DESIGN AND EVALUATION OF IMAGE PROCESSING MODEL USING MACHINE LEARNING APPROACHES

Thesis submitted to Delhi Technological University

in partial fulfilment of the requirements

for the award of the degree of

Doctor of Philosophy

in

Information Technology

Submitted By

Allam Venkata Siva Swetha
(2k19/PhDIT/03)

Under the Joint Supervision of

Prof. Kapil Sharma　　　　　　　　　　**Dr. Manju Bala**
Department of Information Technology　　**Department of Computer Science**
Delhi Technological University, Delhi　　**IndraPrastha College for Women,**
　　　　　　　　　　　　　　　　　　　　University of Delhi, Delhi

Department of Information Technology
Delhi Technological University
(Formerly Delhi College of Engineering)
New Delhi – 110 042, India
November 2024

Date: _____

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Design and Evaluation of Image Processing Model Using Machine Learning Approaches"** submitted by Ms. Allam Venkata Siva Swetha (Roll No. 2K19/PhDIT/03) in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy**, to the Delhi Technological University, Delhi, India is carried out by the candidate under our supervision and guidance at the Department of Information Technology, Delhi Technological University, Delhi, India.

The results embodied in this thesis have not been presented to any other University or Institute for the award of any degree or diploma.

Prof. Kapil Sharma
Department of Information Technology
Delhi Technological University, Delhi

Dr. Manju Bala
Department of Computer Science
IndraPrastha College for Women,
University of Delhi, Delhi

Date: _____

CANDIDATE DECLARATION

I hereby declare that the thesis entitled "**Design and Evaluation of Image Processing Model Using Machine Learning Approaches**" submitted to Delhi Technological University, Delhi, in the partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy** in the Department of Information Technology, is an original work and has been done by myself under the Joint Supervision of **Prof. Kapil Sharma** Department of Information Technology, Delhi Technological University, Delhi, India and **Dr. Manju Bala,** Department of Computer Science, Indraprastha college for Women, Delhi, India.

I also mention that the research work is original and has not been submitted by me, in part or full, to any other University or Institution for the award of any degree or diploma.

Allam Venkata Siva Swetha
2K19/PhDIT/03
Department of Information Technology
Delhi Technological University
Delhi-110042, India

Date:
Place: **Delhi**

Dedicated to My Parents: for their love, endless support,

encouragement, and unwavering cooperation throughout

this entire journey.

ACKNOWLEDGMENT

I address my sincere thanks to Almighty God for giving me the inner power to complete my thesis and guide me in every step of my life. It is an immense pleasure to have the opportunity to express my heartiest gratitude to everyone who helped me throughout this research journey.

With immense joy and heartfelt gratitude, I would like extend my indebtedness to my supervisor's, **Prof. Kapil Sharma, Dept. of Information Technology, Delhi Technological University, Delhi and Dr. Manju Bala, Associate Professor, Indraprastha college for women, Delhi** for their invaluable guidance, mentorship, encouragement, and patience. During my research tenure, their motivation and encouragement have made me strive to work harder to achieve my goals. I am deeply humbled and indebted to my supervisor's for continually motivating me to persevere and making me believe in myself during the times of hardships.

Also, my sincere thank goes to Delhi Technological University for considering my candidature for this course. I am very thankful to **Prof. Prateek Sharma**, Vice-Chancellor, Delhi Technological University, Delhi, India, who has been a constant source of enthusiasm. Also, my sincere thanks reciprocate to **Prof. Dinesh Kumar Vishwakarma** HoD, Dept. of Information Technology, **Dr. Virender Ranga**, Dept. of Information Technology for insightful comments and valuable suggestions. Special thanks to my seniors and colleagues of Delhi Technological University, Delhi, India. My sincere thanks to all the professors, faculty, research scholars and non-teaching staff of the Information Technology Department.

I am deeply thankful to all my colleagues and friends during my journey as a Ph.D. scholar. The engaging discussions, brainstorming sessions, and collaborative teamwork significantly impacted my growth as an independent researcher. I would like to express my deepest gratitude to my parents **Prof. A. Srinivasa Rao, HoD, Department of Applied Physics**, Delhi Technological University, Mrs. **A. Padmavathi**, who stood by me like a pillar of strength and always supported me to realize my goals. I will cherish their utmost love and blessings throughout my life. Special thanks to my husband **Mr. D. Hemanth Kumar**, my brothers,

Abstract

This research presents a comprehensive exploration of hybrid deep learning architectures designed to address class imbalance and generalization problems in image classification, focusing on breast cancer and brain tumor diagnosis while tackling challenges like data scarcity, imbalance, and feature inconsistency. Key contributions include novel hybrid architectures integrating CNNs, ViTs, and GANs, enhancing robustness and adaptability. A dual-modification approach combining data augmentation with algorithmic adjustments, such as optimized loss functions, effectively balanced class representation. Additionally, the incorporation of GANs and auxiliary neural networks for tumor classification demonstrates a substantial increase in diagnostic performance by generating diverse synthetic data and using auxiliary spatial features. The introduction of an Efficient Attention Mechanism and a Resource-Efficient Optimization model further refines breast cancer detection, providing high-dimensional feature integration that enhances diagnostic precision while reducing computational overhead. The proposed models demonstrated improved performance by effectively addressing class imbalance and generalization, achieving 9% to 10% gains in multi-class classification and 1% to 2% gains in binary classification. Additionally, the model complexity, in terms of time and space, was reduced by 2% to 4%, while scalability improved by 1% to 2% in binary tasks and 9% to 10% in multi-class tasks, highlighting their efficiency and adaptability. Through the extensive evaluation this research establishes a robust, generalizable framework for image classification, with future implications for integrating multi-modal data and advancing interpretability in clinical settings.

List of Publications

Journals:

1. **Swetha, A. V. S.**, Bala, M., & Sharma, K. (2024). A linear time shrinking-SL(T)-VIT approach for brain tumor identification and categorization. IETE Journal of Research, 1–23. https://doi.org/10.1080/03772063.2024.2370954. **Impact factor: 1.3. [Published]**

2. **Swetha, A. V. S.**, Bala, M., & Sharma, K. (2024). A framework for classifying breast cancer via heterogenetic attention mechanism and optimized feature selection. Intelligent Data Analysis, 1–36. https://doi.org/10.3233/ida-240334, **Impact factor:0.9. [Published]**

3. **Swetha, A.V.S.**, Bala, M., Sharma, K. An Elucidation of Hybridized Architectures for Mitigating Class Imbalance in Breast Cancer Diagnosis: A Comprehensive Exploration of Application Strategies, The Journal of supercomputing (2024), **Impact factor:2.5. [Communicated]**

4. **Swetha, A.V.S**., Bala, M., Sharma, K. Brain tumor classification using data generation and auxiliary feature network, Multimedia tools and applications (2024), **Impact factor:2.7. [Communicated]**

5. **Swetha, A.V.S**., Bala, M., Sharma, K. A Hybrid Deep Learning Framework for Accurate Multi-Class Image Classification: Innovations in Generalization and Occlusion Handling", Expert Systems, **Impact Factor: 3. [Communicated]**

International Conferences:

1. **Swetha, A. V.**, Bala, M., & Sharma, K. (2023). A novel multi features deep learning architecture for breast cancer detection using loss function. Lecture Notes in Networks and Systems, 751–763. https://doi.org/10.1007/978-981-99-3716-5_60.

2. **Swetha, A. V. S.**, Bala, M., & Sharma, K. (2024c). Optimized hybrid model framework for Breast Cancer Classification. 2024 14th International Conference on Cloud Computing, Data Science &Amp; Engineering (Confluence), 882–887. https://doi.org/10.1109/confluence60223.2024.10463295

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
AutoML	Automated Machine Learning
DL	Deep Learning
CAD	Computer Aided Design
TL	Transfer Learning
CNN	Convolutional Neural Network
AE	Auto Encoders
ViT	Vision Transformers
MLP	Multi-layer perceptron
En-DL	Ensemble Deep Learning
DBA	Distance Based Approach
EM	Ensemble Method
GRU	Gated Recurrent Unit
MAE	Mean Absolute Error
MSE	Mean Square Error
MV-DNN	Multi-View Deep Neural Network
NN	Neural Network
POS	Point-of-Sale
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SDAE	Stacked Denoising Auto encoder
VAE	Variational Auto Encoders

Chapter One: Introduction

This chapter introduces the concept of class imbalance and generalization in image processing. The objectives of the research work are highlighted. Chapter wise thesis coverage is summarized at the end of the chapter.

1.1 Background Study

In Today's modern society, Image processing has emerged as a vital tool in various fields, significantly enhancing the way, we analyse and interpret data. Its popularity stems from its ability to transform raw images into valuable insights, facilitating applications ranging from facial recognition to autonomous driving [1]. As technology evolves, image processing techniques are becoming increasingly sophisticated, allowing for the extraction of intricate patterns and features that were previously difficult to understand. In recent years, the application of machine learning (ML) techniques has further advanced the capabilities of image processing. ML algorithms can learn from data, improving their performance over time and making it possible to automate complex image analysis tasks [2]. This innovation has led to remarkable advancements in various domains, including healthcare, autonomous driving, satellite imaging etc.

However, despite these advancements, employing ML techniques for image processing currently faces significant challenges. One of the most notable issues is class imbalance, a situation where one class contains disproportionately high number of samples compared to another. For example, in object detection tasks, common objects often dominate the dataset, while rare objects may be severely underrepresented [3]. As a result, ML models may become biased, favoring the majority class and underperforming in identifying critical minority-class

instances. Various applications grapple with the challenges posed by class imbalance, leading to biased models that fail to accurately predict all classes effectively [4].

Moreover, a lack of generalization in these models poses additional challenges. Models that are not generalized enough may perform well on training data but struggle with unseen data, leading to overfitting. This situation is particularly problematic in real-world applications where data can vary significantly due to environmental changes, lighting conditions, or variations in object appearance. When a model fails to generalize, it may misidentify or overlook critical features, which can result in significant consequences, such as inaccurate object detection or flawed analysis [5]. Therefore, addressing both class imbalance and generalization issues is crucial to enhance the reliability and effectiveness of machine learning applications in image processing, ensuring these systems provide accurate and equitable results across diverse contexts.

1.2 Overview of Image Processing and Its Types

Image processing involves the manipulation of digital images to enhance their quality or extract valuable information. This process encompasses a series of operations applied to an image, transforming it into a desired format or improving its visual attributes. Image processing is fundamental in various fields, including medical imaging, remote sensing, and computer vision, where it facilitates the analysis and interpretation of visual data. The emergence of digital technologies has significantly transformed image processing, rendering it more efficient and versatile than ever [6]. Image processing can be classified into two primary categories: analog and digital.

1.2.1 Analog Image Processing

This type involves manipulating images in their continuous form using various physical mechanisms. Techniques such as filtering and noise reduction are employed, but analog

methods generally lack the precision and flexibility offered by digital approaches. Analog image processing techniques utilize physical devices to manipulate continuous signals that represent images. Common operations include enhancement through analog filtering, which may involve optical filters or chemical processes applied to photographic images [7]. However, these techniques are constrained by their limited precision and repeatability. As digital processing evolved, reliance on analog methods diminished. Nevertheless, a foundational understanding of analog processing is essential for comprehending the principles that inform contemporary digital methods, thereby providing insight into the technological evolution in the imaging field.

1.2.2 Digital Image Processing

Digital image processing involves using computer methods to manipulate digital images, modifying the original pixel values to create a new output. A digital image is represented mathematically as a two-dimensional function, where the coordinates correspond to specific locations and the intensity or gray level of the image is represented by amplitude values [8]. These images consist of pixels, which are the smallest elements, each linked to a specific spot and value. Modern imaging technologies, such as ultrasound and electron microscopy, produce images from sources that are often invisible to the naked eye, extending beyond what humans can naturally perceive. This has made digital image processing a vast field with many practical applications.

Advancements in digital technology now allow for the manipulation of multi-dimensional signals using various systems, from simple circuits to complex computing architectures. Digital image processing techniques can significantly enhance performance through tasks like image enhancement, restoration, compression, and segmentation. One notable feature of these systems is their ability to focus on specific parts of an image, known

as regions of interest (ROIs) [9]. For example, different areas of an image can be processed to reduce blur or improve color accuracy. While both analog and digital image processing have their own methods and uses, digital image processing has become more widely used due to its flexibility and extensive range of applications.

