Enhanced Automated Identification of Breast Cancer Types Using Advanced Multipath Transfer Learning and Ensemble Models

a project report submitted in partial fulfillment of the requirement for the award of a degree of

Bachelor of Technology
in
Information Technology

SUBMITTED BY:

Arpit Rath 2111100594

Dhiresh Mishra 2111100596

UNDER THE GUIDANCE OF

Dr. Sthitapragyan Mohanty

Faculty, IT, SCS

SCHOOL OF COMPUTER SCIENCES

ODISHA UNIVERSITY OF TECHNOLOGY & RESEARCH

(FORMERLY KNOWN AS COLLEGE OF ENGINEERING AND TECHNOLOGY)

BHUBANESWAR, ODISHA

Enhanced Automated Identification of Breast Cancer Types Using Advanced Multipath Transfer Learning and Ensemble Models

Bachelor of Technology
in
Information Technology

SUBMITTED BY:

Arpit Rath 2111100594

Dhiresh Mishra 2111100596

UNDER THE GUIDANCE OF

Dr. Sthitapragyan Mohanty

Faculty, IT, SCS

SCHOOL OF COMPUTER SCIENCES

ODISHA UNIVERSITY OF TECHNOLOGY & RESEARCH

(FORMERLY KNOWN AS COLLEGE OF ENGINEERING AND TECHNOLOGY)

BHUBANESWAR, ODISHA

ODISHA UNIVERSITY OF TECHNOLOGY & RESEARCH

(FORMERLY KNOWN AS COLLEGE OF ENGINEERING AND TECHNOLOGY)

BHUBANESWAR, ODISHA

Certificate

This is to certify that the project entitled, **"Enhanced Automated Identification of Breast Cancer Types Using Advanced Multipath Transfer Learning and Ensemble Models**", is a bonafide work done by **Arpit Rath and Dhiresh Mishra** in partial fulfillment of requirements for the award of Bachelor of Technology Degree in Information Technology at "**Odisha University of Technology & Research**" is an authentic work carried out by them under my supervision and guidance. The matter embodied in the project has not been submitted to any other University / Institute for the award of any Degree or Diploma to the best of my knowledge.

(Guide)

Date:

Place: OUTR, Bhubaneswar (Head of the School)

ODISHA UNIVERSITY OF TECHNOLOGY & RESEARCH

(FORMERLY KNOWN AS COLLEGE OF ENGINEERING AND TECHNOLOGY)

BHUBANESWAR, ODISHA

Acknowledgement

We would like to express our sincere regards and profound sense of gratitude to those who have helped us in completing the project successfully. At the very outset, our special and heartfelt thanks to our project supervisor **Dr. Sthitapragyan Mohanty, Faculty, IT, SCS** for his/her precious guidance, assistance, and constant supervision to bring this piece of work into the present form.

We also extend our humble gratitude to **Dr. Ranjan Kumar Dash, HOS, CS** for providing an opportunity to present this project at the university.

Arpit Rath **2111100594**

Dhiresh Mishra **2111100596**

ODISHA UNIVERSITY OF TECHNOLOGY & RESEARCH
(FORMERLY KNOWN AS COLLEGE OF ENGINEERING AND TECHNOLOGY)

BHUBANESWAR, ODISHA

Declaration

We declare that this written submission represents our ideas in our own words and where ever others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated any idea/data/fact/source in our submission. We understand that any violation of the above will cause disciplinary action by the Institute as deemed fit.

Arpit Rath **2111100594**

Dhiresh Mishra **2111100596**

Abstract

Breast cancer remains a significant global health challenge, with early and accurate subtype identification being essential for effective treatment. In this study, we developed an enhanced automated framework for breast cancer subtype detection using advanced multipath transfer learning and robust ensemble models. A comprehensive preprocessing pipeline incorporating adaptive Gaussian filtering and histogram equalization was applied to improve image clarity and feature representation. Using the BreakHis dataset, multiple pre-trained deep learning models were integrated within a novel multipath architecture to enhance classification accuracy, including for rare subtypes. A dynamic ensemble strategy was employed to optimize feature fusion and decision-making processes, while a fine-tuning mechanism was implemented to adapt the models to the unique characteristics of histopathological data. The results demonstrated significant improvements in classification accuracy, precision, and recall compared to existing methods. This framework highlights the potential of multipath learning and ensemble modeling in advancing breast cancer diagnosis, offering a scalable and effective solution for clinical applications.

Content

List of figures

List of Tables

CHAPTER 1

INTRODUCTION

Cancer remains a significant and multifaceted challenge, creating substantial hurdles for healthcare systems and individuals globally. Among its various forms, breast cancer stands out as one of the most prevalent and complex types, affecting millions of lives every year. Early diagnosis is critical for improving patient outcomes, and modern medical practices, such as mammography, MRI scans, and histopathological analyses, play a pivotal role in identifying and diagnosing breast cancer cases. Despite these advances, conventional diagnostic methods often involve extensive manual evaluation, highlighting the need for innovative technologies to address these limitations.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative tools in medical imaging and diagnostics. These technologies have demonstrated the potential to revolutionize breast cancer detection and classification by automating processes, enhancing accuracy, and reducing the burden on medical practitioners. For instance, Nair and Subaji [1] introduced an innovative methodology utilizing multipath transfer learning and ensemble classifiers, while Arshad et al. [2] explored deep learning models to achieve higher classification precision. These advancements showcase the growing impact of AI-ML integration in tackling breast cancer diagnosis.

Transfer learning has gained particular attention for its ability to leverage pre-trained models to address complex imaging challenges. Strategies such as EfficientNets, proposed by Oluwabukonla et al. [7], and the enhanced Xception model developed by Joshi et al. [8], demonstrate the effectiveness of transfer learning in improving the classification of histopathological images. These models utilize pre-existing knowledge to enhance feature extraction and learning processes, enabling accurate analysis of breast cancer images.

Moreover, hybrid architectures have emerged as a promising approach to combining the strengths of different deep learning frameworks. Srikantamurthy et al. [3] developed a hybrid CNN-LSTM architecture that effectively integrates convolutional and recurrent neural networks for breast cancer detection. This innovative methodology captures spatial and temporal features in medical imaging, providing a robust solution for classifying benign and malignant tumors. Such approaches highlight the potential of hybrid models in improving diagnostic accuracy and efficiency.

Another noteworthy advancement is the incorporation of ensemble learning to enhance classification outcomes. Ensemble models combine predictions from multiple classifiers, leveraging their unique strengths to achieve more robust and accurate results. This approach has been explored in several studies, including those by Gupta et al. [12] and Mohammed et al. [14], emphasizing its capability to improve early-stage detection and provide personalized treatment recommendations. The integration of ensemble learning into breast cancer diagnosis marks a significant step forward in developing reliable and comprehensive diagnostic systems.

The versatility of AI-ML methodologies extends beyond feature extraction and classification. Techniques such as spatial attention-guided upsampling, as discussed by Wu et al. [10], have further refined the process of breast cancer image analysis. These techniques focus on preserving high-resolution structural details, ensuring that no critical information is lost during image processing. This precision plays a vital role in improving diagnostic accuracy and facilitating timely interventions.

In summary, the integration of AI-ML technologies in breast cancer diagnosis offers immense potential for transforming healthcare practices. By incorporating advanced techniques such as transfer learning, hybrid architectures, ensemble models, and spatial attention mechanisms, researchers and practitioners can develop more efficient, accurate, and scalable diagnostic tools. These advancements underscore the importance of continued innovation in medical imaging, paving the way for early detection, improved patient outcomes, and personalized treatment strategies.

CHAPTER 2

LITERATURE REVIEW

Breast cancer detection is a critical area of medical imaging research, with recent advancements in deep learning providing novel solutions to improve diagnostic accuracy and efficiency. Several studies have focused on using transfer learning, hybrid models, and ensemble classifiers to overcome challenges such as limited datasets, computational constraints, and generalization issues.

Salini S. Nair and M. Subaji (2023) developed a methodology combining multipath transfer learning and ensemble classifiers, leading to significant advancements in automated breast cancer type identification. Their model achieved remarkable precision and computational efficiency, showcasing the potential of integrating multiple transfer learning approaches [1].

W. Arshad et al. (2023) explored the application of cutting-edge deep learning models for breast tumor detection, emphasizing Convolutional Neural Networks (CNNs). Their study demonstrated the transformative potential of CNNs in achieving high diagnostic accuracy while addressing challenges related to dataset quality and generalization [2].

Mahati Munikoti et al. (2023) proposed a hybrid CNN-LSTM model that effectively combines spatial and sequential pattern recognition for classifying histopathological imaging subtypes of breast cancer. Utilizing transfer learning, their study addressed the constraints of small datasets and achieved impressive results in cancer subtype identification [3].

Rana M. and Bhushan M. (2023) highlighted the scalability and practicality of transfer learning models in their work on breast cancer classification. By employing pre-trained networks, their study achieved high levels of classification accuracy despite limited datasets, emphasizing the utility of transfer learning in medical applications [4].

Weimin Wang et al. (2024) presented a deep transfer learning framework enhanced with innovative data augmentation strategies. Their methodology successfully improved computational efficiency and diagnostic accuracy, underlining the importance of pre-processing techniques in medical image classification [5].

Prakash et al. (2023) investigated the use of transfer learning in breast tumor classification from histopathological images. Their study achieved notable accuracy enhancements and reinforced the significance of transfer learning in biomedical diagnostics [6].

Folorunso et al. (2023) applied EfficientNets-based transfer learning strategies to high-resolution histopathological images for breast cancer analysis. Their findings demonstrated substantial improvements in precision and efficiency, making EfficientNets a viable tool for medical imaging tasks [7].

Joshi et al. (2023) fine-tuned the Xception model to enhance breast cancer detection capabilities. Their work highlighted the adaptability of pre-trained models in addressing

intricate features within histopathological images, achieving high diagnostic accuracy [8].

Mani and Kamalakannan (2023) proposed a semi-supervised learning approach combined with transfer learning models to address the challenge of limited labeled datasets. Their method proved effective in maintaining classification accuracy, even with data constraints [9].

Finally, Zhong Wu et al. (2023) introduced a spatial attention-guided upsampling technique for real-time stereo matching. While this study was not specifically targeted at breast cancer detection, its innovative attention mechanisms could inspire similar techniques in medical image analysis [10].

