn-based method for lung cancer detection in whole slide histopathology images,” in *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, IEEE, 2019, pp. 1–4.
- [35] T. Y. Tan, L. Zhang, and C. P. Lim, “Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models,” *Applied Soft Computing*, vol. 84, p. 105725, 2019.
- [36] Z. Xiang, Z. Ting, F. Weiyan, and L. Cong, “Breast cancer diagnosis from histopathological image based on deep learning,” in *2019 Chinese Control And Decision Conference (CCDC)*, IEEE, 2019, pp. 4616–4619.
- [37] S. Alinsaif and J. Lang, “Histological image classification using deep features and transfer learning,” in *2020 17th Conference on Computer and Robot Vision (CRV)*, IEEE, 2020, pp. 101–108.

- [38] W. Ansar, A. R. Shahid, B. Raza, and A. H. Dar, “Breast cancer detection and localization using mobilenet based transfer learning for mammograms,” in *International symposium on intelligent computing systems*, Springer, 2020, pp. 11–21.
- [39] N. Elazab, H. Soliman, S. El-Sappagh, S. R. Islam, and M. Elmogy, “Objective diagnosis for histopathological images based on machine learning techniques: Classical approaches and new trends,” *Mathematics*, vol. 8, no. 11, p. 1863, 2020.
- [40] H. A. Khan, W. Jue, M. Mushtaq, and M. U. Mushtaq, “Brain tumor classification in mri image using convolutional neural network,” *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 6203–6216, 2020.
- [41] R. Man, P. Yang, and B. Xu, “Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks,” *IEEE access*, vol. 8, pp. 155 362–155 377, 2020.
- [42] F. Parvin and M. A. M. Hasan, “A comparative study of different types of convolutional neural networks for breast cancer histopathological image classification,” in *2020 IEEE Region 10 Symposium (TENSYP)*, IEEE, 2020, pp. 945–948.
- [43] D. P. Pati and S. Panda, “Feature extraction and enhancement of breast cancer mammogram noisy image using image processing,” in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, IEEE, 2020, pp. 1–5.
- [44] P. Sampurnima, S. K. Satapathy, S. Mishra, and P. K. Mallick, “Hyperspectral image classification using stochastic gradient descent based support vector machine,” in *International Conference on Biologically Inspired Techniques in Many-Criteria Decision Making*, Springer, 2020, pp. 78–84.
- [45] Y. Yari, T. V. Nguyen, and H. T. Nguyen, “Deep learning applied for histological diagnosis of breast cancer,” *IEEE Access*, vol. 8, pp. 162 432–162 448, 2020.
- [46] S. A. Alanazi, M. Kamruzzaman, M. N. Islam Sarker, *et al.*, “Boosting breast cancer detection using convolutional neural network,” *Journal of Healthcare Engineering*, vol. 2021, 2021.
- [47] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. A. Almansour, “Deep cnn model based on vgg16 for breast cancer classification,” in *2021 International Conference on Information Technology (ICIT)*, IEEE, 2021, pp. 805–810.
- [48] S. Boumaraf, X. Liu, Y. Wan, *et al.*, “Conventional machine learning versus deep learning for magnification dependent histopathological breast cancer image classification: A comparative study with visual explanation,” *Diagnostics*, vol. 11, no. 3, p. 528, 2021.
- [49] M. Divyavani and G. Kalpana, “An analysis on svm & ann using breast cancer dataset,” *Aegaeum J*, vol. 8, pp. 369–379, 2021.
- [50] Z. Fan, J.-k. Xie, Z.-y. Wang, P.-C. Liu, S.-j. Qu, and L. Huo, “Image classification method based on improved knn algorithm,” in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1930, 2021, p. 012 009.

- [51] J. Ferlay, M. Colombet, I. Soerjomataram, *et al.*, “Cancer statistics for the year 2020: An overview,” *International journal of cancer*, vol. 149, no. 4, pp. 778–789, 2021.
- [52] W. Jiao, X. Hao, and C. Qin, “The image classification method with cnn-xgboost model based on adaptive particle swarm optimization,” *Information*, vol. 12, no. 4, p. 156, 2021.
- [53] C. Vo-Le, N. H. Son, P. Van Muoi, and N. H. Phuong, “Breast cancer detection from histopathological biopsy images using transfer learning,” in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, IEEE, 2021, pp. 408–412.
- [54] K. Manasa and D. Murthy, “Skin cancer detection using vgg-16,” *European Journal of Molecular & Clinical Medicine*, vol. 8, no. 1, pp. 1419–1426, 2021.
- [55] V. Sudha and T. Ganeshbabu, “A convolutional neural network classifier vgg-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning,” *CMC-Computers Materials & Continua*, vol. 66, no. 1, pp. 827–842, 2021.
- [56] E. U. H. Qazi, T. Zia, and A. Almorjan, “Deep learning-based digital image forgery detection system,” *Applied Sciences*, vol. 12, no. 6, p. 2851, 2022.
- [57] D. A. D. Shweta Suresh Naik, “Cancer detection using image processing and machine learning,” 2022.
- [58] R. Sipherd, *The third-leading cause of death in us most doctors don’t want you to know about*, *cncb: Modern medicine*, feb. 22, 2018.
- [59] A. Nicklas, “What is explainable ai (xai)?,” September 18, 2022.