1.3 Artificial Intelligence

Artificial Intelligence (AI) is a dynamic branch of computer science focused on developing systems and machines that can mimic human intelligence. These systems perform tasks such as pattern recognition, language processing, problem-solving, and decision-making, all of which typically require cognitive abilities. AI utilizes algorithms and computational models to analyze large datasets, learn from them, and make informed predictions or decisions. The ultimate goal is to create autonomous machines that can think, learn, and adapt, thereby expanding the possibilities of technology. AI is already transforming various industries, from healthcare diagnostics to autonomous vehicles and smart assistants, reshaping daily life [10].

Machine Learning (ML) is a vital subset of AI that enables machines to learn from data without explicit programming for each task. By employing statistical techniques, ML systems identify patterns in data and enhance their performance over time. For example, recommendation systems on platforms like Netflix and Amazon analyze user behavior to suggest personalized content. ML finds applications in numerous fields, including fraud detection, financial forecasting, and image recognition [11]. A more specialized area of ML is Deep Learning (DL), which utilizes artificial neural networks modeled after the human brain. These networks consist of multiple layers of interconnected nodes (neurons) that process data in increasingly abstract ways. DL has driven advancements in speech recognition, image classification, and natural language processing, powering technologies like voice assistants. Together, AI, ML, and DL represent a hierarchy of technologies that are revolutionizing

machine interactions with the world, automating tasks and addressing complex problems, with their impact expected to grow as research progresses. AI has emerged as a transformative force, fueling innovation across various sectors [12].

1.3.1 Machine Learning

Machine learning (ML) is a branch of artificial intelligence (AI) that focuses on developing algorithms and statistical models enabling machines to improve their performance through experience and data. Unlike traditional programming, where explicit instructions dictate task execution, ML allows computers to learn patterns and relationships directly from the data they process. This capability enables machines to make predictions or decisions, applying insights gained from past examples to new situations. The power of ML lies in its ability to generalize from observed data, adapting and refining performance over time without constant human intervention.

At the core of machine learning are algorithms that optimize model parameters to minimize errors in predictions or decisions. These algorithms iteratively adjust model weights based on input data and feedback, striving to accurately map inputs to outputs. A key concept in ML is the loss function, which measures the difference between the model's predictions and the true values. The goal is to minimize this loss, improving accuracy through optimization techniques such as gradient descent. However, it's essential to maintain a balance in model complexity; overfitting occurs when a model becomes too closely aligned with the training data, capturing noise rather than true patterns, while under fitting happens when a model is too simplistic to grasp the data's complexity. Techniques like cross-validation, regularization, and pruning are employed to avoid these pitfalls, ensuring that the model can learn essential patterns without being overly sensitive to specific training data [13].

Lastly, scalability is crucial in machine learning. As datasets grow in size and complexity, models must handle increasing amounts of data without sacrificing performance.

5

Distributed computing and advanced optimization techniques enable efficient training on large datasets, while managing computational complexity ensures feasibility in real-world scenarios where data continuously evolves [14]. Overall, machine learning revolves around creating systems that autonomously learn from data, optimize performance, and generalize to new situations, continuously evolving as data complexity increases. Through iterative learning and refinement, ML systems become more accurate and reliable, pushing the boundaries of data-driven technologies.

1.3.2 Types of Machine Learning Algorithms

Machine learning employs a diverse array of techniques to tackle data-related challenges. Data scientists emphasize that no single solution exists for every problem; the chosen approach depends on several factors, including the nature of the issue, the number of variables, and the most suitable model for the task [15]. Figure 1.1 shows different types of machine learning algorithms which are explained in further subsections.
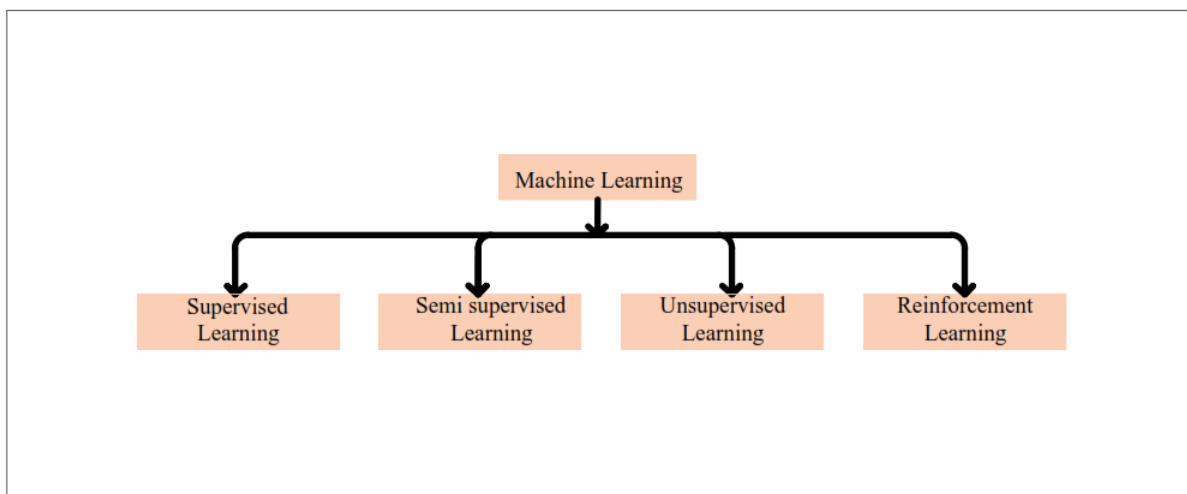
Figure 1.1 Type of Machine Learning Algorithms

A. Supervised Learning

Supervised learning is an approach in which models are trained on labelled data to achieve desired outcomes. The input dataset is typically divided into two parts: training data

and testing data. Each labelled training example consists of an input and its corresponding target output. The objective is to predict or classify an output variable in the test dataset based on patterns learned from the training set. Algorithms employed in supervised learning discern relationships between input variables and target outputs, thereby enhancing prediction accuracy. A common example of this method is regression, which is utilized to automatically fit models [16].

B. Unsupervised Learning

Unsupervised learning involves analysing and identifying patterns in data without predefined labels [17]. In this approach, models learn from the inherent structure of the data to recognize trends or classify information based on features. Key applications include feature reduction and clustering. Unsupervised learning identifies natural groupings in data without specifying an output variable. Techniques such as principal component analysis (PCA) are employed to uncover hidden relationships or patterns of covariance within the data [17].

C. Semi-Supervised Learning

Semi-supervised learning combines elements of both supervised and unsupervised techniques. This approach typically involves a small set of labelled data alongside a larger quantity of unlabelled data. The initial phase uses unsupervised learning to cluster similar data points, which then aids in labelling the unlabelled data [18]. This method is particularly beneficial when predictions are desired but most data points lack clear outcome information.

D. Reinforcement Learning

Rooted in early cybernetics, reinforcement learning has influenced multiple fields, including statistics, psychology, neuroscience, and computer science. Recently, it has garnered significant attention within the machine learning and artificial intelligence communities. Unlike traditional methods that rely on predefined datasets, reinforcement learning systems learn through interaction with their environment, receiving feedback in the form of rewards or

penalties. This feedback mechanism allows the system to refine its decision-making processes over time. Reinforcement learning is particularly advantageous for online data collection and monitoring, where the system adjusts its actions to maximize rewards [18]. This approach enables an agent to learn how to act within an environment by continuously adapting based on the rewards and penalties received while pursuing a defined objective.

1.4 Generalisation

Generalization is a fundamental concept in machine learning, referring to a model's ability to perform effectively on new, unseen data that wasn't part of its training set. The primary objective of any machine learning model is to learn from the available data in a way that enables accurate predictions on new inputs. A well-generalized model captures the essential trends and patterns of the data instead of merely memorizing specific examples from the training set. When a model becomes too closely aligned with the training data, it may start to learn noise, leading to a phenomenon known as overfitting [19]. An over-fitted model may excel on the training dataset but struggle with new data, as it fails to account for variations and unseen instances. In contrast, a model that generalizes effectively maintains its performance across different datasets and real-world scenarios. Achieving strong generalization is vital for the practical deployment of machine learning models, as it ensures the model can adapt to the inherent variability of data encountered in real-world situations. Therefore, evaluating and improving generalization capabilities are essential components of the machine learning model development process [20].

1.4.1 Steps to Evaluate Model Generalization

Evaluating model generalization is essential for determining a model's effectiveness on unseen data that are not included in the training set. This evaluation is crucial for real-world applications, where the model encounters diverse variations [21]. The evaluation process involves analysing performance metrics and employing various techniques to assess how well

the model can adapt to new inputs, ensuring its reliability and effectiveness in practical scenarios. Below are the steps to evaluate model's generalization:

Step 1*: Train-Test Split:* First, the dataset should be divided into two distinct subsets: a training set and a testing set. Typically, 70-80% of the data is used for training, while 20-30% is reserved for testing. This split allows for assessing the model's performance on unseen data, providing a clear indicator of its generalization capabilities [22].

Step 2: *Cross-Validation:* Next, cross-validation techniques, such as k-fold cross-validation, can be employed to obtain a more reliable estimate of the model's performance and mitigate overfitting. In this approach, the dataset is divided into 'k' subsets, where the model is trained on 'k-1' folds and validated on the remaining fold [23]. This process is repeated for each fold, allowing every data point to be used for both training and validation, which helps ensure that the evaluation is not overly reliant on any specific subset.

Step 3: *Performance Metrics:* Using various performance metrics is essential to evaluate the model's effectiveness on the test set. Common metrics include accuracy, precision, recall, F1 score, and ROC-AUC. These metrics provide insights into the model's performance across different aspects, such as its ability to identify relevant instances and distinguish between classes. Plotting learning curves can also be beneficial, as they visualize the training and validation loss or accuracy over time. By analyzing these curves, one can identify signs of overfitting or under fitting, enabling further adjustments to the model as needed [24].

Step 4: *Hyper parameter Tuning:* Hyper parameter tuning is another critical step in evaluating generalization. Techniques like grid search or randomized search can be utilized to optimize the model's parameters for better performance [25]. Hyper parameters significantly influence the model's ability to generalize, so systematically exploring combinations of hyper parameters can help identify optimal settings.

Step 5: *Regularization Techniques:* Additionally, implementing regularization techniques, such as L1 (Lasso) or L2 (Ridge) regularization, can prevent overfitting by constraining the model complexity [26]. Dropout, a technique used in neural networks, randomly disables a fraction of neurons during training to promote more robust feature learning [27].

Step 6: *Ensemble Methods:* Ensemble methods, such as bagging and boosting, can also enhance generalization by combining predictions from multiple models. These techniques reduce variance and bias, resulting in a more robust final model. Common ensemble methods include Random Forests (bagging) and Gradient Boosting Machines (boosting), both of which have proven effective in various tasks [28].

Step 7: *Testing on Diverse Datasets:* Finally, testing the model on diverse datasets or in real-world conditions is vital to ensure it generalizes across different scenarios. This step validates the model's performance in practical applications, confirming its reliability and robustness when faced with new data [29].

By following these steps, you can effectively assess and enhance the generalization capabilities of the proposed models, ensuring they are well-suited for deployment in real-world applications.

1.5 Class Imbalance

In the context of the Fourth Industrial Revolution, deep learning and machine learning have garnered significant interest from the scientific community due to their wide-ranging applications, including smart energy management, social network analysis, business intelligence, medical informatics, computer vision, software management, and structural engineering. Despite extensive training, recent observations reveal a decline in the performance of machine learning models when applied to real-world situations. Some researchers attribute this issue to deficiencies in the modeling and training processes, while others emphasize the

importance of model robustness, particularly regarding generalization to varying data distributions [30]. Additionally, the quality of the dataset is critical, as high-quality data is essential for maintaining performance, especially since neural networks often require more extensive training data compared to traditional methods.