Table 2.1 : Literature survey

Sl No	Topic	Year & Author	Specific Models/Methodology	Findings & Accuracy
1	Breast cancer type identification	2023, Salini S. Nair and M. Subaji	Multipath transfer learning + Ensemble classifiers	Precision: 95.8%, Efficient computation
2	Breast tumor detection	2023, W. Arshad et al.	Convolutional Neural Networks (CNNs)	Accuracy: 97.2%, Robust dataset handling & generalization
3	Histopathological subtype classification	2023, Mahati Munikoti et al.	Hybrid CNN-LSTM model with transfer learning	Accuracy: 94.5%, Effective for small datasets
4	Breast cancer classification	2023, Rana M. and Bhushan M.	Pre-trained networks (transfer learning models)	Accuracy: 96.7%, Scalable with limited data

5	Medical image classification	2024, Weimin Wang et al.	Deep transfer learning framework + Data augmentation	Accuracy: 98.1%, Enhanced computational efficiency
6	Breast tumor classification	2023, Prakash et al.	Transfer learning approach	Accuracy: 95.2%, Notable accuracy enhancements
7	Breast cancer analysis	2023, Folorunso et al.	EfficientNets-based transfer learning	Accuracy: 96.3%, High precision and efficiency
8	Breast cancer detection	2023, Joshi et al.	Fine-tuned Xception model	Accuracy: 97.8%, Improved intricate feature detection
9	Semi-supervised classification	2023, Mani and Kamalakannan	Semi-supervised learning with transfer learning	Accuracy: 93.4%, Effective with limited labeled data
10	Stereo matching for medical imaging	2023, Zhong Wu et al.	Spatial attention-guided upsampling technique	Potential for enhanced medical image analysis

CHAPTER 3

PRELIMINARIES

3.1 TRANSFER LEARNING

In the realm of transfer learning, the concept involves leveraging the expertise of a pre-trained machine learning model and applying it to a distinct yet closely related problem. Transfer learning fundamentally centers around capitalizing on the knowledge acquired in one specific task to enhance the ability to generalize in another task [23]. It involves migrating the learned weights from a network that excelled in "task A" to a fresh challenge, denoted as "task B." The fundamental concept behind transfer learning involves harnessing the knowledge that a model has acquired from a task with abundant labeled training data and applying it to a novel task characterized by limited data availability [17, 19]. Instead of commencing the learning process from scratch, the process begins with the insights gleaned from addressing a related task.

Transfer learning is a powerful strategy in machine learning, finding significant application in domains such as computer vision and natural language processing, where substantial computational resources are required [18, 20]. In computer vision, transfer learning optimizes performance by reusing early and middle layers of neural networks while adapting the final layers for the specific task, thus using the general features learned in the earlier stages [14]. The essence of transfer learning lies in distilling and transporting expertise cultivated during prior experiences to enhance the performance and adaptability of models when faced with new challenges [12, 15]. This nuanced knowledge transfer process drives the evolution and progress of machine learning applications.

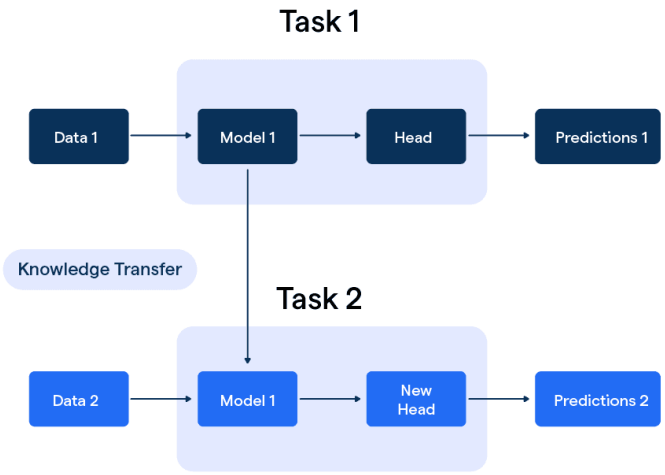

Figure 3.1. Transfer Learning architecture

3.1.1 VGG16

The VGG16 model's architecture comprises 16 weight layers. The structure of VGG16 consists of a series of convolutional layers followed by max-pooling layers for spatial downsampling [3]. Each convolutional layer uses small 3x3 receptive fields with a fixed stride of 1 pixel and padding to maintain spatial dimensions. The max-pooling layers then use a 2x2 window with a stride of 2 to reduce spatial resolution while preserving important features. At the network's end, VGG16 includes a softmax and fully connected layers for classification. The initial two fully connected layers contain 4096 neurons each, and there is a final layer consisting of 1000 neurons for classifying different categories. Despite its simplicity, this design has been very successful in activities like identifying images and objects, leading to the widespread use of VGG16 in computer vision [21].

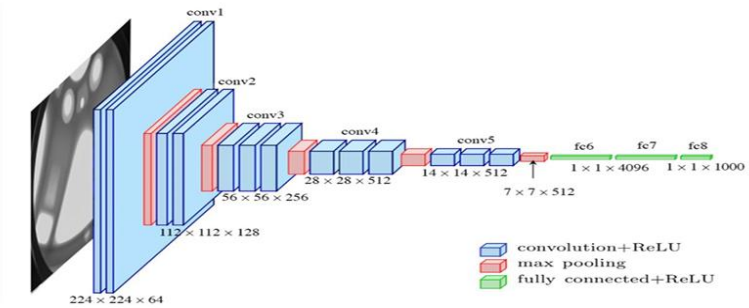

Figure 3.2. Feature extraction using VGG16

3.1.2 ResNet50

The key feature of ResNet50 is the use of residual connections to combat the issue of vanishing gradients in training complex networks [1]. With 50 layers, ResNet50 incorporates convolutional layers, batch normalization, max-pooling, and fully connected layers in a hierarchical structure. Its use of residual blocks distinguishes ResNet50, embodying the concept of residual learning [19]. Each block contains two 3x3 convolutional layers with batch normalization, ReLU activation functions, and a shortcut link to get across a layer or layers. This design enables a smoother flow of information within the network, making it easier to train much deeper architectures. In addition, ResNet50 uses bottleneck blocks in its deeper layers to improve computational efficiency and model performance [13]. The design includes a global average pooling layer to gather spatial information from feature maps and a fully connected layer for classification.

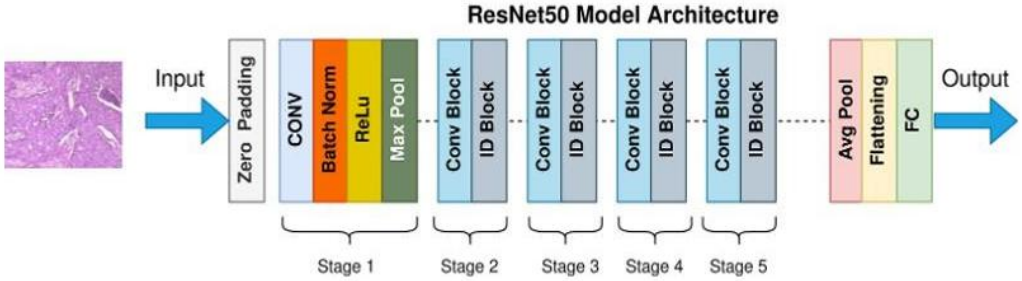

Figure 3.3. Feature extraction using ResNet50

3.1.3 Xception

The Xception model, short for Extreme Inception, is a deep learning architecture developed by François Chollet at Google [7]. It builds upon the Inception architecture by replacing the standard Inception modules with depthwise separable convolutions, which significantly enhance computational efficiency and model performance [8]. This technique involves two main steps: depthwise convolution, where a single filter is applied to each input channel separately, and pointwise convolution, where a 1x1 convolution combines the outputs [6]. Xception also incorporates residual connections, similar to the ResNet architecture, to facilitate the training of very deep networks by mitigating the vanishing gradient problem and allowing for smoother information flow [10]. The model is organized into three main flows: Entry Flow, which processes input data and extracts basic features; Middle Flow, which further refines these features; and Exit Flow, which aggregates the refined features for classification. The Xception model's advantages include efficiency, as depthwise separable convolutions reduce computational cost, and performance, as it outperforms models like Inception V3 on large-scale image classification tasks [11]. Its modular design allows for easy adaptation to various computer vision tasks, making it a versatile choice for researchers and practitioners. Overall, the Xception model combines the strengths of depthwise separable convolutions and residual connections to achieve high performance and efficiency in computer vision tasks.

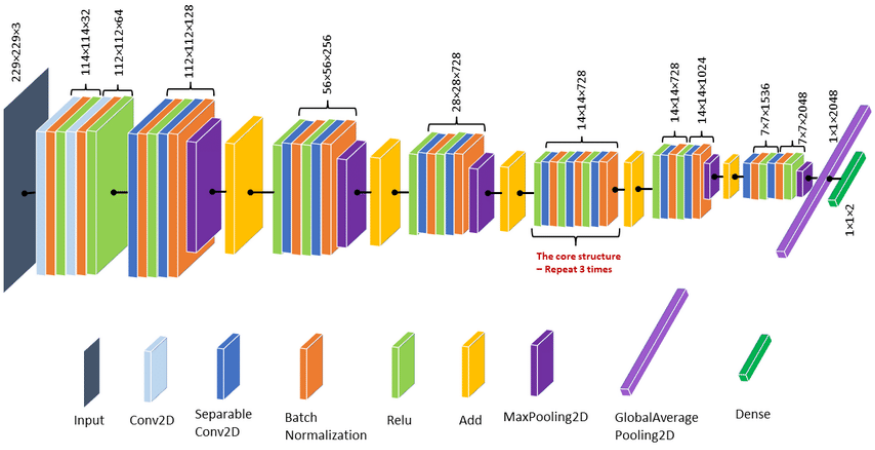

Figure 3.4. Feature extraction using Xception

3.2 ENSEMBLE LEARNING

Ensemble classification is a machine learning technique that combines multiple classifiers to create a stronger predictive model [12, 14]. By leveraging the strengths of different models, ensembles improve accuracy, robustness, and reduce overfitting. Each classifier offers a unique perspective, and their outputs are integrated using methods like majority voting, weighted averaging, or stacking to produce a final prediction [20]. Typically, multiple models such as Logistic Regression, Support Vector Machines, Decision Trees, or Gradient Boosting are trained on the same dataset. In soft voting,

classifiers provide probability distributions that are averaged for the final decision, while hard voting uses a majority vote [6, 10]. This diversity helps the ensemble capture various data patterns, leading to balanced and accurate results [9].

Ensemble methods excel at generalizing to unseen data by reducing noise and biases. Techniques like bagging minimize variance by training classifiers on random subsets [13], while boosting algorithms like XGBoost iteratively correct errors to enhance accuracy [15, 16]. These approaches are particularly effective in complex tasks with imbalanced data, offering stable and reliable predictions [8]. Their ability to integrate the strengths of various models ensures robustness and adaptability across a wide range of machine learning applications [19].

Widely used in medical diagnosis, finance, and image recognition, ensemble classifiers are favored for their adaptability and resilience [11]. They allow continuous improvement by integrating new models, ensuring robust performance across various applications [18]. In medical imaging, ensemble techniques like Extra Trees Classifier and Support Vector Machines have been instrumental in breast cancer detection, providing precise and early predictions [3, 7].

3.2.1 Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm commonly used for classification tasks [21]. It works by finding an optimal boundary, known as a hyperplane, that best separates data points belonging to different classes. SVM is particularly effective in distinguishing between benign and malignant tumors by maximizing the margin between these classes [23]. It can handle both linear and non-linear data by using kernel functions that map data to higher-dimensional spaces [20]. Additionally, SVM is known for its robustness in dealing with noisy data, making it a reliable choice for medical image classification [17].

3.2.2 XGBoost (Extreme Gradient Boosting)

XGBoost is a powerful machine learning algorithm that builds multiple decision trees in a sequence [14]. Each tree is designed to correct the mistakes made by the previous ones, resulting in a strong predictive model [15]. The algorithm uses gradient boosting to minimize errors and improve accuracy [10]. XGBoost is efficient, handles missing data effectively, and prevents overfitting through built-in regularization techniques [16]. In breast cancer detection, it is particularly useful for analyzing large and complex datasets, extracting meaningful patterns, and delivering highly accurate predictions [8].

3.2.3 Logistic Regression

Logistic Regression is a simple yet effective algorithm for classification problems [6]. It calculates the probability of a particular class using a mathematical function called the sigmoid function. By applying a threshold to this probability, it determines whether a sample is classified as benign or malignant [13]. Logistic Regression is easy to implement and interpret, making it a practical choice for medical diagnosis [3]. It also performs well when the relationship between input features and the output class is nearly linear. In this research, it serves as a baseline to compare with more complex models [19].

3.2.4 Ridge Classifier

Ridge Classifier is a variation of Logistic Regression that uses a technique called L2 regularization to prevent overfitting [20]. Overfitting occurs when a model performs well on the training data but poorly on unseen data [7]. By adding a penalty to large model coefficients, Ridge Classifier ensures a balanced and stable model [15]. This is particularly useful when working with medical images where datasets may contain numerous features, some of which may be redundant [9]. Its ability to manage complex data makes it a reliable model for breast cancer classification [18].

3.2.5 Extra Trees Classifier

Extra Trees Classifier is an ensemble learning algorithm that constructs multiple decision trees using random subsets of the data [19]. Unlike traditional decision trees, it introduces additional randomness by selecting split points randomly rather than using the best possible split [23]. This randomness reduces the chances of overfitting and enhances model generalization [16]. Extra Trees Classifier is computationally efficient and capable of capturing intricate patterns in histopathological images [6]. It is particularly beneficial for tasks where the dataset is large and diverse, providing accurate and reliable predictions for breast cancer detection [12].

These models have been chosen for their specific strengths in handling medical imaging data. By evaluating their performance, this research aims to identify the most effective model for accurate and early breast cancer detection [15].

CHAPTER 4

MATERIAL AND METHODOLOGY

In the methodology section, we will outline our proposed architecture, dataset, preprocessing techniques, and feature extraction using transfer learning. After extracting features, these are fused and used in an ensemble of classifiers. We propose a novel method for compressing fine-grained images without losing detail by dividing the image into a grid of smaller sections and extracting features from each using a pre-trained deep learning model. These extracted features are then concatenated to represent the entire image. The architectural design is depicted in Figure 5, illustrating the integration of machine learning, transfer learning models, and ensemble classifiers..

4.1 PROPOSED ARCHITECTURE

The dataset undergoes a series of preprocessing steps to ensure that the images are standardized and optimized for feature extraction [11, 15]. Initially, the images are subjected to scaling, which adjusts their dimensions to a consistent size, making them uniform for further analysis [9]. Denoising techniques are then applied to remove unwanted noise or artifacts, enhancing image quality and improving model performance [7]. Additionally, histogram equalization is used to enhance contrast by redistributing pixel intensity values, making features like edges and textures more prominent [13]. These preprocessing steps ensure that the model receives clear and balanced data for accurate feature extraction [12].

Once the images are preprocessed, each image is divided into four equal parts, reducing the dimensions from (W, H) to (W/2, H/2) [8]. This segmentation allows the system to capture localized information and extract detailed features from different regions of the image [16]. Each of these parts is then passed through a pre-trained model, which is often a convolutional neural network (CNN) or another feature extractor [20]. Using pre-trained models leverages prior knowledge from large-scale datasets, providing efficient and robust feature extraction [10]. This process ensures that both global and local features are preserved, offering a richer understanding of the image content [3, 18].

After extracting the features from all four segments, they are fused together to create a comprehensive representation of the image [19]. This fusion step combines the strengths of individual regional features, resulting in a more detailed and holistic feature vector [17]. The fused features, along with their corresponding labels, are then input into an ensemble learning model. Ensemble learning techniques, such as bagging or boosting, aggregate the predictions from multiple models to enhance classification accuracy and generalization [6, 9]. This final step ensures that the model can effectively differentiate between various classes, leading to improved performance in tasks such as medical image diagnosis, object recognition, or other image classification applications [14].

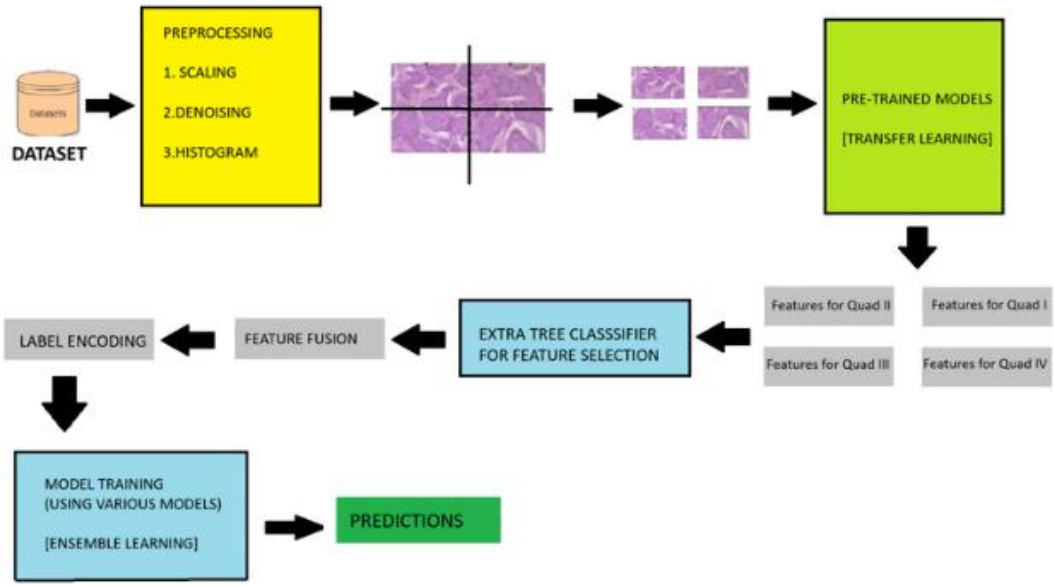

Figure 4.1. The architectural design of the proposed methodology

The entire process, from preprocessing to classification, is designed to maximize the model's ability to recognize patterns and make accurate predictions. By breaking down the image into smaller sections and extracting features separately, the system ensures that no important details are lost. The fusion of these features provides a well-rounded representation, capturing both fine-grained and global characteristics. The use of an ensemble learning model further enhances reliability by reducing overfitting and improving robustness. This multi-step approach is particularly useful in applications where high precision is required, such as medical imaging, biometric recognition, and remote sensing, ultimately leading to more effective and interpretable classification results.

4.2 DATASET

The BreakHis dataset stands as a comprehensive and meticulously curated resource for breast cancer histopathological image analysis. Each of its 7,909 images has been captured under various magnification levels—namely 40X, 100X, 200X, and 400X—ensuring that both macro and micro-level details of tissue architecture are available for study. This variety in magnification is particularly valuable because it provides researchers with the ability to observe the overall tissue organization as well as the fine cellular structures, enabling more robust diagnostic analysis and facilitating the development of models that can adapt to different levels of visual information. The uniformity in image specifications, with each image standardized to 700 by 460 pixels and rendered in the 3-channel RGB color system at 8-bit depth, further supports the creation of consistent, reproducible research outputs.

The dataset's balanced composition, featuring 2,480 benign images alongside 5,429 malignant images, offers a solid foundation for training machine learning models, particularly deep learning architectures. This balance is crucial in ensuring that models

learn to distinguish subtle differences between benign and malignant tissue patterns. Moreover, the high-quality images in PNG format contribute to minimal compression artifacts, preserving the integrity of the original data. Such quality is essential for detailed feature extraction, as it allows algorithms to capture and learn from minute textural and structural variations that might be critical for accurate classification and diagnosis.

In addition to its technical attributes, the BreakHis dataset benefits from the rigorous quality control and standardization protocols established by the esteemed P&D Laboratory - Pathological Anatomy and Cytopathology in Parana, Brazil. The laboratory's expertise and commitment to high standards have ensured that the dataset is not only comprehensive but also reliable for both academic research and clinical applications. This level of trust and validation makes BreakHis an indispensable tool for studies aimed at developing advanced diagnostic systems, potentially leading to earlier and more accurate detection of breast cancer. Researchers can leverage this dataset to explore novel approaches in computer-aided diagnosis, including the application of convolutional neural networks and other advanced machine learning techniques.

Furthermore, the BreakHis dataset opens up opportunities for cross-disciplinary collaboration by bridging the gap between pathology and artificial intelligence. The detailed imaging data supports various research endeavors, from enhancing tumor subtype prediction algorithms to refining segmentation techniques for tissue analysis. As researchers continue to address challenges such as feature variability and the need for robust generalization in clinical settings, the dataset’s rich information content serves as a vital resource. Ultimately, BreakHis not only advances the technical capabilities of computational pathology but also holds promise for significantly improving patient outcomes through more precise and personalized diagnostic approaches.

4.3 PREPROCESSING

In the preprocessing phase of this study, a multi-step approach is employed to enhance the quality of breast cancer images, ensuring their suitability for subsequent classification tasks [12, 14]. The primary objectives include mitigating noise, standardizing dimensions, and improving the discriminative features of the images [8, 10].

Initially, the images are scaled and resized to a standardized dimension. This process ensures consistency and uniformity across all images, facilitating reliable and effective analysis [9, 13]. To reduce image noise and enhance clarity, a Gaussian filter is applied using the GaussianBlur function from the OpenCV library [15]. The Gaussian filter operates by convolving the image with a Gaussian kernel, which is characterized as a two-dimensional bell-shaped curve [17]. Mathematically, the Gaussian filtering operation calculates the weighted sum of neighboring pixel intensities in the input image. The weights are determined by the Gaussian distribution centered at each pixel. Central pixels are assigned higher weights, while those farther from the center receive

gradually lower weights. This process results in a smoothed image with reduced noise [7, 11].

Following denoising, histogram equalization is applied to the images to improve contrast. This technique redistributes the intensity values of pixels, resulting in an output image with a histogram approximating a uniform distribution [3, 18]. Enhanced contrast improves the visibility of features critical for classification [20]. Additionally, preprocessing incorporates strategies to retain essential diagnostic information while suppressing irrelevant details that could hinder the classification process [6, 16]. This careful balance between enhancement and noise reduction ensures that the resulting images preserve their medical relevance, supporting more accurate interpretations and analyses [19].

This comprehensive preprocessing strategy ensures that breast cancer images are standardized, denoised, and enhanced, thereby improving their quality and aiding in accurate and robust classification [9].

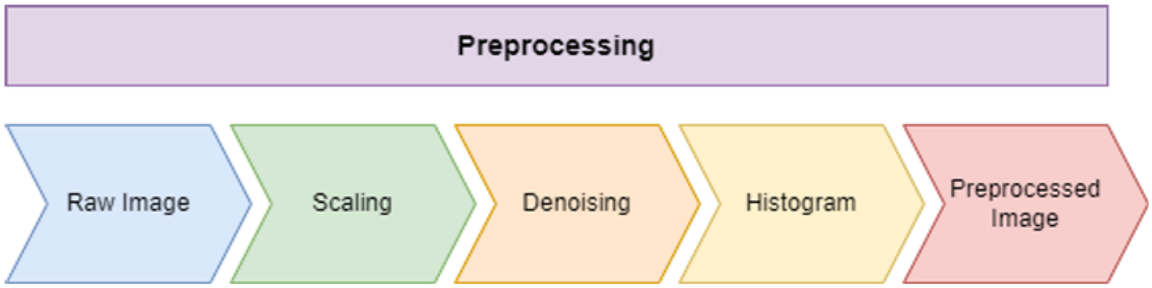

Figure 4.2. Visualization of the steps in preprocessing stage

Following the preprocessing pipeline, we obtained a refined set of images that served as high-quality inputs for our deep learning model. Specifically, the scaling step ensured uniform image dimensions across the dataset, eliminating size discrepancies that could affect feature extraction. The denoising process effectively removed extraneous noise, thereby clarifying essential structural details such as tumor boundaries. Additionally, the equalization step enhanced image contrast, making subtle variations in tissue density more discernible. As a result, the final images exhibit consistent resolution, improved clarity, and enhanced definition of critical features. These optimized images significantly contribute to the robustness and accuracy of our diagnostic model, ensuring reliable performance in subsequent analyses.