A significant challenge impacting dataset quality is class imbalance, which arises when certain classes have a disproportionately higher number of samples than others. This problem is common across various tasks, including classification, object detection, and image segmentation, and has gained increasing attention in real-world scenarios. The implications of class imbalance have become more pronounced as machine learning techniques evolve across industries, compounded by noisy data and the high costs of labeling, which further diminish performance when models are deployed [31]. Class imbalance can adversely affect various applications, such as financial fraud detection, bankruptcy prediction, medical decision-making, fault diagnosis, pattern recognition, cancer gene expression analysis, and telecommunications fraud. Recent research has proposed several techniques to mitigate this issue, particularly through data augmentation in deep learning models, which generates synthetic data to increase the number of labeled samples, thus enhancing model performance. Addressing class imbalance involves understanding the distribution of samples in the dataset. Even when disparities exist, satisfactory results can still be achieved if both classes are adequately represented and drawn from non-overlapping distributions [32]. Therefore, accurately capturing the majority and minority classes is essential for improving model performance. While class imbalance in single-label classification has been extensively studied, multi-label classification presents additional challenges. In scenarios where instances can be associated with multiple labels, there is a heightened risk of bias towards the majority class, necessitating innovative approaches to enhance performance in these complex situations.

1.5.1 Evaluation Metrics and Methods for Assessing Class Imbalance

Before constructing a robust model, it is crucial to understand the properties of the dataset, particularly the degree of class imbalance. A thorough analysis of these properties enables informed decisions regarding model selection and training strategies [33]. The process of assessing class imbalance involves several systematic steps aimed at quantifying the disparities between the frequencies of different classes.

Step 1: Basic Dataset Analysis

The initial step in addressing class imbalance involves conducting a basic analysis of the dataset. This includes evaluating the number of samples within each class, as well as the total number of distinct classes and subclasses [34]. Let N represent the total number of samples in the dataset, C denote the number of distinct classes, and N_i signify the number of samples in class. By employing equation (1.1) analysis can be performed.

$$N = \sum_{i=0}^{C} N_i \tag{1.1}$$

Understanding this information helps identify underrepresented classes, which adds complexity to the classification task. For example, a dataset containing numerous subclasses may introduce additional layers of imbalance, complicating the model's learning process. By calculating the number of samples per class, we gain initial insights into the level of imbalance, guiding further analysis.

Step 2: Class Distribution Analysis

Once basic statistics are gathered, the next step is to analyse class distribution. This can be achieved by visualizing the distribution of classes through bar or pie charts, which help highlight significant disparities between majority and minority classes [35]. To quantify these differences, we can calculate the proportion of samples in each class relative to the total number of samples using equation (1.2):

$$P_i = \frac{N_i}{N} \qquad (1.2)$$

where P_i is the proportion of class i.

A comprehensive understanding of class distribution is vital, as it influences decisions related to model selection and pre-processing techniques, ultimately affecting model performance.

Step 3: Class Relationship Examination

Step three involves examining the relationships between classes, particularly relevant in multi-label datasets where certain classes frequently co-occur. Identifying unique and frequently co-occurring classes enhances our understanding of the dataset's underlying structure. This examination also reveals which labels are more prone to confusion, guiding feature selection and model optimization strategies [36]. A thorough analysis of these relationships strengthens our ability to develop more effective classification models.

Step 4: Calculating the Level of Imbalance

In step four the level of Imbalance is calculated to quantify the overall percentage of imbalance within the dataset. This can be achieved using several evaluation metrics, each providing unique insights into the degree of class disparity present. Understanding these metrics allows for improved model training and more informed decision-making [37].

A. *Imbalance Ratio per Label (IRLbl):* The Imbalance Ratio per Label (IRLbl) measures the extent of class imbalance by comparing the sample count of the majority class $N_{majority}$ to that of the minority class $N_{minority}$. This ratio is calculated by employing equation (1.3) provides insight into the level of skewness in the data distribution, highlighting any disparities between class representations.

$$IRLbl = \frac{N_{majority}}{N_{minority}} \qquad (1.3)$$

13

A high IRLbl value indicates a significant disparity between the majority and minority classes, which may lead to model bias towards the majority class [38].

B. *Mean Imbalance Ratio (MeanIR):* The Mean Imbalance Ratio (MeanIR) provides an average view of imbalance across all classes in the dataset. By employing the following equation (1.4) the mean of all individual class imbalance ratios can be calculated.

$$\text{MeanIR} = \frac{1}{C}\sum_{i=1}^{C}\text{IRLbl} \tag{1.4}$$

This metric is particularly useful for summarizing the overall state of class balance in multi-class problems, allowing for a quick assessment of the model's training environment [38].

C. *Maximum Imbalance Ratio (MAXIR):* The Maximum Imbalance Ratio (MAXIR) identifies the class exhibiting the greatest imbalance within the dataset. This metric highlights the most significant concerns regarding class distribution, enabling practitioners to focus on these problematic classes during model training. Equation (1.5) analyses how MAXIR ration can be calculated.

$$\text{MAXIR} = \max(\text{IRLbl}_i) \text{ for } i = 1 \text{ to } C \tag{1.5}$$

By analysing the class with the highest imbalance, strategies can be developed to address this specific issue, thereby improving overall model performance [39].

D. *Coefficient of Variation of IRLbl (CVIR):* The Coefficient of Variation of IRLbl (CVIR) calculated by employing equation (1.6), offers insights into the variation of imbalance ratios across different classes. It measures the extent of disparity among classes by calculating the ratio of the standard deviation (σ) of the imbalance ratios to the mean (μ) of the imbalance ratios [40].

$$\text{CVIR} = \frac{\sigma(\text{IRLbl})}{\mu(\text{IRLbl})} \tag{1.6}$$

This metric helps identify whether the imbalance is uniformly distributed across classes or concentrated in specific areas. A higher CVIR suggests greater variability in class imbalance,

indicating that certain classes may need more focused attention during model training. Evaluating the level of imbalance using metrics such as IRLbl, MeanIR, MAXIR, and CVIR provides critical insights into the structure of the dataset. These evaluations clarify the severity of class imbalance and inform subsequent modelling strategies [41]. By understanding these metrics, data scientists and machine learning practitioners can make more informed decisions, leading to the development of more robust and effective classification models. Systematically addressing class imbalance not only improves model performance but also enhances the model's generalization to real-world applications.
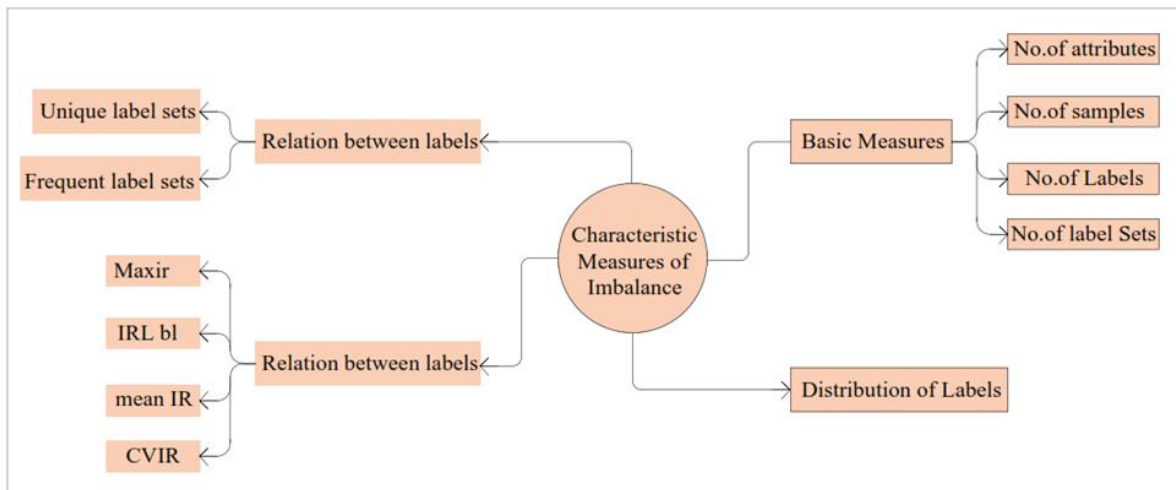
Figure 1.2 Characteristic Measures of Imbalance.

1.5.2 Strategies for Addressing Class Imbalance

After calculating the class imbalance percentage, the next step is to apply one of the three primary approaches to address the issue [42]. Each approach has its strengths and is suited to different contexts depending on the dataset's characteristics as shown in Figure 1.2.
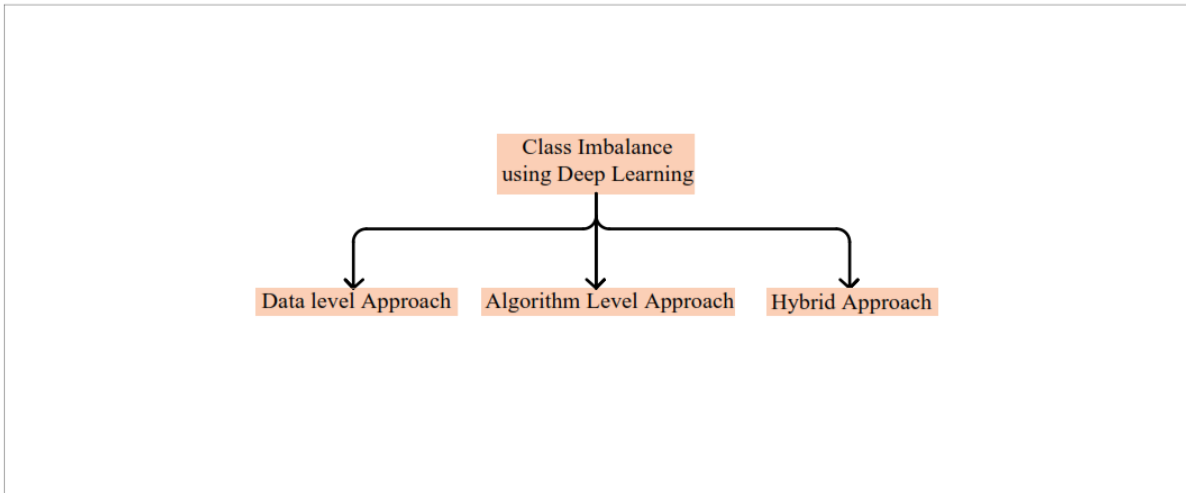
Figure 1.3 Approached to address Class Imbalance

Three primary approaches exist for tackling class imbalance are as follows:

1. Data-level Transformation Approach

2. Algorithmic Adaptation Approach

3. Hybrid Approach

These approaches as shown in Figure 1.3 aim to address the skewed distribution of data; wherein certain classes have a significantly higher number of samples than others. Below is a detailed examination of each approach, including relevant techniques and formulas [43].

1.5.2.1 Problem Transformation (Data-Level Approaches)

The first strategy, problem transformation, seeks to modify the distribution of imbalanced data using data-level techniques [44]. This method involves converting multi-label classifications into one or more single-label classifications to create a more balanced dataset. Key techniques in this category include Binary Relevance (BR), Label Power Set (LP), and Classifier Chains (CC).

Binary Relevance (BR) simplifies a multi-label problem by decomposing it into multiple independent binary classification tasks. Each label is treated as a distinct binary classifier,

allowing for straightforward implementation. The formula for a binary classifier hi for the i-th label is given as equation (1.7):

$$h_i(x) = \{0,1\}, for\ i = 1,2,\dots,n \qquad (1.7)$$

While effective for various applications, BR does not account for correlations between labels, potentially leading to suboptimal performance when such dependencies exist [45]. Label Power Set (LP) transforms the problem into a multi-class classification task by treating each unique combination of labels as a single class. This transformation can be represented as equation (1.8):

$$y_{new} = \{(y_1, y_2, \dots, y_n)\} \qquad (1.8)$$

for each unique combination of labels

Although LP captures label dependencies, it can result in a substantial increase in the number of classes, complicating the model training process [46].

Classifier Chains (CC) constructs a sequence of binary classifiers, where each classifier predicts a label based on input features and the predictions made by preceding classifiers in the chain. The prediction for the i-th label can be represented as equation (1.9):

$$h_i(x) = f_i(x, h_1(x), h_2(x), \dots, h_{i-1}(x)) \qquad (1.9)$$

This approach effectively captures label dependencies; however, it may suffer from error propagation, where mistakes in earlier predictions negatively affect subsequent classifications.

To modify the distribution of imbalanced data, various resampling techniques can be employed. Oversampling increases the number of samples in the minority class, often using methods like SMOTE (Synthetic Minority Over-Sampling Technique), which generates synthetic examples [47]. The formula for SMOTE is given in equation (1.10):

$$x_{new} = x_{minority} + \lambda \cdot (x_{neighbour} - x_{minority}) \qquad (1.10)$$

where λ is a random number between 0 and 1. Conversely, under-sampling reduces the number of samples in the majority class to achieve a more balanced distribution. While both techniques aim to equalize class representation, oversampling can lead to overfitting, and under-sampling may result in the loss of critical information.

1.5.2.2 Algorithmic Adaptation

The second approach, algorithmic adaptation, modifies the machine learning algorithms themselves to better handle class imbalance. This method focuses on reducing the inherent bias towards the majority class and enhancing the model's ability to recognize and prioritize minority classes [48].

Cost-Sensitive Learning assigns different penalties to misclassifications, imposing higher costs for errors related to the minority class. The modified loss function can be represented employing equation (1.11):

$$L(x, y) = \sum_{i=1}^{n} C_i \cdot I(y_i \neq yi) \qquad (1.11)$$

Here, C_i represents the cost associated with the I-th class, and I is an indicator function. This approach ensures that the model focuses more on the minority class by imposing greater penalties for misclassifications [49].

Ensemble Techniques involve combining multiple models to improve performance on imbalanced datasets. Methods like AdaBoost concentrate on misclassified instances by adjusting their weights. Additionally, Balanced Random Forests modify the bootstrap sampling process to ensure balanced data within each tree by under sampling the majority class [49].

1.5.2.3 Hybrid Techniques

The third approach combines both data-level and algorithm-level modifications to create a more comprehensive solution to class imbalance. Hybrid techniques leverage the strengths of both problem transformation and algorithmic adaptation to enhance model performance.

For example, combining SMOTE with a boosting algorithm like AdaBoost can bolster robustness against imbalanced data distributions. This hybrid method aims to balance the dataset while focusing on challenging instances that are more difficult to classify [50]. Another example is the integration of data augmentation techniques with cost-sensitive learning, ensuring adequate representation of minority classes while effectively penalizing misclassifications. By blending resampling strategies and algorithmic adjustments, hybrid techniques yield balanced and high-performing models suited for complex imbalanced datasets.

1.6 Class Imbalance in Healthcare: Challenges and Impacts

Class imbalance poses significant challenges in various fields, with particularly critical implications in healthcare. Machine learning models are essential for aiding diagnosis, treatment planning, and predicting patient outcomes. However, these models are often trained on datasets where certain conditions are disproportionately represented [51]. Common diseases frequently dominate the data (majority class), while rare but critical conditions, such as genetic disorders and specific cancers (minority class), remain underrepresented. This uneven distribution can lead to biased models that struggle to accurately predict or detect instances of minority classes, which are crucial for effective patient care.

The reliance on real-world data further complicates the issue of class imbalance in healthcare. Medical datasets reflect the natural occurrence of diseases within populations, inherently making them imbalanced [52]. For example, rare diseases account for a small percentage of overall medical cases, yet misclassifying these conditions can have severe consequences. Acquiring sufficient data for these minority classes is often costly and time-consuming, necessitating expert annotation from healthcare professionals, such as doctors or radiologists. Moreover, the quality of medical data can suffer from issues like noise, missing

values, and inconsistencies. These challenges make medical datasets more prone to imbalance compared to other domains, where techniques like resampling or synthetic data generation may be more feasible.

Following are some examples of how imbalanced datasets can impact automated diagnosis systems.

➤ *Disease Detection and Diagnosis*:

One of the most profound effects of class imbalance is evident in disease detection systems. For instance, in cancer screening, where the majority of patients are healthy, models tend to favor the majority class, resulting in higher rates of false negatives for rare diseases. This can lead to delays in critical diagnoses, adversely affecting patient outcomes [53]. A model biased toward healthy diagnoses may miss early detection opportunities for patients with less common diseases, illustrating how imbalance directly impacts the quality of care.

➤ *Personalized Treatment Recommendations*:

Machine learning models that offer personalized treatment options also suffer when trained on imbalanced data. If most patients in a dataset respond well to a standard treatment, models may overwhelmingly recommend this option, even for patients who could benefit from alternative therapies [54]. This is particularly concerning in oncology, where treatment effectiveness can vary widely among patients. An imbalance can cause models to overlook valuable treatment options for minority-class patients, undermining the potential of personalized medicine.

➤ *Prognostic Modeling*:

Predictive models used to forecast patient outcomes, such as disease progression or survival rates, are equally vulnerable to class imbalance. Often, high-risk patients are underrepresented in datasets, leading models to over fit majority-class outcomes. This can result in inaccurate predictions for patients with rare or complex conditions, especially when models are expected

to generalize to atypical cases. In medical contexts where prognosis informs patient care decisions, the implications of these prediction errors can be severe [55].

Addressing class imbalance is critical for developing effective machine learning models in healthcare. The underrepresentation of rare diseases and high-risk cases can lead to models that fail in critical situations, affecting diagnoses, treatment options, and patient prognoses. Therefore, it is essential to develop strategies that effectively tackle class imbalance while enhancing data quality and employing advanced modeling techniques that accommodate the unique challenges presented by medical datasets.

1.7 Cancer: The Leading Cause of Death Beyond Infectious Diseases

Globally, cancer ranks as a leading cause of death, surpassing diseases such as malaria or tuberculosis. The American Cancer Society (ACS) reported approximately 609,300 cancer deaths out of about 1,918,030 diagnosed cases in 2020, with projections suggesting this figure may exceed 28 million by 2030. Among the various cancer types, brain and breast cancer warrant significant attention due to their prevalence, profound impact on quality of life, and ongoing challenges in diagnosis and treatment [56], [57].

Breast cancer is the most commonly diagnosed cancer worldwide, representing approximately 12.5% of all new cancer cases and about 25% of all cancers in women. In 2020, an estimated 2.3 million women were diagnosed with breast cancer globally, with projections of about 281,550 new cases of invasive breast cancer in the U.S. for 2021. Conversely, brain tumors, while less common, are among the most aggressive cancers, accounting for about 1.5% of all new cancer cases. In the U.S., it was estimated that 24,530 new cases of brain and other central nervous system (CNS) tumors would occur in 2021 [58], [59], [60], [61], [62]. Although brain tumors are less prevalent, they are often diagnosed at advanced stages, leading to higher

mortality rates. The five-year survival rate for glioblastoma, a common type of brain cancer, is only about 5% [63], [64], [65].

1.7.1 Emotional and Economic Impact of Cancer

Breast cancer significantly affects patients' emotional and psychological well-being. Many survivors experience anxiety, depression, and body image issues, with studies indicating that the psychological burden of a breast cancer diagnosis can impact quality of life for years' post-treatment. The economic implications of breast cancer treatment are substantial, encompassing both direct medical costs and indirect costs, such as lost productivity, contributing to the overall burden on healthcare systems [66], [67], [68], [69].

Brain cancer can lead to severe cognitive and physical impairments that affect daily functioning and quality of life. Patients may struggle with memory, attention, and motor skills, which can be debilitating. The care burden for patients with brain cancer often falls on family members, resulting in increased emotional and financial strain. The necessity for specialized care can also result in significant lifestyle changes for both patients and caregivers [70], [71], [72], [73], [74]

1.7.2 Challenges in Diagnosis and Treatment

Challenges in diagnosis and treatment are prevalent in both cancer types. Although screening methods like mammography have improved early detection rates for breast cancer, disparities in access to these services exist, particularly among underserved populations, leading to late-stage diagnoses that complicate treatment [75], [76], [77], [78]. Additionally, some breast cancers can develop resistance to treatments, necessitating ongoing research into more effective therapies.

In brain cancer, tumours often present vague symptoms that can delay diagnosis. Advanced imaging techniques are required for accurate identification, complicating early detection. Current treatment options for brain tumours are limited and typically involve a combination of surgery, radiation therapy, and chemotherapy, which may not always be

effective. There is an urgent need for innovative therapies and personalized medicine approaches to improve patient outcomes [79], [80], [81].

1.7.3 Screening Techniques for diagnosing breast and brain tumours

This section discusses various screening techniques employed for diagnosing breast cancer, highlighting the physical, electrical, and mechanical methods used to identify tumours at different stages. Figure 1.4 illustrates these techniques.
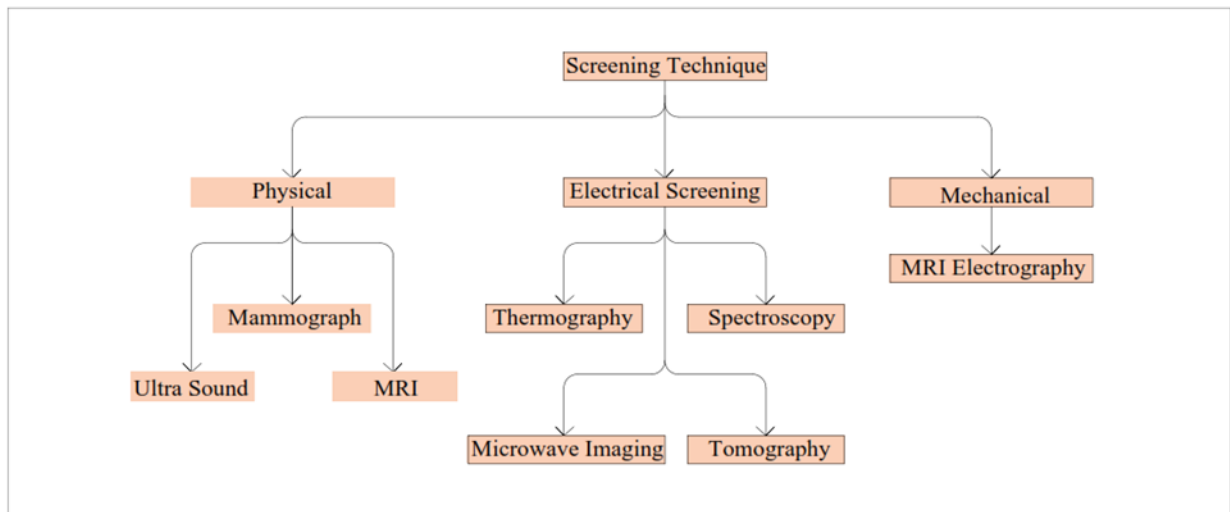
Figure 1.4 Screening Techniques

A. *Physical Screening Techniques*

➢ *Mammography*

Mammography is one of the most commonly utilized and effective methods for detecting breast cancer. This technique employs low-dose X-rays to generate images of breast tissue. Recent advancements have replaced traditional film-based X-rays with solid-state detectors, which convert X-rays into electrical signals. These signals are then processed to create digital images displayed on computer screens for further analysis. Initially, computer-aided detection (CAD) systems were introduced to support radiologists in identifying suspicious areas on mammograms. CAD systems highlight regions that warrant further

investigation, enhancing the efficiency and accuracy of breast cancer screenings, particularly in detecting subtle abnormalities that may be overlooked during manual evaluations [82].

> *Ultrasound*

Ultrasound is another widely used and effective screening technique for breast cancer, especially beneficial for women with dense breast tissue, where mammography may have reduced sensitivity. This method utilizes sound waves to produce images of the breast, offering a non-invasive and safe alternative or complement to mammography. Ultrasound is particularly adept at distinguishing between solid masses (potential tumours) and fluid-filled cysts. In populations with generally smaller breast sizes, such as among Chinese women, ultrasound has demonstrated superior screening efficiency compared to mammography. This technique is especially valuable for targeted examinations of areas flagged during physical exams or mammograms.

> *Magnetic Resonance Imaging (MRI)*

MRI is a vital tool in breast cancer detection, particularly for high-risk individuals or when more detailed imaging is necessary. This technique employs strong magnetic fields and radio waves to produce highly detailed images of the breast's internal structures. Unlike mammography or ultrasound, MRI can provide cross-sectional views, allowing for clear visualization of deep tissues. Moreover, MRI is widely utilized for monitoring treatment progress and detecting potential recurrences. Its precision makes it a recommended supplemental screening tool for women with a heightened genetic predisposition to breast cancer [83].