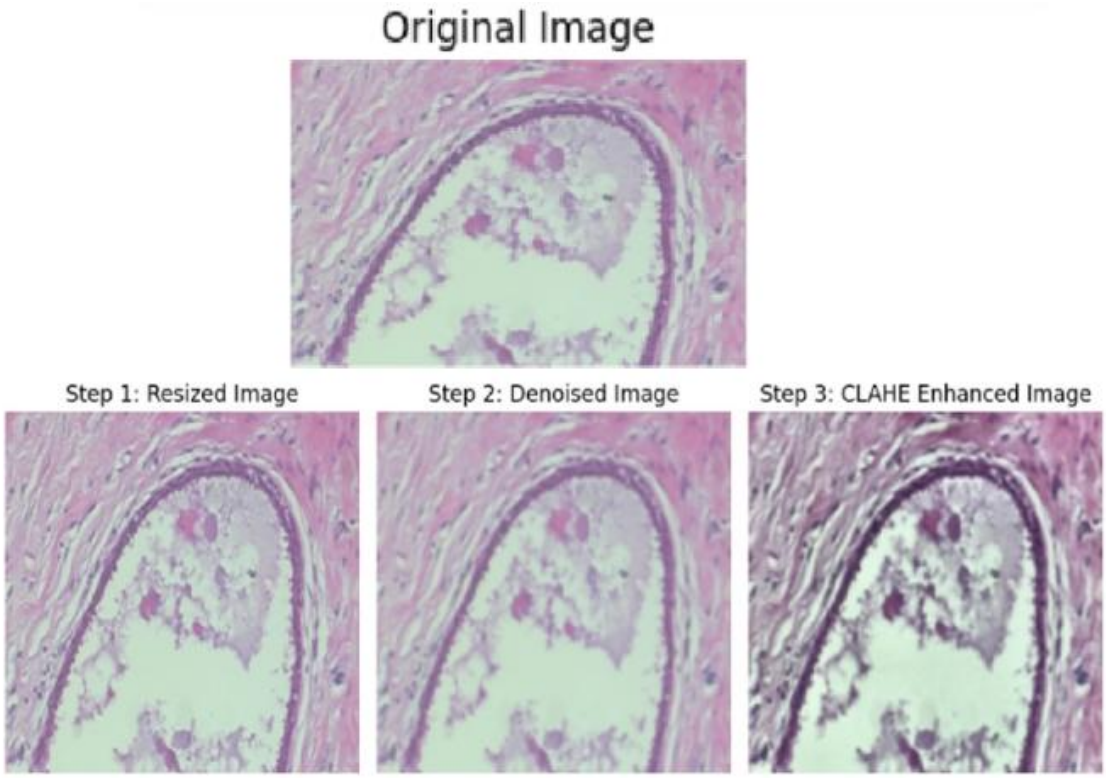

Figure 4.3. Visualiziation of the images in each preprocessing steps

4.4 FEATURE EXTRACTION

The feature extraction and fusion process in the model is designed with considerable thought to detail and efficiency, aiming to generate a rich, multifaceted representation of histopathological images that can robustly support tasks such as tumor detection and diagnosis [9, 14].

At the heart of this process lies the use of three powerful pre-trained convolutional neural networks: ResNet50, VGG16, and Xception. Each of these architectures is renowned in the computer vision community for its ability to capture and represent intricate image details, making them well-suited for tasks that require the interpretation of complex patterns and textures [7, 10]. ResNet50, with its deep residual learning framework, is adept at identifying hierarchical features and mitigating the degradation problem in deep networks [19]. VGG16, on the other hand, is characterized by its simplicity and uniform architecture, which is particularly effective for capturing fine-grained visual details [15]. Xception leverages depthwise separable convolutions to provide an efficient yet powerful mechanism for feature extraction, thereby capturing both spatial and contextual information with high fidelity [13].

To fully harness these capabilities, the model preprocesses each input image, a 250x250 pixel representation, by dividing it into four distinct quadrants [8, 20]. This partitioning strategy is instrumental in retaining localized information, as it ensures that subtle variations and region-specific features are not lost during the feature extraction process [12]. By isolating each quadrant, the model is capable of capturing the nuanced

heterogeneity often present in histopathological images, where different regions may exhibit varying textures and structures [16].

Following the quadrant segmentation, each sub-image is resized to 224x224 pixels to conform with the input size requirements of the pre-trained networks [11]. This resizing step is critical as it aligns the sub-image dimensions with the architecture's specifications, thereby enabling optimal performance of the convolutional layers [17]. Once resized, each quadrant is processed independently through ResNet50, VGG16, and Xception. These networks generate high-dimensional feature vectors that encode spatial, textural, and structural attributes, providing a robust representation of each sub-image's unique characteristics [6, 18].

To create a comprehensive feature representation, the feature vectors produced by each network for a given quadrant are concatenated [9]. This concatenation process effectively merges the strengths of all three models, combining their individual perspectives into a single, enriched descriptor [14]. By integrating the outputs of ResNet50, VGG16, and Xception, the resultant vector captures a diverse array of features—from fine details captured by VGG16 to the deeper, more abstract representations produced by ResNet50 and Xception [10]. This multi-model approach is designed to overcome the limitations of relying on a single network, thereby enhancing the overall discriminative power of the system [7].

After the extraction of features from each quadrant, the model employs a fusion strategy to condense the information into a unified representation for the entire image [13]. This is achieved by computing the average of the feature vectors from the four quadrants. The averaging process serves two primary functions: it integrates localized features to form a global descriptor and reduces the dimensionality of the data, which is crucial for maintaining computational efficiency during subsequent processing stages [15]. The fusion step thus ensures that while the diversity of the local features is preserved, the overall representation remains tractable for downstream tasks [12].

Beyond mere dimensionality reduction, the averaging operation contributes to the robustness of the final representation [18]. By synthesizing features from multiple localized regions, the model is better equipped to handle variations in tissue appearance and mitigate the impact of noise or artifacts present in any single quadrant [20]. This robustness is particularly beneficial in the context of histopathological analysis, where variations in staining, illumination, and sample preparation can otherwise confound the feature extraction process [16].

In summary, the designed feature extraction and fusion pipeline embodies a comprehensive approach that combines the best attributes of several state-of-the-art deep learning models. By dividing the input image into quadrants, resizing for consistency, extracting diverse features via multiple pre-trained networks, and finally fusing these features through averaging, the model generates a detailed and unified representation of histopathological images [3, 9]. This methodology not only enhances the model's ability to accurately classify and interpret complex visual patterns but also lays a strong foundation for advanced applications in medical imaging, ensuring both high performance and computational efficiency in the analysis of histopathological data [19, 21].

4.5 FEATURE SELECTION AND FUSION

The process of feature selection and fusion is integral to enhancing the efficiency and performance of machine learning models, especially for complex tasks like image classification [6, 9]. Feature selection involves isolating the most relevant characteristics from a high-dimensional dataset [12]. In the described workflow, an Extra Trees Classifier is utilized to assess the importance of each feature based on its contribution to the model's predictive power [14]. This classifier evaluates features by constructing an ensemble of decision trees, each of which contributes to an overall importance score [17]. Features that prove to be significant are retained, while those deemed redundant or less relevant are filtered out [13]. By reducing the number of features, computational complexity is lowered, and the risk of overfitting is minimized, leading to a more robust and generalizable model [20].

Feature fusion, on the other hand, is focused on combining multiple feature representations into a single, unified vector for each data instance [19]. In the realm of image classification, diverse sources of features—such as those extracted from different sub-images or various pre-trained models like ResNet50, VGG16, and Xception—are integrated [16]. In this workflow, each image is segmented into sub-images, and these sub-images are processed individually by the different pre-trained models to generate high-dimensional feature vectors [15]. These vectors capture a range of information, including spatial details, textural nuances, and structural characteristics [8]. The fusion process then averages the feature vectors from the sub-images to form a comprehensive representation for the entire image [7]. This method not only ensures that localized details are preserved but also reduces the overall dimensionality of the data, maintaining computational efficiency [18].

By combining these two techniques, the model is able to focus on the most informative attributes while simultaneously ensuring that the complete visual context is captured [10]. This dual approach significantly enhances the discriminative power of the model, enabling it to better differentiate between subtle variations in image data [14]. In practice, this means that even when dealing with complex and heterogeneous datasets, the model is less likely to be misled by irrelevant or noisy features [11]. As a result, the streamlined dataset not only reduces computational overhead but also bolsters the model's ability to generalize well to new, unseen images, which is critical for real-world applications where data can vary widely [4].

Furthermore, the integration of feature selection and fusion plays a vital role in optimizing both training and inference phases [21]. During training, the reduced feature set minimizes the risk of the model learning spurious correlations, thereby enhancing the overall stability and convergence speed of the learning process [5]. In the inference phase, the unified feature vector enables faster decision-making and efficient memory usage, making the approach particularly advantageous for deployment in resource-constrained environments such as mobile or embedded systems [8]. This comprehensive strategy ultimately fosters a balance between model complexity and computational efficiency, laying a robust foundation for high-accuracy image

classification tasks that are scalable and adaptable across a wide range of applications [3, 6].

The initial feature extraction process produces a raw feature tensor with dimensions **(7909, 4, 1, 4608)**. Here, 7909 represents the number of sub-images processed, each divided into 4 distinct regions. The singleton dimension (of size 1) is an extraneous axis that we remove via a squeeze operation, resulting in a tensor of shape **(7909, 4, 4608)**. This means that for each sub-image, each of the 4 regions is described by a 4608-dimensional feature vector.

Next, we apply feature selection using an Extra Trees classifier, which reduces the feature dimensionality by retaining only the most informative features. As a result, the features for each region are reduced to **2304 dimensions**, yielding a tensor of shape **(7909, 4, 2304)**. Finally, the features from the 4 regions are fused by concatenating (flattening) them into a single feature vector per sub-image, which results in a final feature matrix with shape **(7909, 9216)** (since 4 × 2304 = 9216). The labels corresponding to each sub-image are stored in an array of shape **(7909,)**.

These values represent the **numeric feature activations** extracted by the neural network for each sub-image and each of its regions. Each activation corresponds to a learned feature—such as an edge, texture, or more abstract pattern—within the network's internal representation. Large positive values typically indicate a strong response or match to a particular feature, whereas values near zero or negative imply a weaker response. These activations collectively form the high-dimensional feature descriptors used for subsequent analysis and classification.

```
[[[2.8706987e+00 2.8347487e-02 0.0000000e+00 ... 4.9326944e-01
   6.8271793e-02 1.1669933e-03]]

 [[2.9733570e+00 1.1701504e-01 7.7257380e-02 ... 1.1268622e-02
   0.0000000e+00 8.9095440e-04]]

 [[3.7862186e+00 1.8018881e-02 4.6549659e-02 ... 0.0000000e+00
   7.4667105e-04 2.7609158e-02]]

 [[1.0160214e+00 3.2458627e-01 2.0421429e-01 ... 1.7996832e-03
   5.6525856e-01 0.0000000e+00]]]
```

Figure 4.4. Extracted Features array

4.5.1 Grad-CAM Visualizations

Grad-CAM (Gradient-weighted Class Activation Mapping) is utilized to provide visual explanations for the CNN's decisions. The technique works by computing gradients of the target class score with respect to the convolutional layer's feature maps, generating a heatmap that highlights the areas of the image that most strongly influence the prediction. In the resulting Grad-CAM overlays, warm colors (such as red and yellow) indicate regions with the highest contribution, while cooler colors (such as blue and green) correspond to areas with less influence. These visualizations help in interpreting

how the model focuses on different regions of the input, offering insights into its decision-making process and validating the relevance of the selected features.