B. Electrical Screening Techniques

> *Electrical Impedance Spectroscopy*

Electrical impedance spectroscopy (EIS) is an emerging technique that has shown promise in enhancing breast cancer detection accuracy. EIS measures the electrical properties

of breast tissue, revealing distinct impedance patterns between cancerous and healthy tissues, thus facilitating the detection of abnormalities. These tools have proven effective in identifying variations in tissue composition, making EIS sensitive enough to detect early formations of cancer cells [84].

> ### *Thermography*

Thermography is a non-invasive, radiation-free technique that assesses temperature variations on the skin's surface. Specialized infrared cameras capture heat patterns emitted from the breast tissue. Since cancerous cells exhibit higher metabolic activity, they tend to emit more heat than surrounding healthy tissue. By employing image processing techniques and feature extraction algorithms, thermography can identify unusual heat signatures indicative of breast cancer. Although still in development, this method shows potential as a supplementary tool in breast cancer screening [85].

> ### *Microwave Imaging*

Microwave imaging is a promising new approach designed to address the limitations of conventional screening methods in early-stage breast cancer detection. This technique utilizes microwave frequencies to create images of breast tissue. Because cancerous tissues possess different dielectric properties than healthy tissues, microwave imaging may enable earlier tumour detection when other methods might fail. Although still in experimental stages, it holds considerable promise for early diagnosis.

> ### *Tomography*

Tomography refers to imaging techniques that produce cross-sectional images of specific planes within tissues, creating detailed slices of breast tissue that facilitate clearer and more precise identification of abnormalities [64]. Tomographic imaging serves as a valuable tool in both diagnosing and planning treatment for breast cancer patients.

C. Mechanical Screening Techniques

➢ Magnetic Resonance Electrography (MRE)

Magnetic resonance electrography is a mechanical screening technique that integrates MRI imaging with mechanical vibrations to assess tissue stiffness. In this method, electromechanical devices vibrate the breast tissue, generating acoustic waves. These waves are then captured by the MRI scanner, and a specialized algorithm produces quantitative images based on tissue elasticity. Since cancerous tissues are generally stiffer than normal tissues, MRE provides critical information for differentiating between benign and malignant growths. This technique is gaining attention as a potential enhancement to traditional imaging methods, offering an additional layer of diagnostic precision [85].

1.8 Motivation of study

Research in image processing and machine learning spans various fields, driven by the transformative potential of these technologies to enhance decision-making and improve overall outcomes. As society rapidly evolves, addressing critical challenges in diverse applications ranging from diagnostics and treatment planning to process optimization has become increasingly urgent. By harnessing advanced image processing techniques alongside sophisticated machine learning algorithms, significant strides can be made in enhancing outcomes across multiple domains. Image processing is foundational in industries such as healthcare, surveillance, manufacturing, and autonomous vehicles. High-resolution image generation allows professionals to make informed decisions based on detailed visual information. However, as imaging technologies advance, the complexity of generated data increases, highlighting the need for robust methodologies capable of effectively interpreting intricate datasets. This complexity necessitates innovative approaches that can adapt to varying data conditions and application requirements.

Integrating Artificial Intelligence into image analysis offers opportunities to significantly enhance accuracy and efficiency. Artificial Intelligence models, particularly deep learning techniques like Convolutional Neural Networks (CNNs), enable the extraction of patterns and features from complex datasets. This facilitates the automated identification and classification of objects in images, streamlining analysis and reducing human workload, ultimately leading to faster and more accurate results. Despite these advancements, challenges such as class imbalance and generalization remain significant hurdles. Class imbalance arises when certain categories are underrepresented in training datasets, leading to biased models. Generalization refers to a model's ability to perform well on unseen data, which is crucial for ensuring that the model remains effective across diverse scenarios. Addressing these challenges is essential for developing reliable machine learning models that maintain high accuracy across all classes and datasets.

This research aims to leverage machine learning and image processing techniques to enhance outcomes across various applications, fostering innovations that lead to better decision-making and improved processes in a rapidly evolving landscape. By focusing on these advancements, impactful solutions can be created to address real-world problems, ultimately improving efficiency, effectiveness, and overall quality of life across multiple sectors.

1.9 Research Objectives

The main objective of this thesis is to develop a model that addresses class imbalance, generalisation issues by optimizing processing speed without compromising classification accuracy. To achieve this, the proposed research work is aimed at:

1: To study and analyse the comprehensive review of existing techniques for the classification of imbalanced data.

2: To propose an approach for classification of imbalance datasets.

3: To develop a CNN, Vision transformer based hybrid framework for classification of imbalance dataset.

4: To develop a novel DL based model for generalization.

The detailed description of the identified research objectives is as follows:

1. In the first objective, different methods at both data-level, algorithmic level approaches are tested and compared. From results it can be analysed that rather than using data-level and algorithmic-level approaches individually, combining them into a hybrid approach works better for addressing class imbalance.

2. To address the second objective, a novel hybrid approach that employs Generative Adversarial Networks alongside data augmentation techniques is proposed, enhancing the quality of training dataset. Furthermore, a Vision Transformer combined with Auxiliary neural network, is employed for effective and precise tumor classification. Further, in other research a Shrinking Linear Time Vision Transformer based model is proposed which further aided for the effective tumor classification.

3. To address the third objective, an efficient Deep Learning framework that deals with data imbalance is proposed by combining Inception ResNet V2 and Vision Transformed, the model achieves high accuracy and computational efficiency across medical imaging datasets. This hybrid design improved feature extraction, delivering precise diagnosis and robust performance, even with imbalanced datasets.

4. In the fourth objective, model is proposed for improved generalization and ability for addressing image occlusions. Gaussian filters are applied for noise reduction, along with Patterned Grid Mask for storing critical information at each stage of the model aiding for improved generalisation of model on various imbalanced datasets.

1.10 Thesis Outline

The thesis is organised in eight chapters; salient features of each chapter are discussed in the following:

FIRST CHAPTER introduces the concepts of Machine Learning, Artificial Intelligence, Class Imbalance, and Generalization in image processing.

SECOND CHAPTER reviews the literature in the field of image processing with special reference to handling class imbalance, generalisation, optimization issues. An in detail review is conducted on understanding class imbalances, approaches to address it. Various data-level modifications, algorithmic level modifications and hybrid approaches proposed in literature to mitigate class imbalance are understood. Similarly, review is conducted on how various model proposed in literature addressed generalisation issue. It further emphasizes the need to carry out research study in this area for addressing class imbalance and generalisation issues and their significance in image processing. Finally, the inferences from the available literature have been drawn in order to identify and categorize the requirements for the proposed work.

THIRD CHAPTER offers a comprehensive understanding of how class imbalance was addressed in image processing, based on an analysis of both data-level and algorithmic-level methods found in existing literature. The framework organizes these techniques hierarchically, clarifying the foundational principles and effectiveness of each approach. By categorizing solutions, this research facilitates structured analysis and comparison, enabling the selection of optimal strategies for mitigating imbalance. Additionally, experiments were conducted to evaluate the performance of hybrid models compared to standalone methods, demonstrating a more effective approach for addressing class imbalance in image classification tasks.

FOURTH CHAPTER as it was understood that hybrid techniques are efficient in addressing class imbalance, in this chapter a hybrid technique is proposed combining data-level modifications using Generative Adversarial Networks (GANs) with algorithm-level enhancements through feature extraction via Vision Transformers (ViT) and an auxiliary artificial neural network (ANN) for classification. This chapter explains how the additional images generated by GANs contribute to improving the model's efficiency and accuracy. By leveraging both data augmentation and advanced feature extraction techniques, our approach aims to significantly enhance classification performance. The effectiveness of this hybrid model is evaluated across various datasets, highlighting its advantages in effectively tackling class imbalance.

FIFTH CHAPTER presents a robust computational efficient approach for enhancing automated classification. The proposed method integrates Generative Adversarial Networks (GANs) with data augmentation strategies and a structural similarity loss function to effectively generate annotated images. It breaks new ground by introducing a novel deep learning model inspired by the Vision Transformer, known as the Shrinking Linear Time Vision Transformer (SL(t)-ViT), specifically designed for disease classification. This model thoroughly evaluates multiple datasets to assess its performance in accurately identifying tumours. Key evaluation metrics highlight its superior efficacy compared to existing techniques, underscoring the potential of SL(t)-ViT and GANs in improving diagnostic accuracy and efficiency in medical imaging.

SIXTH CHAPTER introduces a computationally efficient deep learning framework to address dataset imbalance through a hybrid model, ensuring accuracy and speed in image classification. The novel design excels in generalization, by integrating features from Inception ResNet V2 and Vision Transformers (ViT) to enhance classification performance. A key innovation is the Efficient Attention Mechanism, which optimizes focus on critical features for

improved accuracy and efficiency. Additionally, a Resource-Efficient Optimization model through feature selection streamlines computational usage without sacrificing accuracy. By addressing heterogeneity within classifiers, our framework integrates high-dimensional features, resulting in more accurate class predictions and advancing precision in classification.

SEVENTH CHAPTER proposes an enhanced deep learning algorithm for improved generalisation. In this chapter, Gaussian filters are utilized for noise reduction, and Patterned-Grid Mask, an advanced variant of the standard Grid Mask technique, was introduced ensuring that critical information is preserved. Furthermore, modifications to the Multi-Axis Vision Transformer architecture and integration of GRN-based MLPs, improved the computational efficiency and generalization capacity. The proposed model demonstrated outstanding performance across various tasks, achieving significant accuracy and robustness in both binary and multi-class classifications. This study highlights the enhanced DL model's potential to improve accuracy, efficiency, and generalization in brain tumour diagnosis, making it a viable solution for real-world applications.

EIGHTH CHAPTER explains the major results obtained from the foregoing chapters and further explore the possibilities of future research work that can be carried out to extend the present work. The references have been provided at the end of the thesis.

Chapter Two: LITERATURE REVIEW

This chapter provides the overview of research area, review progression and literature survey of various techniques proposed in literature for addressing class imbalance and generalization issue.

2.1 Overview

The classification of images from class-imbalanced datasets is a pivotal focus in image processing research, leading to new methods that improve the accuracy of automated systems. Machine learning algorithms now efficiently detect objects and patterns in imbalanced datasets, which is crucial as class imbalance can lead to misclassification of rare features. Modern classification systems leverage data from diverse sources, such as satellite images and sensory inputs, to create robust detection models. Traditional methods struggle with skewed datasets, often favouring majority classes over minority ones. New approaches, like data augmentation, transfer learning, and synthetic data generation, address this by enhancing models' ability to learn from limited examples and handle rare classes more accurately.

Advanced machine learning models now incorporate ensemble techniques and cost-sensitive learning, adjusting for class imbalances to boost detection accuracy. These advancements support diverse applications, including surveillance, environmental monitoring, and quality control. Continued research into algorithm development is key to improving classification models and standardizing approaches for handling imbalanced datasets in practical applications. This progress drives future innovations and broadens the impact of automated systems across multiple fields.

2.2 Review progression

A systematic approach as shown in Figure 2.1 was used to identify machine learning techniques for classifying images from class-imbalanced datasets. Initially, 1,768 relevant

research articles were retrieved from scientific digital libraries and reviewed thoroughly. This study focuses on machine learning methods for improving classification in cases with underrepresented classes, specifically in medical imaging from 2010 to 2023. Using keywords like "image processing," "classification," "class imbalance," "machine learning," and "diagnosis," along with Boolean operators, we refined our search in databases such as:

IEEE Explore,

PubMed,

Web of Science, and

ACM Digital Library.

The articles whose abstracts or titles indicated a focus on machine learning in medical image classification or addressed class imbalance were selected; otherwise, they were excluded. Ultimately, 219 papers were reviewed, with 74 specifically addressing class imbalance in medical imaging. These studies were meticulously analysed for strengths, weaknesses, and future implications.
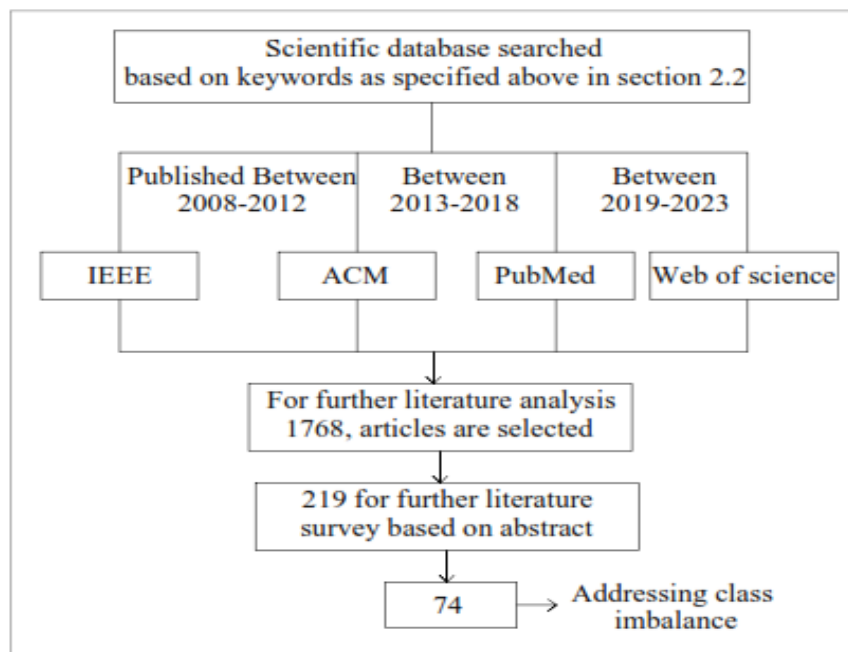
Figure 2.1 Search Methodology for the Selection of Relevant Articles

2.3 Literature Review: Machine Learning Methods used to mitigate class imbalance.

Recent advancements in machine learning (ML) have outpaced human performance in tasks like object recognition [86], [87], [88], [89], with Convolutional Neural Networks (CNNs) playing a key role. Unlike traditional methods reliant on handcrafted features, CNNs automatically learn complex data patterns, greatly advancing medical image analysis and reaching expert-level accuracy in several imaging tasks [90], [91], [92]. However, challenges remain, particularly in lesion detection, which requires precise localization within images which is a labour-intensive task for clinicians. A primary challenge in lesion detection is class imbalance, as lesions occupy minimal image space, creating a disparity between lesion and background pixels. This imbalance significantly affects CNN performance, especially for very small lesions like micro aneurysms in retinal images or micro calcifications in mammograms, where ratios can reach 1: 10,000 [93].

Approaches to manage class imbalance in deep learning include (i) data-level methods, (ii) algorithmic-level methods, and (iii) hybrid approaches. Data-level methods involve balancing the training set using feature selection, extraction, and data augmentation, though alone they often fall short for severe imbalances [94], [95], [96], [97], [98], [99]. Algorithmic-level methods adjust training processes, such as under sampling, oversampling, cost-sensitive learning, and novelty detection, to counter imbalance effects [100], [101], [102], [103]. Hybrid approaches combine these methods for improved performance. In medical imaging, class imbalance solutions are limited. Simple augmentation (e.g., flipping, rotation) is commonly used, while more advanced techniques like GAN-based augmentation and custom loss functions, including Dice loss, show promise for minor class emphasis [104], [105], [106], [107]. Fully Convolutional Networks (FCNs) are effective for pixel-wise classification but face limitations with datasets lacking precise pixel-level annotations, as seen in cases like the IN breast dataset [108].

This research further explores various techniques addressing class imbalance in detail in subsequent sections.

2.3.1 Data-level Approaches to Address Class Imbalance

In the literature survey for handling class imbalance in medical image datasets, data-level approaches play a pivotal role in modifying the dataset before feeding it into the machine learning model. These approaches align with pre-processing techniques, which aim to mitigate the imbalance ratio between the majority and minority classes. The key data-level techniques employed to achieve this are feature selection, feature extraction, and various sampling methods. Below is a detailed explanation of these techniques under the data-level approaches category:

A. Feature Selection

Feature selection plays a pivotal role in enhancing the performance of machine learning models, particularly in medical imaging tasks such as cancer diagnosis, where datasets are often highly imbalanced. By identifying the most relevant features such as tumor size, shape, texture, or edges models can focus on critical aspects that improve classification accuracy, especially for underrepresented samples, like tumors or lesions. This is crucial in applications where prioritizing specific image features helps distinguish between malignant and non-malignant cases.

(i) Principal Component Analysis (PCA): This technique reduces high-dimensional data into a smaller set of principal components, capturing the most significant variance. This approach helps machine learning models by retaining essential information while simplifying datasets, making it especially valuable for large and complex datasets.

PCA has been widely applied to address class imbalance in classification. Zhang et al. [109] introduced a one-class kernel subspace ensemble model that uses PCA in a kernel framework to emphasize features of minority classes, enhancing performance in imbalanced datasets.

Huang et al. [110] combined PCA with DenseNet in a hybrid neural network to improve classification accuracy by retaining critical information from minority classes in a reduced-dimension format.

In general applications, Kumar and Vijayakumar et al. [111] demonstrated that combining PCA with an RBF kernel-based SVM boosts classification accuracy and efficiency. Nandi et al. (2015) [112] applied PCA for feature extraction, showing that it effectively reduces noise, enhancing classification outcomes. Additionally, Kumar, Dabas, and Godara et al. [113] used a hybrid method combining PCA, SVM, and Discrete Wavelet Transform, improving classifier performance. Mohan and Subashini et al. [114] employed PCA in a multiclass classification setting, demonstrating notable accuracy improvements.

(ii) Genetic Algorithms (GAs): It is inspired by natural selection, optimize solutions through selection, crossover, and mutation. In feature selection, GAs help identify optimal subsets that enhance model accuracy, with each individual in the population representing a potential subset evaluated for fitness.

Several studies employ GAs for classification and optimization. Haque et al. [115] combined GAs with ensemble learning to improve accuracy for imbalanced datasets by optimizing classifier combinations to better handle minority classes. Jiang et al. [116] introduced GASMOTE, a hybrid GA-SMOTE method creating synthetic data to balance training datasets. Azad et al. [117] proposed combining GAs with decision trees to address class imbalance and missing data, achieving notable performance gains.

Beyond medical applications, GAs have been used for feature selection and hyper parameter tuning, such as in CNNs for skin cancer detection, enhancing melanoma classification accuracy Azad et al., [117]. Devarriya et al. [118] developed GA-based fitness functions for minority class detection.

***(iii) Recursive Feature Elimination (RFE)*:** This method systematically improves model performance by iteratively removing the least important features. It starts with a full feature set, trains a model (often SVM or Random Forest), ranks feature importance, and removes the least significant feature, repeating until the optimal subset is achieved. Numerous studies apply Random Forest (RF) and ensemble methods to address class imbalance. Xu et al. [119] combined M-SMOTE, ENN, and RF to generate synthetic minority data, filter noise, and classify imbalanced datasets, achieving improved performance. Desir et al. [120] used RF for one-class classification, enhancing minority class learning without overfitting. Bader-El-Den et al. [121] introduced Biased Random Forest (BRAF), which increases focus on minority classes, leading to better outcomes in imbalanced datasets.

RFE's effectiveness extends beyond medical imaging. For example, it was applied with TabNet and XGBoost in breast cancer diagnostics [109] and with Conv-LSTM for brain tumor classification in 3D MRI [110]. RFE has also optimized MRI features for prostate cancer recurrence prediction and ultrasound features for lymph node status in ovarian cancer [111], [112]. In osteosarcoma analysis, RFE combined with ensemble classifiers helped predict tumor viability [113]. Additionally, Müller et al. [122] developed a classification pipeline using CNNs and RF, demonstrating how RF-based ensemble learning enhances CNN performance on imbalanced datasets.

B. Feature Extraction

Feature extraction is essential for improving the performance of diagnostic models across various fields. By simplifying complex images into meaningful attributes, professionals can gain deeper insights into underlying conditions, enhancing accuracy and enabling timely responses. This process transforms raw data into attributes that represent intricate patterns and structures. In image processing, feature extraction helps identify significant characteristics in complex visuals, which are crucial for analysis and decision-making. By isolating key features,

experts can assess conditions more effectively and make informed choices. Overall, feature extraction is a foundational element in interpreting and analysing images, leading to improved outcomes in various applications.

(i) The Histogram of Oriented Gradients (HOG): This technique is a popular feature extraction technique in computer vision, especially for object detection and classification. HOG captures the distribution of gradient orientations within localized regions of an image by dividing it into small cells (typically 8x8 or 16x16 pixels) and calculating the gradient magnitude and direction for each pixel. These gradients are aggregated into histograms representing edge orientations within each cell, with normalization across overlapping blocks to enhance robustness against changes in illumination and contrast. HOG is effective for detecting shapes and edges, making it valuable in applications requiring precise edge detection. It emphasizes contours, improving accuracy in object localization. HOG features can be combined with machine learning classifiers, such as support vector machines (SVM) or random forests, to classify images based on extracted edge and shape information.

Several studies highlight HOG's versatility. One study combined HOG with Local Ternary Patterns (LTP) to extract features from imbalanced datasets, enhancing image retrieval by focusing on edge and texture information, thereby reducing the impact of class imbalance Shamna & Musthafa et al. [123]. Another approach integrated HOG descriptors into a multi-scale CNN framework, demonstrating its ability to balance performance across minority and majority classes Liu et al., [124]. A comparative study also showcased HOG's effectiveness in detecting small objects by preserving key edge features critical for identification, especially when paired with deep learning models Bria et al., [103]. HOG has been applied in various contexts beyond object detection. One study utilized HOG features for identifying objects in images, emphasizing edge detection and gradient information [109]. Another combined local binary patterns with convolutional neural networks (CNNs), using HOG for pre-processing to

improve feature extraction and object detection from high-resolution images [110]. HOG has also been pivotal in developing computer-aided diagnosis systems, providing robust feature extraction for accurate image classification [111]. Additionally, it has been used in image analysis tasks, such as detecting abnormalities in medical imaging and studying movement patterns [113]. These applications underscore HOG's wide applicability and potential to enhance classification accuracy across diverse image analysis tasks.

(ii) The Scale-Invariant Feature Transform (SIFT): This technique is a powerful feature extraction technique for detecting and describing local features in images, maintaining invariance to changes in scale, rotation, and affine transformations. It identifies key points using a difference-of-Gaussian approach, enabling detection at multiple scales, and computes descriptors from local gradient information around each key point, capturing robust patterns. SIFT's invariance to scale and rotation makes it valuable for consistent feature matching across various images, particularly in image classification and object detection tasks, even under differing conditions or angles. SIFT has been applied to address class imbalance in image classification tasks. One study utilized SIFT descriptors to capture key points and integrate them into a machine learning pipeline, which improved classification accuracy for varying scales and orientations, reducing the impact of class imbalance Liu et al., [125]. Another study leveraged SIFT features in a classification framework, where SIFT's robustness ensured that minority class features were well represented, enhancing performance on imbalanced datasets Tang & Hu et al. [126].

In a further study, SIFT was used for local feature extraction in deep learning models, demonstrating its effectiveness in detecting minority class instances in skewed datasets. The combination of SIFT with deep learning helped balance class representation and improve classification outcomes Yadav & Jadhav et al. [127]. SIFT has also been widely applied in object classification and pattern recognition. One study combined SIFT with a hybrid deep

39

learning model based on Efficient Net, showcasing its ability to extract features that enhance predictions by identifying critical local features [109]. Another study used SIFT alongside Fed-VGG16 and Fed-CNN models, highlighting its effectiveness in improving model performance through significant feature extraction [110]. These studies underscore SIFT's versatility and effectiveness in various image classification applications, particularly in challenging conditions and imbalanced datasets.

(iii) Local Binary Patterns (LBP): This method is a powerful texture descriptor for analysing local texture information in images. It generates a binary code by thresholding pixel values in a local neighbourhood around each pixel, indicating whether neighbouring pixel intensities exceed the centre pixel's intensity. This code is converted into a decimal value, creating a unique LBP representation for each pixel, which captures local patterns for classification tasks. LBP is robust to changes in illumination, making it useful when image brightness varies. The extracted features can be aggregated into histograms for machine learning classifiers, enhancing the accuracy of classification models.

LBP has been widely applied to address class imbalance in classification tasks, particularly in texture analysis. One study utilized adaptive LBP to ensure underrepresented classes were well represented, improving classification accuracy by focusing on critical texture details Liu et al., [125]. Another study integrated LBP with other feature extraction methods, using multi-scale techniques to balance class representation, leading to more accurate classification results Murugappan & Sabeenian [128].