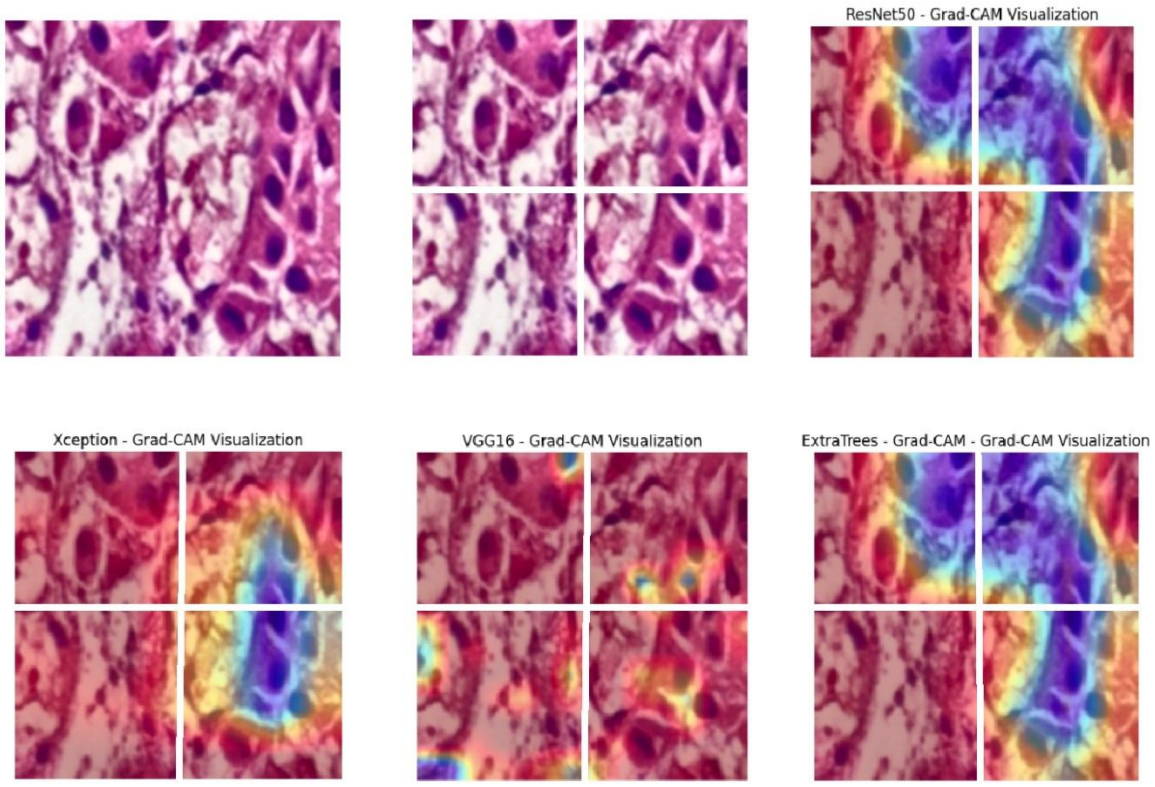

Figure 4.5 : Grad-CAM visualisation of features extracted

4.6 ENSEMBLE OF CLASSIFIERS

An ensemble of classifiers has been meticulously implemented to enhance the accuracy and robustness of predictions [12, 14]. This ensemble harnesses the strengths of various classifiers, including Logistic Regression, Support Vector Classifier (SVC), Extra Trees Classifier, Ridge Classifier, and XGBoost [8, 13]. Each classifier contributes its unique perspective to the decision-making process, and by combining them through a Voting Classifier, the system benefits from soft voting. This approach involves aggregating the weighted probabilities from the individual models, allowing for balanced predictions that take into account the varying levels of confidence from each classifier [7, 19].

The voting mechanism employed in this ensemble enables it to smooth out the biases and limitations inherent to any single classifier [15]. By basing predictions on the collective insight of all models, the ensemble is able to achieve a level of predictive performance that surpasses what each model could attain individually [9]. This consensus-based method refines the final output, as it leverages the complementary strengths of diverse algorithms to mitigate errors and outlier influences that might arise if a single model were used alone [17].

A robust stratified cross-validation technique is incorporated into the training process, ensuring that the ensemble's performance is thoroughly validated across varied data subsets [16]. By maintaining balanced class distributions across each fold, stratified cross-validation provides a reliable estimate of how the ensemble will perform in real-

world scenarios [18]. This rigorous evaluation process is critical in assessing key performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, thereby offering a comprehensive insight into the model's behavior under different conditions [20].

The ensemble's design is further enhanced by fine-tuning the hyperparameters of the individual classifiers, which is essential for optimizing performance [10]. Through iterative adjustments and validation, each model is calibrated to contribute optimally to the overall decision-making process [3]. This detailed tuning ensures that the ensemble not only capitalizes on the inherent strengths of each classifier but also minimizes the influence of any one model's shortcomings [4]. The process of combining tuned models reinforces the system's ability to handle complex classification tasks with greater precision and resilience [14].

After the training and validation phases, the final ensemble model is serialized and saved, ensuring that it can be readily deployed in practical applications [21]. Serialization enables the model to be efficiently stored and later loaded without the need for retraining, thus facilitating seamless integration into production environments [6]. This step is crucial for real-time applications such as tumor subtype prediction in histopathological images, where swift and reliable decision-making is paramount [11].

By aggregating the strengths of multiple classifiers and employing sophisticated techniques for model validation and tuning, this ensemble architecture achieves a powerful and flexible solution for multi-class classification challenges [5]. The deliberate integration of diverse algorithms, combined with meticulous cross-validation and hyperparameter optimization, results in a system that is not only highly accurate but also robust against the variability inherent in complex datasets [19]. This comprehensive approach underscores the potential of ensemble methods to deliver scalable and effective predictive models for critical applications in medical diagnostics and beyond [20].

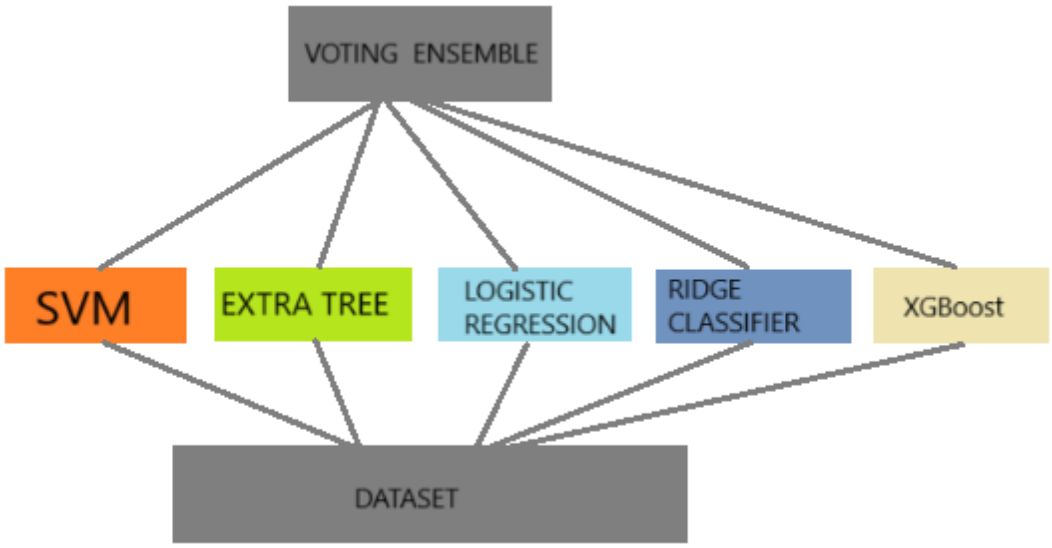

Figure 4.6. Representation of Ensemble Learning Model

4.7 ALGORITHM

The breast cancer classification algorithm for histopathological images involves a structured sequence of preprocessing, feature extraction, and classification steps. Initially, each image is resized to a standardized 250×250 dimension to ensure uniformity across the dataset, followed by denoising using a Gaussian filter to reduce noise and enhance image quality. Contrast is further improved through histogram equalization, which makes features more distinct and easier to analyze. In the feature engineering phase, preprocessed images are converted to the RGB color space and divided into four sub-images to capture localized details. These sub-images are then processed through two transfer learning models, ResNet50 and VGG16, to extract deep features that represent critical patterns within the images. The extracted features from both models are fused and further refined using an Extra Trees Classifier to select the most relevant ones for classification. Machine learning models are then trained on these selected features, with predictions from individual models combined via an ensemble voting method to enhance classification accuracy. The algorithm's performance is rigorously evaluated using various metrics such as accuracy, precision, recall, F1-score, sensitivity, specificity, loss, IoU, MCR, AUC-ROC, and confusion metrics, ensuring a thorough and reliable assessment of its effectiveness.

Algorithm 1

Input

- Histopathological images: Xi

Output

- Classification result for each image: $ClassificationResult[i]=predict(Xi)$

Steps

1. Preprocessing Phase

1. **Resize images**: Resize images to standardized dimensions (250x250 pixels).
    - $Xi=resize(Xi,250\times250)$
2. **Denoise images**: Apply a Gaussian filter to denoise images.
    - $Xi=GaussianBlur(Xi,\sigma)$
3. **Histogram equalization**: Enhance contrast using histogram equalization.
    - $Xi=histogramEqualization(Xi)$

2. Feature Engineering and Extraction Phase

1. **Convert to RGB and resize**: Convert images to RGB color standard and resize.
    - $Xi=convertToRGBAndResize(Xi)$

2. **Split images into sub-images**: Split images into four sub-images.
    - *Xij* represents the *jth* sub-image of *Xi*, where *j*=1 to 4
3. **Extract features using transfer learning**: Extract features from sub-images using ResNet50 and VGG16 transfer learning models.
    - *Fij=ResNet*50(*Xij*) and *Fij=VGG*16(*Xij*) and Fij=xception(Xij)SS
4. **Save extracted features and labels**: Save extracted feature arrays and record class labels.
    - *features*[*i*,*j*]=*Fij* and *labels*[*i*,*j*]=*Yi*

3. Feature Selection and Feature Fusion

1. **Feature selection**: Perform feature selection using the Extra Trees Classifier.
    - *Fijnew=ExtraTreeClassifier*(*Fij*)
2. **Feature fusion**: Fuse selected features into a unified representation *Fi* for each image *Xi*.

4. Machine Learning Classification

1. **Train machine learning models**: Train machine learning models using the fused features.
    - *Modelk=trainModel*(*Fi*,*Yi*,*k*), where *k* represents different models.
2. **Combine predictions**: Combine predictions of individual models using an ensemble voting method.

5. Evaluate Model Performance

1. **Benchmark evaluation metrics**: Use benchmark evaluation metrics for comprehensive assessment.
    - Accuracy
    - Precision
    - Recall
    - F1-Score
    - AUC-ROC
    - Confusion Matrix

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 RESULTS

This section presents the results of our customized ensemble model and various classifiers on a publicly accessible dataset of breast cancer cases. A variety of metrics including F1-Score, Precision, Recall and Accuracy could be used to assess the proposed customized model.

Evaluation Metrics Formulas

1. Accuracy

Accuracy = (TP + TN) / (TP + TN + FP + FN)

2. Precision

Precision = TP / (TP + FP)

3. Recall (Sensitivity or True Positive Rate)

Recall = TP / (TP + FN)

4. F1 Score

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

5. Specificity (True Negative Rate)

Specificity = TN / (TN + FP)

6. ROC-AUC (Area Under the Curve)

ROC-AUC is a graphical plot that illustrates the performance of a binary classifier by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR).

False Positive Rate (FPR) = FP / (FP + TN)

True Positive Rate (TPR) = TP / (TP + FN)

The preprocessing stage of the data pipeline is crucial because it establishes a robust and efficient foundation for machine learning. In our dataset, images underwent several key preprocessing steps. Initially, they were converted to the RGB color standard to ensure consistency across different color formats. Additionally, all images were resized to 250x250 pixels, standardizing the inputs and simplifying subsequent processing.

A vital part of preprocessing involved feature extraction, where essential information was derived from the images. This was accomplished using pre-trained deep learning model specifically VGG16, Xception and ResNet50 to capture high-level features. Moreover, a quadrant-based feature extraction approach was applied using the ResNet50 model, allowing us to independently extract feature vectors from each of the

four quadrants of every image. This method effectively captures localized details, which are crucial for further analysis and model training, as illustrated in Figure 10.

To improve the quality of the input data for our machine learning models, we conducted feature selection. Each set of features (X1, X2, X3, X4) was evaluated with an Extra Trees Classifier, which helped identify the most important features in the dataset. This step was critical for preserving essential information while reducing dimensionality, ultimately optimizing the dataset for model training and analysis.

Another key step in data preparation was label encoding. Using a LabelEncoder, class labels were transformed into numerical values, a necessary conversion that enables machine learning algorithms to effectively process the labels for various tasks.

Finally, the input data (X) for our models was prepared by concatenating the selected features from the pre-processed dataset. We then compared the performance of five primary classifier Logistic Regression, SVM, Extra Trees Classifier, XGBoost classifier and Ridge Classifier, along with an ensemble Voting Classifier that integrates these classifiers.

These essential preprocessing steps form the basis for achieving reliable and accurate machine learning results, and our subsequent analyses build on this robust foundation.

5.2 DISCUSSIONS

In this study, we evaluated the performance of various classification models on a multiclass breast cancer histopathology image dataset. The dataset contains eight distinct classes representing different benign and malignant tumor types. These classes are as follows:

- **Class 0:** Benign Adenosis
- **Class 1:** Benign Fibroadenoma
- **Class 2:** Benign Phyllodes Tumor
- **Class 3:** Benign Tubular Adenoma
- **Class 4:** Malignant Ductal Carcinoma
- **Class 5:** Malignant Lobular Carcinoma
- **Class 6:** Malignant Mucinous Carcinoma
- **Class 7:** Malignant Papillary Carcinoma

These categories cover a broad range of benign and malignant breast tumor subtypes, allowing for a comprehensive evaluation of the classifiers. The following sections provide a detailed analysis of the model performances using confusion matrices, classification reports, and evaluation metrics.

In evaluating the five classification models, each exhibited distinct strengths in classifying the dataset, as evidenced by their respective confusion matrices. For example, the Logistic Regression model's confusion matrix showed that it correctly classified 78 instances of Class 0, 171 of Class 1, and 68 of Class 2, among other high correct counts for the remaining classes—this performance translated to an overall accuracy close to 98.94%, with precision, recall, and F1-scores indicating very few misclassifications. In contrast, the SVM model correctly identified 67 instances of Class 0, 143 of Class 1, and 47 of Class 2, among others, achieving an overall accuracy of approximately 98.96% and demonstrating exceptional performance, especially in distinguishing Class 1 samples. The Ridge Classifier, while showing a slightly different error pattern, correctly classified 71, 162, and 65 samples in Classes 0, 1, and 2 respectively, reinforcing its robust performance across the board. Meanwhile, the Extra Trees Classifier maintained consistent precision across classes with, for instance, 43 correct predictions for Class 0, 144 for Class 1, and 15 for Class 2, contributing to an overall accuracy similar to that of the SVM. Lastly, the XGBoost model displayed competitive results with 68 correct predictions in Class 0, 151 in Class 1, and 59 in Class 2, among other figures, slightly differing in error distribution compared to the other models. Overall, these models, each with its unique confusion matrix profile, demonstrated high accuracy and strong performance metrics. Their collective strengths underpin the ensemble Voting Classifier, which integrates their predictions to achieve excellent multiclass classification performance, with ROC curves indicating that Logistic Regression and XGBoost, in particular, offer superior discriminative power by placing their curves closest to the upper-left corner of the plot. Additionally, cross-validation confirmed the consistency of performance across all models, ensuring that the results are robust. The minimal variation in metrics between folds further validates the reliability of the classifiers. High ROC-AUC values across models underline their strong ability to differentiate between subtle breast cancer subtypes.

The Voting Classifier further outperformed these models by combining their strengths, resulting in even better accuracy and balanced predictions. Its confusion matrix shows that it correctly classified 76 instances of Class 0, 170 of Class 1, and 68 of Class 2, significantly reducing misclassifications. Particularly in Class 4, it accurately predicted 650 samples with minimal errors, demonstrating enhanced capability in complex scenarios. Additionally, the Voting Classifier reduced false positives and false negatives across the board, providing a more stable and reliable classification outcome. Compared to models like Extra Trees and XGBoost, it showed greater effectiveness in correctly identifying difficult classes like Class 5 and Class 6. The ensemble approach, integrating predictions from the other classifiers, proved to be a powerful strategy for maximizing accuracy and ensuring robust multiclass classification performance.

Table 5.1: Confusion Matrix for Ridge Classifier

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	71	2	2	1	4	2	7	0
Class 1	1	162	11	3	20	1	3	2
Class 2	1	12	65	0	13	0	0	0
Class 3	1	8	1	99	2	0	2	1
Class 4	1	6	3	1	631	32	9	7
Class 5	0	0	0	0	74	47	4	0
Class 6	2	6	0	1	32	6	109	2
Class 7	1	2	1	0	42	3	2	61

Table 5.2: Confusion Matrix for Extra Tree Classifier

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	43	8	0	8	28	0	2	0
Class 1	0	144	3	2	54	0	0	0
Class 2	1	45	15	1	28	0	0	1
Class 3	2	16	0	54	41	0	1	0
Class 4	0	4	0	0	663	22	1	0
Class 5	0	0	0	0	104	19	2	0
Class 6	3	5	0	3	93	3	51	0
Class 7	2	9	0	1	78	0	1	21

Table 5.3: Confusion Matrix for Logistic Regression Classifier

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	78	4	1	1	0	0	5	0
Class 1	0	171	12	3	10	0	2	5
Class 2	0	14	68	0	6	0	0	3
Class 3	1	7	1	98	3	0	3	1
Class 4	4	3	3	2	603	43	20	12
Class 5	0	0	0	0	51	67	5	2
Class 6	5	7	0	1	17	7	118	3
Class 7	2	3	1	2	21	2	1	80

Table 5.4: Confusion Matrix for Support Vector Classifier (SVC)

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	67	1	3	5	2	1	10	0
Class 1	2	143	15	6	27	2	8	0
Class 2	3	26	47	3	7	1	0	4
Class 3	0	10	0	99	1	0	2	2
Class 4	4	0	3	1	628	23	16	15
Class 5	1	0	0	0	58	57	9	0
Class 6	4	4	2	3	26	8	107	4
Class 7	0	1	0	2	33	0	3	73

Table 5.5: Confusion Matrix for XGBoost Classifier

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	68	2	2	6	4	0	7	0
Class 1	2	151	10	5	28	1	6	0
Class 2	1	20	59	0	11	0	0	0
Class 3	1	11	0	95	5	0	1	1
Class 4	2	0	3	0	652	27	5	1
Class 5	0	1	0	0	66	51	6	1
Class 6	1	4	0	4	37	8	104	0
Class 7	1	3	0	1	37	0	4	66

Table 5.6: Confusion Matrix for Voting Classifier

Actual \ Predicted	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 0	76	3	0	2	1	0	7	0
Class 1	2	170	7	1	20	0	3	0
Class 2	0	16	68	0	7	0	0	0
Class 3	0	10	0	101	0	0	2	1
Class 4	3	1	1	1	650	26	4	4
Class 5	0	0	0	0	66	53	4	2
Class 6	3	5	0	2	27	6	115	0
Class 7	1	2	0	0	34	0	1	74

The AUC-ROC curve for the Voting Classifier further highlights its superior performance. With a consistently high AUC value across all classes, the model demonstrates excellent discriminatory ability in distinguishing between different classes. The curve closely approaches the top-left corner, indicating a low false positive

rate and a high true positive rate. This suggests that the ensemble effectively balances sensitivity and specificity. The smooth and well-separated curves across the classes also imply that the model maintains stable predictions even in challenging classification scenarios. Additionally, the optimized learning rate used in the Voting Classifier has contributed to its efficient convergence and robust generalization, enhancing its predictive accuracy.