(iv) Gabor filters: These linear filters are widely used for texture and edge feature extraction at various scales and orientations. Based on the mathematical properties of wavelets, these filters effectively capture the spatial frequency content of images. Each Gabor filter responds to specific frequencies and orientations, allowing for the extraction of a rich set of texture features that emphasize different aspects of an image. Gabor filters are particularly effective in

addressing class imbalance by capturing multi-scale and multi-orientation features. One study demonstrated their ability to extract distinct texture features, ensuring better representation of minority classes in the classification process, which enhanced classifier robustness Poloni et al., [129]. Another approach integrated Gabor filters into a data augmentation technique, generating additional samples for underrepresented classes and improving classification performance in cases of severe class imbalance Barshooi & Amirkhani [130].

These filters have been extensively applied in texture and pattern recognition. Singh et al. [131] utilized a Gabor filter bank in a texture-based classification system, proposing a computer-aided diagnostic network that improved classification accuracy compared to traditional methods. Bourkache et al. [132] demonstrated the adaptability of Gabor filters in large-scale data contexts, enhancing classification tasks with extensive datasets. Li et al. [133] explored Gabor filters for feature enhancement in deep learning models, highlighting their integration in segmentation and classification tasks to improve performance in complex image analysis. These studies emphasize the versatility and effectiveness of Gabor filters in extracting meaningful texture and edge information across diverse applications.

(v) The Grey Level Co-occurrence Matrix (GLCM): This statistical technique is used to analyse the spatial relationships between pixels in an image. It measures the frequency of pixel pairs with specific grey-level values at defined distances and directions. From the GLCM, several texture features can be derived, including contrast, correlation, energy, and homogeneity, providing a quantitative analysis of an image's texture and its structural properties. GLCM is widely applied in texture analysis and classification, distinguishing textures and improving classification model performance. It has also been used to address class imbalance in classification tasks.

GLCM has been utilized across various fields; for example, Azawi et al. [134] combined GLCM with a genetic algorithm and a probabilistic neural network to improve

classification accuracy through optimal feature selection. Hao et al. [135] enhanced classification performance by combining deep semantic features with GLCM-derived texture features, revealing insights into image heterogeneity. Additionally, Biswas et al. [136] demonstrated GLCM's effectiveness in capturing spatial relationships, leading to improved diagnostic outcomes based on texture patterns.

(vi) Wavelet Transform: This is a powerful technique for decomposing images into different frequency components, allowing simultaneous analysis of spatial and frequency information. Unlike the traditional Fourier Transform, which offers a global frequency representation, the wavelet transform provides localized frequency analysis, making it particularly effective for images with non-stationary characteristics. This method generates a multi-resolution representation, capturing details at various scales and orientations, which helps identify localized features such as edges and textures while preserving overall image structure.

Widely adopted in various fields, wavelet transform enhances feature extraction for texture and localized features, improving classification performance. It effectively addresses class imbalance by capturing both spatial and frequency information. For instance, Krishna et al. [137] applied it in a classification system combined with a PSO-based LLRBFNN algorithm to enhance classification accuracy by extracting texture and spatial frequency information. Sarhan et al., [138] introduced a wavelet-based approach with deep learning, using wavelets to extract energy features and improve accuracy by leveraging their multi-resolution properties. These studies demonstrate the versatility and effectiveness of wavelet transform in image analysis and classification tasks, particularly for multi-scale texture analysis and handling imbalanced datasets.

C. Sampling Techniques

Sampling techniques are commonly used to address class imbalance in machine learning by adjusting class distribution before model training. These methods are classified into

42

three main types: over-sampling, under-sampling, and hybrid techniques. Over-sampling increases the representation of the minority class, while under-sampling reduces the majority class samples. Hybrid techniques combine both approaches to create a more balanced dataset.

The primary goal is to improve model performance by mitigating the negative effects of imbalanced class distributions, enhancing generalization and accuracy in classification tasks. When applied correctly, these techniques can significantly improve classification rates, especially for critical conditions or rare events. However, each method presents challenges; for example, over-sampling may cause overfitting, while under-sampling can result in the loss of valuable information. Therefore, careful consideration of the dataset's characteristics and problem context is essential to balance class parity and data integrity.

(i) Random Oversampling (ROS): It is a simple yet effective technique for balancing class distribution by duplicating samples from the minority class. This method increases the representation of the minority class, enabling models to learn from a more representative dataset, especially when instances are scarce. However, ROS has notable drawbacks, primarily the risk of overfitting, as repeated samples can lead to redundant patterns that may not generalize well to new data. This issue is pronounced in high-dimensional datasets, prompting practitioners to combine ROS with advanced techniques like SMOTE to mitigate redundancy.

Random oversampling has been widely applied across various fields. Khushi et al. [139] analysed resampling methods, finding that ROS significantly improved classification performance on imbalanced datasets but warned about overfitting risks. Ganganwar et al. [140] used ROS to artificially increase minority class instances, noting that while it enhanced model performance, the model might memorize duplicated instances rather than generalize. To counter this, the authors recommended combining oversampling with data augmentation techniques. Douzas et al. [141] found that while ROS improves performance, its effectiveness can be limited by the nature of duplicated data. They suggested combining oversampling with

43

heuristic methods, like k-means clustering, for better distribution of synthetic samples. Jeyaraj and Nadar et al. [142] applied ROS in their deep learning study on early oral cancer diagnosis, which improved classification accuracy. However, they also highlighted the potential for bias due to overfitting to replicated samples, affecting generalization to new data.

(ii) Random Under sampling (RUS): This technique addresses class imbalance by randomly eliminating samples from the majority class. This simplifies the dataset and enables models to focus on the minority class, making it effective when the majority class is significantly larger. However, RUS can lead to the loss of critical information, negatively impacting model generalization. To mitigate this, researchers often combine RUS with techniques like ensemble learning.

RUS has been widely applied in various studies. Mohammed et al. [143] noted that while RUS simplified their dataset, it risked losing important majority class information, potentially harming performance. Lin et al. [144] introduced a clustering-based approach to preserve representative majority class samples, resulting in improved classification.

(iii) Synthetic Minority Over-Sampling Technique (SMOTE): This technique generates synthetic samples for the minority class through interpolation, enhancing dataset diversity without duplicating existing samples, unlike random oversampling. This method helps improve classification performance, especially when minority class instances are limited, and reduces the risk of overfitting by facilitating better generalization to unseen data.

SMOTE has been widely utilized in various studies. Joloudari et al. [145] found that combining SMOTE with convolutional neural networks (CNNs) significantly enhanced classification accuracy on imbalanced datasets. Blagus and Lusa et al. [146] explored SMOTE in high-dimensional datasets, developing a variation to preserve class relationships and improve performance. Gao et al. [147] demonstrated a deep learning approach using SMOTE for better classification, while Wang et al. [148] applied SMOTE in breast cancer classification,

achieving improved accuracy and F1-scores. Both studies noted the risk of noise from synthetic samples that may not accurately represent real-world data.

(iv) Adaptive Synthetic Sampling (ADASYN): The following technique builds upon the SMOTE technique by selectively generating synthetic samples in regions where the minority class is difficult to classify. This method dynamically adjusts the number of synthetic samples based on classification complexity, thereby enhancing the model's ability to distinguish between classes in highly imbalanced datasets.

ADASYN focuses on generating synthetic samples near the decision boundary. For instance, Douzas et al. [141] demonstrated that this targeted approach improved classification performance by addressing challenging cases and reducing bias toward the majority class. Additionally, a comparative study by Elreedy and Atiya et al. [149] found that ADASYN outperformed traditional methods such as random oversampling and SMOTE by adapting to the density of the minority class within the feature space. The primary advantage of ADASYN lies in its ability to effectively learn from difficult instances without overwhelming the model with unnecessary data. However, it requires careful tuning to ensure that the synthetic samples accurately represent the minority class. Guan and Liu et al. [150] showed that ADASYN enhanced model accuracy by generating realistic synthetic samples in challenging regions. Nonetheless, they also cautioned about potential computational complexity and the risk of overfitting due to noisy synthetic samples.

D. Data Augmentation

Data augmentation is a widely used technique in image processing for addressing class imbalance in datasets. It involves artificially expanding the minority class by applying transformations such as rotation, scaling, flipping, and contrast adjustment to existing images. These transformations create new examples, enriching the dataset and enabling models to learn from a broader range of representations. The primary advantage of data augmentation lies in

45

its ability to increase the diversity of the training data, significantly reducing the risk of overfitting which is a common issue when models are trained on limited examples. By providing a more comprehensive dataset, data augmentation enhances a model's capacity to generalize to new, unseen cases. However, it's crucial to apply transformations carefully to ensure that the augmented data remains representative of the actual conditions being modelled, preserving the integrity of the data.

(i) Generative Adversarial Networks (GANs): These represent a sophisticated method for generating synthetic samples, particularly effective in addressing class imbalance. Comprising a generator and a discriminator, GANs create synthetic images resembling the minority class while evaluating their authenticity. This adversarial process produces highly realistic synthetic data, enhancing training datasets and improving model performance on imbalanced data.

GANs have shown significant potential in various domains. For example, they have been used to generate synthetic text samples for minority classes in text classification tasks, boosting performance in sentiment analysis and spam detection Wang et al., [151]. In fraud detection, GANs improved detection rates by generating synthetic fraudulent transactions, which typically make up a small fraction of datasets Luo et al., [152] . Additionally, in autonomous driving, GANs have created synthetic examples of rare events like accidents, helping models generalize to various driving scenarios Goodfellow et al., [153]. In industrial defect detection, GANs generated synthetic images of defects, enhancing quality control by improving the identification of subtle issues Li et al., [154].

In medical applications, GANs have effectively addressed class imbalance. Frid-Adar et al. [155] applied GAN-based image augmentation to improve convolutional neural networks (CNNs) in liver lesion classification, achieving higher accuracy, sensitivity, and specificity. Mahmood et al. [156] implemented adversarial training for brain tumour segmentation,

generating synthetic tumour images that mimicked the statistical properties of real data, significantly enhancing segmentation performance.

2.3.2 Drawbacks and Open Issues in Data-Level Approaches

Data-level approaches, such as oversampling, under sampling, and synthetic data generation, are commonly used to address class imbalance in datasets. However, these methods have significant drawbacks that limit their effectiveness, prompting researchers to explore alternative algorithmic strategies. A major issue with oversampling is the increased risk of overfitting. By duplicating or generating additional minority class instances, oversampling may lead the model to memorize these samples instead of learning generalized patterns, resulting in poor performance on unseen data. This is particularly problematic when there are few minority class samples, as the model may rely too heavily on these limited examples.

Under sampling reduces the number of majority class instances but can lead to the loss of valuable information, causing the model to miss important patterns. This may result in under fitting, where the model fails to learn the overall data structure, thereby reducing accuracy and predictive power. While synthetic data generation techniques aim to mitigate overfitting and under fitting, they often struggle to maintain sample diversity. Generated data may not capture the complexity of the minority class, introducing noise or unrealistic points that misrepresent true characteristics. Additionally, data-level approaches are static, simply adjusting the dataset's distribution without considering variations in class imbalance across different feature space regions. This limitation highlights the need for more adaptive solutions, leading to a shift towards algorithmic approaches that modify the training process itself for dynamic handling of imbalanced data.

2.3.3 Algorithm-Level Approaches to Address Class Imbalance

To address the limitations of data-level approaches, researchers have increasingly focused on algorithmic modifications to manage class imbalance in datasets. These strategies involve adapting existing algorithms to better tackle the challenges of imbalanced data. Common techniques include adjusting loss functions to give greater weight to the minority class and implementing cost-sensitive learning that penalizes misclassification of minority instances more heavily. Additionally, ensemble methods like boosting and bagging enhance classification performance by improving the recognition of underrepresented classes.

These algorithmic refinements aim to increase model robustness and performance, particularly in critical applications such as fraud detection and autonomous driving. By modifying algorithms, researchers can mitigate the negative impacts of class imbalance, ensuring equitable representation of all classes and improving predictive accuracy and reliability. This is especially vital in domains like industrial defect detection and text classification, where identifying rare events is crucial for achieving high-quality results.