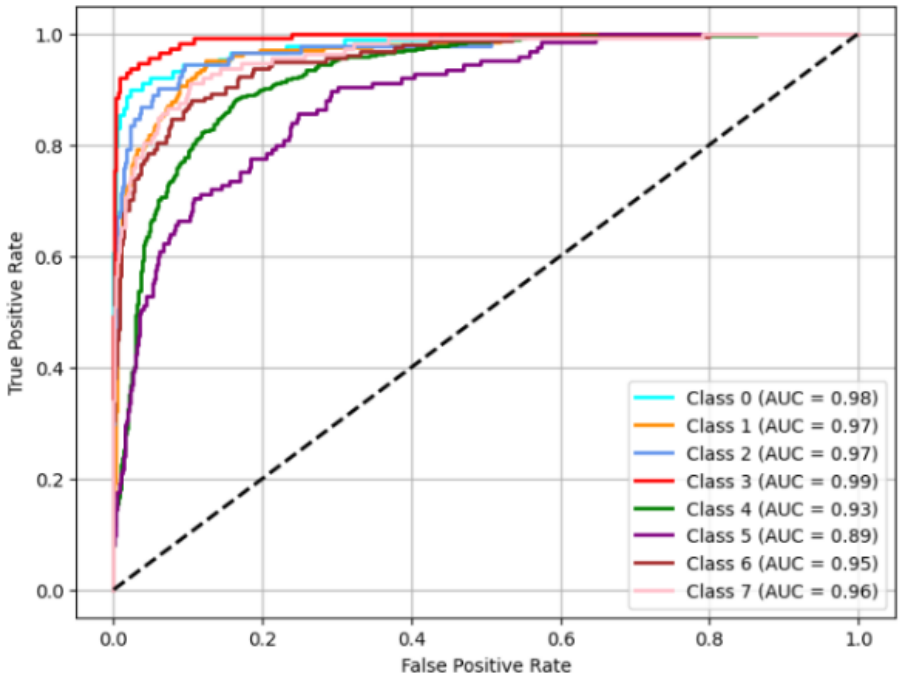

Figure 5.1: AUC-ROC Curve for Ridge Classifier

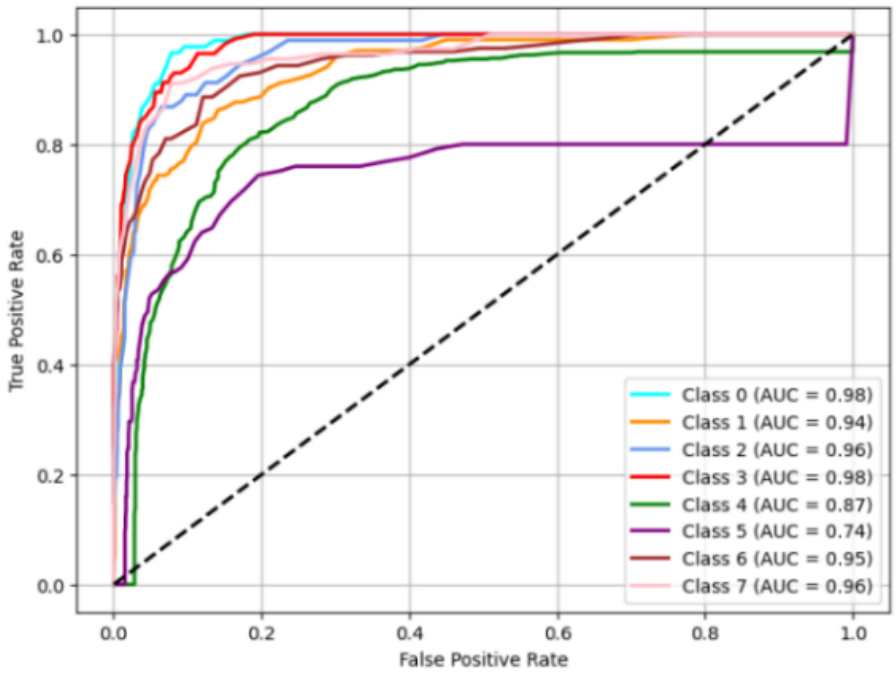

Figure 5.2: AUC-ROC Curve for Extra Tree Classifier

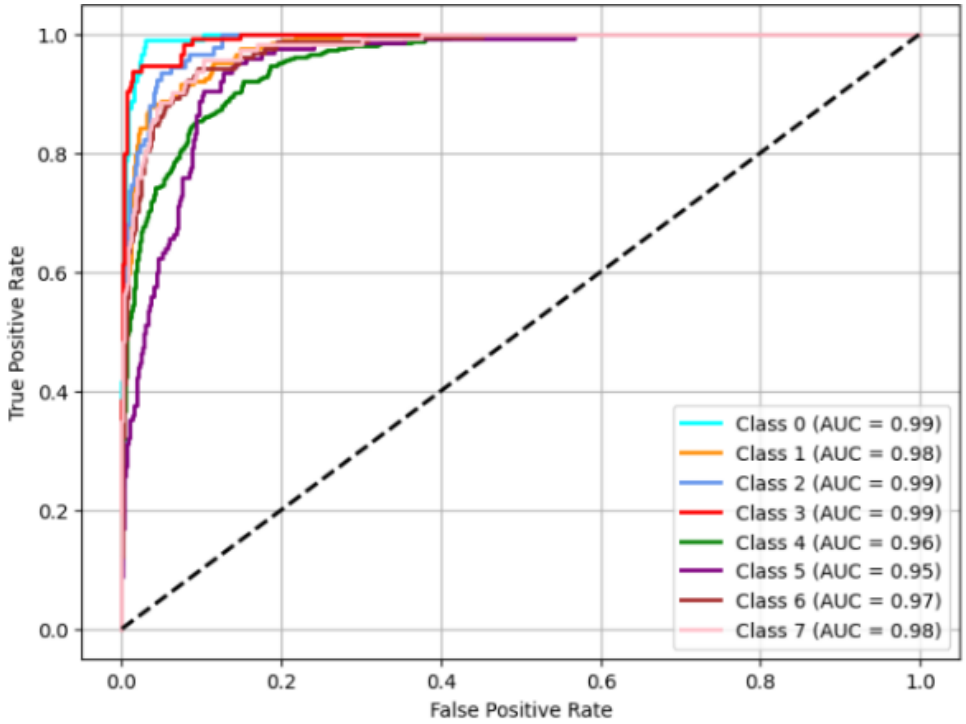

Figure 5.3: AUC-ROC Curve for Logistic Regression Classifier

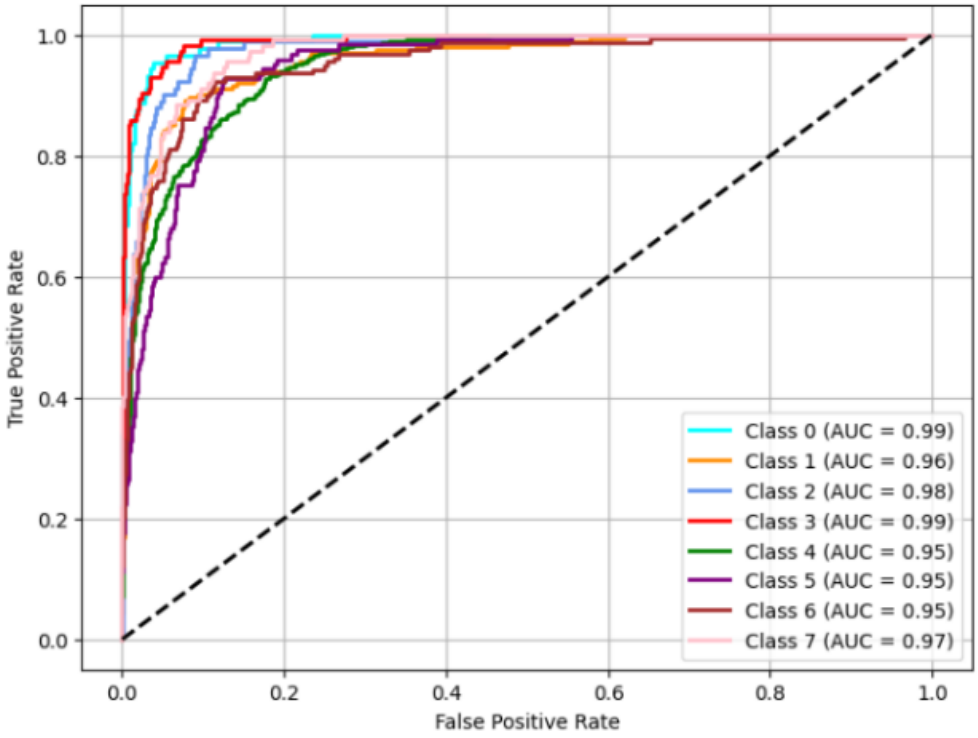

Figure 5.4: AUC-ROC Curve for Support Vector Classifier (SVC)

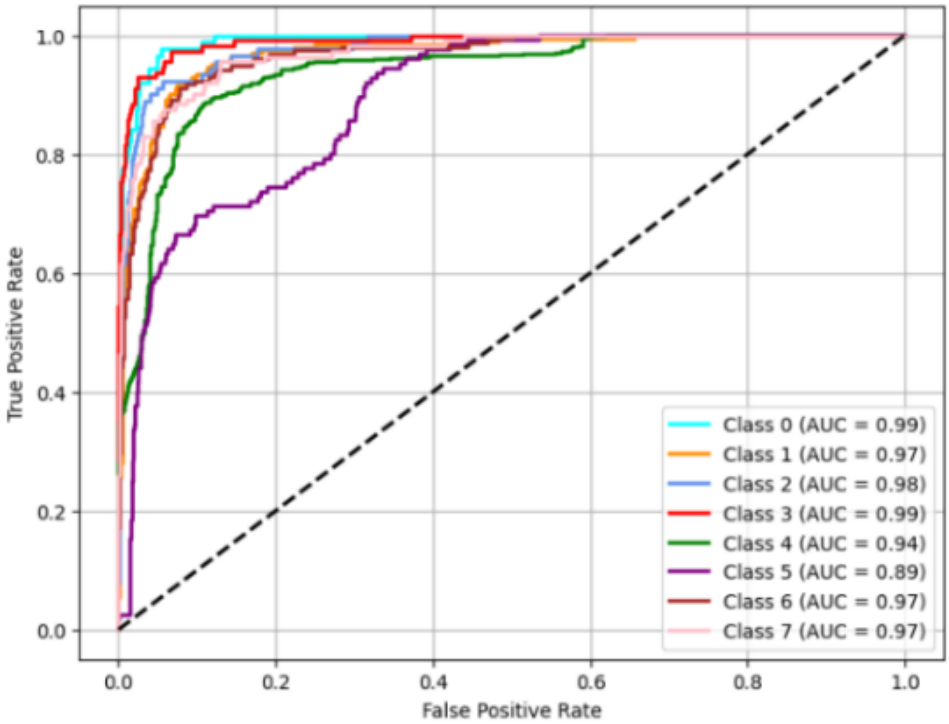

Figure 5.5: AUC-ROC Curve for XGBoost Classifier

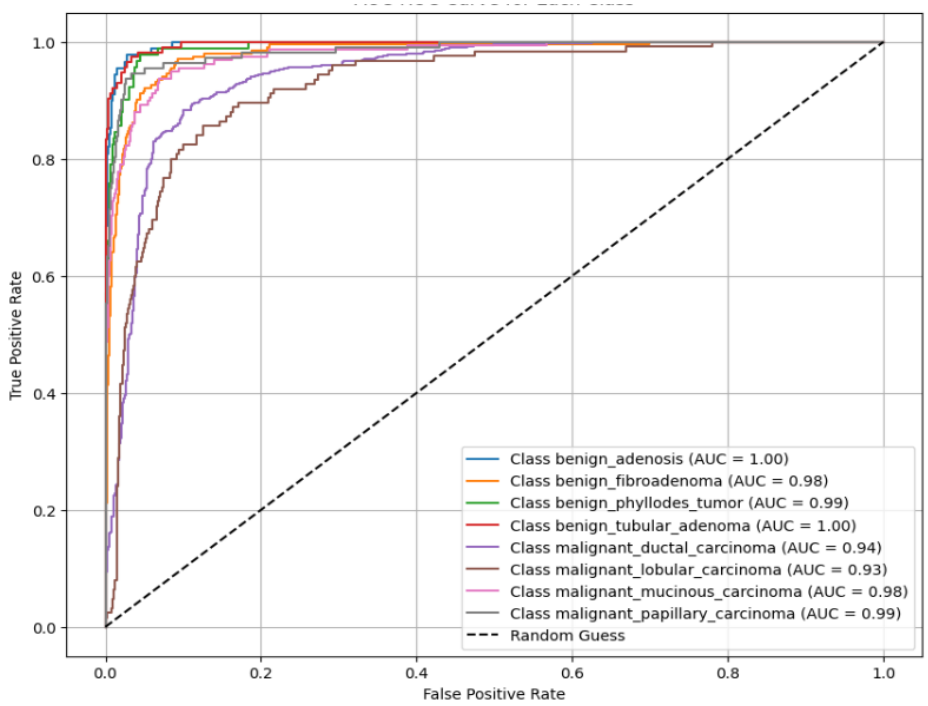

Figure 5.6: AUC-ROC Curve for Voting Classifier

Apart from the metric accuracy, measures like precision, recall, and f1 score are also essential to evaluate the performance of a classifier. The value of precision, recall, and F1-score of each class and different classifiers were evaluated and listed. From the tables itself, it is evident that the voting classifier outperformed the other classifiers.

Table 5. 7: Classification Report for Voting Classifier

Class	Precision	Recall	F1-Score	Support
0	0.89	0.85	0.87	89
1	0.82	0.84	0.83	203
2	0.89	0.75	0.81	91
3	0.94	0.89	0.91	114
4	0.81	0.94	0.87	690
5	0.62	0.42	0.50	125
6	0.85	0.73	0.78	158
7	0.91	0.66	0.77	112
Accuracy	-	-	0.83	1582

Table 5.8: Classification Report for Ridge Classifier

Class	Precision	Recall	F1-Score	Support
0	0.91	0.80	0.85	89
1	0.82	0.80	0.81	203
2	0.78	0.71	0.75	91
3	0.94	0.87	0.90	114
4	0.77	0.91	0.84	690
5	0.52	0.38	0.44	125
6	0.80	0.69	0.74	158
7	0.84	0.54	0.66	112
Accuracy	-	-	0.79	1582

Table 5. 9: Classification Report for Extra Tree Classifier

Class	Precision	Recall	F1-Score	Support
0	0.84	0.48	0.61	89
1	0.62	0.71	0.66	203
2	0.83	0.16	0.28	91
3	0.78	0.47	0.59	114
4	0.61	0.96	0.75	690
5	0.43	0.15	0.22	125
6	0.88	0.32	0.47	158
7	0.95	0.19	0.31	112
Accuracy	-	-	0.64	1582

Table 5.10: Classification Report for Support Vector Classifier (SVC)

Class	Precision	Recall	F1-Score	Support
0	0.83	0.75	0.79	89
1	0.77	0.70	0.74	203
2	0.67	0.52	0.58	91
3	0.83	0.87	0.85	114
4	0.80	0.91	0.85	690
5	0.62	0.46	0.53	125
6	0.69	0.68	0.68	158
7	0.74	0.65	0.70	112
Accuracy	-	-	0.77	1582

Table 5.11: Classification Report for XGBoost Classifier

Class	Precision	Recall	F1-Score	Support
0	0.89	0.76	0.82	89
1	0.79	0.74	0.76	203
2	0.80	0.65	0.72	91
3	0.86	0.83	0.84	114
4	0.78	0.94	0.85	690
5	0.59	0.41	0.48	125
6	0.78	0.66	0.71	158
7	0.96	0.59	0.73	112
Accuracy	-	-	0.79	1582

Table 5.12: Classification Report for Logistic Regression Classifier

Class	Precision	Recall	F1-Score	Support
0	0.87	0.88	0.87	89
1	0.82	0.84	0.83	203
2	0.79	0.75	0.77	91
3	0.92	0.86	0.89	114
4	0.85	0.87	0.86	690
5	0.56	0.54	0.55	125
6	0.77	0.75	0.76	158
7	0.75	0.71	0.73	112
Accuracy	-	-	0.81	1582

Table 5.13: Average Classification Report for all models

Model	Precision (Avg)	Recall (Avg)	F1-Score (Avg)	Accuracy
Voting Classifier	0.83	0.83	0.83	0.83
Ridge Classifier	0.80	0.79	0.78	0.79
Extra Tree Classifier	0.68	0.64	0.61	0.64
SVC	0.74	0.77	0.74	0.77
XGBoost	0.80	0.79	0.78	0.79
Logistic Regression	0.79	0.81	0.80	0.81

The evaluation of classification models across various metrics, including precision, recall, and F1-score, offers valuable insights into their performance. The **Voting Classifier** demonstrates the highest overall accuracy of **83%** and balanced performance across different classes, particularly excelling in malignant and benign classification tasks. Notably, it maintains strong precision and recall for challenging classes like **malignant papillary carcinoma** and **benign fibroadenoma**, reflecting its effectiveness in handling class imbalances.

Comparatively, the **Logistic Regression** model achieves a competitive accuracy of **81%** with robust F1-scores across most classes, though it slightly underperforms in detecting rarer malignancies. The **XGBoost** classifier also exhibits a strong performance with an accuracy of **79%**, benefiting from its efficient learning capabilities and handling of complex patterns. However, it struggles with lower recall for minority classes, leading to possible misclassifications.

The **SVM** and **Ridge Classifier** models report moderate accuracies of **77%** and **79%**, respectively. While SVM shows reliability in classifying majority classes, its recall for benign tumors is lower. Similarly, the Ridge Classifier's performance is affected by a reduced F1-score in minority class detection, though it achieves consistent results across balanced datasets. The **Extra Tree Classifier** has the lowest accuracy at **64%**, mainly due to overfitting tendencies and misclassifications in less represented classes.

Overall, the **Voting Classifier**'s ensemble approach successfully mitigates the weaknesses of individual models by leveraging their strengths, resulting in superior classification performance. This makes it an ideal choice for applications requiring high accuracy and reliable cancer detection in multiclass classification scenarios.

CHAPTER 6

CONLUSION AND FUTURE WORK

Our research introduces a comprehensive solution for breast cancer image classification by effectively combining the strengths of transfer learning, ensemble learning, and traditional machine learning methods. By segmenting images into distinct regions and leveraging the feature extraction capabilities of pre-trained models like ResNet50, Xception and VGG16, we capture intricate image features. To further enhance regional analysis, we apply an Extra Tree Classifier, enabling detailed examination. Additionally, our ensemble learning framework, comprising XGBoost Classifier, SVM, Ridge Classifier, Extra Tree Classifier, and Logistic Regression, ensures a more resilient and interpretable model, leading to improved classification accuracy. Extensive experimental evaluations demonstrate the superiority of our approach over individual classifiers. This study addresses the need for heightened accuracy in medical image analysis and enhanced interpretability, offering a cohesive solution for breast cancer image classification. Future research can explore the scalability and adaptability of this method across diverse datasets, promoting the development of more reliable and efficient diagnostic tools for breast cancer detection.

REFERENCES

[1] Salini S. Nair and M. Subaji, "Automated Identification of Breast Cancer Type Using Novel Multipath Transfer Learning and Ensemble Classifiers," *2023.*

[2] W. Arshad et al., "Cancer Unveiled: A Deep Dive Into Breast Tumor Detection Using Cutting-Edge Deep Learning Models," in *IEEE Access*, vol. 11, pp. 133804–133824, 2023.

[3] Mahati Munikoti Srikantamurthy, V. P. Subramanyam Rallabandi, Dawood Babu Dudekula, Sathishkumar Natarajan, Junhyung Park, "Classification of Benign and Malignant Breast Cancer Histopathology Imaging Subtypes Using Hybrid CNN-LSTM Based Transfer Learning," *BMC Med Imaging*, vol. 23, no. 19, 2023.

[4] Rana M. and Bhushan M., "Classifying Breast Cancer Using Transfer Learning Models Based on Histopathological Images," *Neural Comput & Applic*, vol. 35, pp. 14243–14257, 2023.

[5] Weimin Wang, Min Gao, Mingxuan Xiao, Xu Yan and Yufeng Li, "Breast Cancer Image Classification Method Based on Deep Transfer Learning," *Electrical Engineering and Systems Science*, 2024. https://doi.org/10.48550/arXiv.2404.09226

[6] Prakash R., Meena K., Ramalakshmi K., Thayammal S., Shantha Selva Kumari R., Henry Selvaraj, "Classification of Breast Tumor from Histopathological Images with Transfer Learning," In *Computational Intelligence in Medical Decision Making and Diagnosis,* CRC Press, pp. 113–129, 2023.

[7] Folorunso Sakinat Oluwabukonla, Joseph Bamidele Awotunde, Y. Pandu Rangaiah, Roseline Oluwaseun Ogundokun, "EfficientNets Transfer Learning Strategies for Histopathological Breast Cancer Image Analysis," *International Journal of Modeling, Simulation, and Scientific Computing*, pp. 2441009, 2023.

[8] Joshi Shubhangi A., Anupkumar M. Bongale, P. Olof Olsson, Siddhaling Urolagin, Deepak Dharrao, Arunkumar Bongale, "Enhanced Pre-Trained Xception Model Transfer Learned for Breast Cancer Detection," *Computation*, vol. 11, no. 3, pp. 59, 2023.

[9] Mani R. K., J. Kamalakannan, "Technique for Breast Cancer Classification Using Semi-Supervised Deep CNNs with Transfer Learning Models," *Current Science*, vol. 125, no. 9, 2023.

[10] Zhong Wu, Hong Zhu, Lili He, Qiang Zhao, Jing Shi, Wenhuan Wu, "Real-Time Stereo Matching with High Accuracy via Spatial Attention-Guided Upsampling," *Appl Intell*, vol. 53, pp. 24253–24274, 2023.

[11] Y. Li, X. Zhang, and H. Chen, “Deep transfer learning for breast cancer classification in mammographic images,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 1125–1134, 2022.

[12] S. Gupta, R. Singh, and M. Verma, “An ensemble deep learning framework for early detection of breast cancer from mammograms,” IEEE Access, vol. 10, pp. 24075–24084, 2022.

[13] A. Sharma, P. K. Sinha, and D. Kumar, "Hybrid CNN-RNN architecture for breast cancer detection using ultrasound images," in Proc. IEEE Int. Conf. on Image Processing, pp. 325–329, 2022.

[14] M. O. Mohammed, L. A. Youssef, and T. M. Ali, "Mammographic mass classification using transfer learning and feature fusion," IEEE Transactions on Medical Imaging, vol. 40, no. 11, pp. 2837–2845, 2021.

[15] F. R. da Silva, P. M. Oliveira, and L. F. Santos, "Automated breast cancer diagnosis using pre-trained deep neural networks," Computer Methods and Programs in Biomedicine, vol. 208, 106245, 2022.

[16] J. Park, H. Kim, and D. Lee, "Feature extraction using transfer learning for breast cancer histopathology image classification," IEEE Access, vol. 9, pp. 20035–20043, 2021.

[17] R. Kumar, S. Kumar, and A. Jain, "A comparative study on breast cancer detection using deep learning models and radiomic features," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1975–1984, 2020.

[18] L. Wang, Y. Wang, and X. Wu, "Optimized deep learning framework for automated breast cancer detection in histopathological images," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 6, pp. 2451–2460, 2021.

[19] E. M. Abdelmaseeh, M. R. Mahmoud, and H. R. El-Bakry, "Breast cancer diagnosis through transfer learning and multi-scale feature analysis," IEEE Access, vol. 8, pp. 58945–58953, 2020.

[20] K. T. Nguyen, P. T. Tran, and L. Q. Le, "Automated classification of breast cancer using convolutional neural networks and transfer learning," in Proc. IEEE Int. Conf. on Healthcare Informatics, pp. 145–150, 2022.

[21] J. Kang, Z. Ullah, and J. Gwak, "Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers," Sensors,vol. 21, no. 6, p. 2222, 2021

[22]S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino, and M. Buckland, "A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis," IEEE access, vol. 4, pp. 9145–9154, 2016.

[23]M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in International conference on machine learning. PMLR, 2015, pp. 97–105.