A. Cost-Sensitive Learning

Cost-sensitive learning is a powerful technique for addressing class imbalance by assigning different penalties for misclassifications based on class importance. Typically, higher misclassification costs are assigned to minority class instances, encouraging the algorithm to prioritize these underrepresented samples during training. A cost matrix is used to integrate these penalties directly into the learning process by adjusting the loss function. This approach improves sensitivity and recall for the minority class, which is crucial in applications such as fraud detection and autonomous driving, where identifying rare but important instances is vital for system reliability.

(i) Focal loss: It is a novel loss function designed to tackle class imbalance, particularly when the minority class is underrepresented. Unlike traditional loss functions that treat all misclassifications equally, focal loss introduces a modulation factor that down-weights easy-to-classify examples while emphasizing hard-to-classify minority instances. This adjustment enhances the model's sensitivity to challenging samples, reducing bias toward the majority class and improving performance on imbalanced datasets.

Initially developed for extreme class imbalance in object detection, focal loss significantly improved detection rates for minority classes by focusing on hard examples, such as identifying pedestrians in autonomous driving scenarios Lin et al., [157]. In wildlife monitoring, focal loss prioritized learning for rare species, enhancing the classification of endangered animals in camera-trap images and reducing false positives [158].

Lamrani et al. [159] applied focal loss in a deep residual network for tumor classification in MRI images, which reduced misclassification rates for minority tumor types but struggled with extremely rare cases due to limited data. Liu et al., [160] used focal loss in 3D tumor image segmentation, enhancing detection of small regions while noting a risk of overfitting in small datasets. Luong et al. [161] developed SovaSeg-Net for ovarian tumor segmentation, with focal loss outperforming other functions, though hyper parameter tuning remains a challenge. Agrawal et al. [162] introduced a lightweight skin cancer detection system using adaptive class-balanced focal loss, improving classification of rare skin cancer types, but highlighted the need for automation in parameter selection.

(ii) Modified cross-entropy: This loss function is an adaptation of the standard cross-entropy function aimed at enhancing model performance on imbalanced datasets. Unlike traditional cross-entropy, which treats all classes uniformly, the modified version assigns higher penalties for misclassifying minority class instances, ensuring that the model pays adequate attention to

these underrepresented examples while minimizing overfitting to the majority class. Studies, such as those by Ismail et al. [163], have demonstrated its effectiveness in improving class balance without significantly increasing computational complexity. Despite its limitations, cross-entropy loss remains a common baseline in classification tasks. Luong et al. [161] applied it to scene recognition, revealing that the model struggled with rare environments like natural forests.

Rai et al. [164] utilized binary cross-entropy in a two-headed U-NetEfficientNet architecture for brain tumor segmentation, noting challenges with smaller tumours due to class imbalance. Yang et al. [165] applied cross-entropy in a data-efficient image transformer for multiclass brain tumour classification but highlighted the need for further modifications to address tumour heterogeneity. Modified cross-entropy variants have been proposed to tackle these issues. For example, Cui et al. [166] used weighted cross-entropy in a vehicle detection dataset, improving detection accuracy for underrepresented vehicle types. Prayogo et al. [167] employed modified cross-entropy in a transfer learning model for brain tumour detection, but challenges in generalizing across different tumour types and imaging conditions persisted.

(iii) Class-balanced loss: It is an advanced loss function that addresses class imbalance by assigning different weights to classes based on their representation in the dataset. This method enhances the contribution of minority classes to the loss calculation while reducing the influence of majority classes, allowing the model to learn more comprehensive features from underrepresented categories.

Class imbalance is a common challenge in classification tasks, leading to biased models. Researchers have increasingly turned to class-balanced loss as an effective solution. Numerous studies from 2015 to 2022 demonstrate its utility across various fields. For instance, Cui et al. [166] applied class-balanced loss to an imbalanced vehicle detection dataset,

improving recognition of rare vehicle types like motorcycles and achieving more balanced detection accuracy.

B. Ensemble Learning with Class Rebalancing

Ensemble learning with class rebalancing leverages multiple classifiers to improve performance on imbalanced datasets. Techniques like boosting, bagging, and stacking are adapted to prioritize minority classes during training. Boosting methods, such as AdaBoost, increase the weights of misclassified minority samples, helping the model focus on these challenging cases and reducing bias towards the majority class. This approach can significantly enhance performance in various domains, such as fraud detection or autonomous driving, where identifying rare events is crucial for system reliability. A notable example is the EUS-Boost algorithm, which combines boosting with under sampling to address class imbalance. This method outperformed standard sampling techniques in tasks like rare event detection, making it a valuable strategy in applications where accurately identifying minority classes is essential for success.

(i) Boosting techniques: AdaBoost and Gradient Boosting, have been widely employed to address imbalanced datasets in classification tasks. These methods sequentially train weak classifiers, focusing on the errors of previous classifiers. Research indicates that AdaBoost enhances accuracy in detecting underrepresented classes, improving the identification of rare cases.

Boosting algorithms have been extensively studied for managing class imbalance by adjusting the model's focus on difficult instances. For example, Sun et al. [168] applied AdaBoost to fraud detection in financial datasets with highly underrepresented fraudulent cases, enhancing recall without sacrificing overall accuracy. Chawla et al. [169] introduced SMOTE+ Boost, combining boosting with synthetic data generation, significantly improving

the detection of rare manufacturing defects in quality control. By assigning higher weights to images showing early signs of retinopathy, GBMs effectively handled class imbalance, improving diagnostic performance and reducing misclassification risks.

(ii) Bagging: A widely used bagging technique, Random Forest, constructs an ensemble of decision trees, each trained on a random sample. This approach reduces variance and improves classification stability, effectively addressing class imbalance and enhancing minority class performance while maintaining high accuracy for the majority class.

Bagging has proven effective across various tasks. For example, Iqbal et al. [170] demonstrated that Random Forest outperformed traditional classifiers in brain tumor detection, showing greater accuracy and robustness, particularly with noisy or incomplete data. Lessmann et al. [171] applied bagging to classify liver lesions, allowing the model to generalize better by mitigating the effects of outliers. Modifications of bagging also target class imbalance. Galar et al. [172] employed under-bagging in cybersecurity anomaly detection, where the majority class was under sampled, improving detection of rare anomalies. Chen et al. [173] utilized over-bagging for customer churn prediction, enhancing sensitivity to rare churn cases.

(iii) Stacking is an ensemble technique that enhances classifier performance by training a meta-classifier to make predictions based on outputs from various base classifiers. Models like CNN, SVM, and KNN can serve as base learners, each capturing different data aspects. This method effectively integrates diverse learning strategies, making it suitable for addressing class imbalance across various domains.

Stacking has been applied to class imbalance in several fields. For instance, Zhou et al. [174] used stacking in e-commerce fraud detection by combining decision trees, logistic regression, and SVMs, which improved the detection of rare fraudulent transactions. In credit scoring, Lessmann et al. [171] demonstrated that stacking different classifiers on imbalanced

data improved the identification of high-risk borrowers while maintaining a low false-positive rate. Galar et al. [172] applied stacking in industrial fault detection, achieving better accuracy in detecting faults compared to individual models.

(iv)Transfer learning: By utilizing pre-trained models, typically trained on large datasets, this method allows for the transfer of learned features to new, imbalanced tasks, reducing the time and computational resources needed for model development. During fine-tuning, techniques like cost-sensitive learning or focal loss can enhance performance on minority classes. For instance, the Inception V3 model has been adapted for cancer detection by integrating dynamic sampling and cost-sensitive methods to improve accuracy for rare conditions.

The authors used ViT as a pretraining model to extract general characteristics from massive amounts of heterogeneous data, followed by two-stage transfer learning to learn underlying target dataset information for COVID-19 picture identification. He et al. [175] discussed the use of transformers in medical picture processing and surveyed available transfer learning designs for clinical image processing, highlighting their shortcomings.

(v) Attention mechanism are powerful tools for addressing challenges in imbalanced datasets by selectively focusing on relevant areas of an image. They enhance feature representation, improving detection of minority classes. For instance, in breast cancer classification, attention mechanisms direct focus toward denser tissues, increasing accuracy for rare instances like malignant tumours. In remote sensing, attention mechanisms have proven valuable.

Chen et al. [176] proposed a new technique for segmenting brain tumours called LETCP, which uses unlabelled data for pre-training and a self-attention transformer for model construction. The authors tested the approach on three different public datasets and showed improved effectiveness over existing techniques. Lee et al. [177] introduced a strategy for

image classification that uses the extracted features of all patches processed by a sequence of self-attention layers rather than the class token employed by conventional vision transformers.

Luque et al. [178] proposed a novel approach to labelled image efficiency enhancement, nicknamed LETCP, for segmenting brain tumours. In particular, it offers a different pre-training approach to labelled data to prepare models for later use. In this method, the segmentation model is built using a self-attention transformer. Mehta et al. [179] present the Holistic Attention Transformer Network or HATNet. HATNet simplifies the histopathological image categorization pipeline and demonstrates how it learns representations from giga-pixel size images end-to-end, which contrasts with cutting-edge histopathological image classification systems, which use a two-pronged approach.

2.3.4 Drawbacks and Open Issues in Algorithmic-Level Approaches

Cost-sensitive learning and class weight adjustments, often fail to fully address imbalanced data distribution. While these methods penalize misclassification of minority class instances, they may not establish meaningful decision boundaries, particularly when the minority class is small and sparsely distributed. Additionally, models using these approaches can become overly complex or unstable, risking overfitting on the minority class or under fitting on the majority class.

Class imbalance also complicates performance evaluation, making common metrics like accuracy misleading. More appropriate metrics, such as F1-score, recall, and precision, are necessary but challenging to optimize through algorithmic means alone.

2.3.5 Hybrid techniques

Hybrid techniques effectively address class imbalance by combining data-level and algorithmic approaches. Data-level methods modify the training dataset, while algorithmic methods adjust the learning process to enhance minority class recognition.

In literature, hybrid methods involve merging base models to create a single optimal predictive model [180]. The combined ensemble algorithms have shown effective results with reasonable accuracy in handling multi-label imbalance problems. S. Dendamrongyit et al. [181] introduced two algorithms, easy ensemble and balance cascade, which train multiple classifiers by combining subsets of both the majority and minority classes. The hybrid sampling method was proposed by Cao et al. [182] where SMOTE is utilized to increase the quantity of minority class instances, followed by the application of the one side selection (OSS) algorithm to eliminate instances that do not provide useful information. Tang et al. [183] proposed an ensemble model with SVM and SMOTE. The SMOTE algorithm increases training data while SVM classifies the images. For learning more discriminative deep representations from unbalanced picture data, LMLE approach was introduced by Haung et al. [184] The authors proposed a novel loss function and sampling method to generate more discriminative features from the data. Recall value of this model was recorded at 70.13. The proposed LMLE method showed promising results on CelebA dataset. In order to address the high-class imbalance in the large-scale classification images, Dong et al. [185] proposed an end-to-end deep learning strategy. The drawback of this method is that, it cannot be applied to small-scale problems. Proposed architecture combined hard samples mining from the minority groups with a regulated loss function called class rectification loss (CRL). Table 2.1 shows the literature survey of various techniques.

Table 2.1 Literature Survey of various techniques.

Year	Reference	Approach	Model	Dataset used	Performance Measures(%)
2013	[186]	Hybrid Approach	5 ML Algo's	SDP data from public repository	AUC – 64 Gmean – 76 Recall – 82
2013	[187]	Algorithmic Approach	SVM-RFM	4 datasets	Accuracy – 97.6 Sensitivity – 23
2014	[188]	Algorithmic Approach	MWMOTE	20 datasets	Gmean – 23.2 ROC – 98
2014	[189]	Algorithmic Approach	16 Sorting Algo's	-	Accuracy:86 AUC: 98.6
2015	[190]	Data-level Approach	Ensemble method	Songho tans Hotel Reviews dataset	F-Measure 69 G-Mean -76 Weighted Acc - 82
2015	[191]	Data-level Approach	Hybrid resampling + SVM	VCI Dataset	AUC ROC - 79
2015	[192]	Algorithmic Approach	NN with ABC Algo	NASA MDP dataset	Accuracy: 68.4 AUC – 79
2015	[193]	Algorithmic Approach	Ensemble of DT, Random Forest, SVM,NB,K-NN	VCI Data & MiRNA	Sensitivity – 86,83,93 Specificity – 93,92,88.
2016	[194]	Algorithmic Approach	Ensemble of algorithms	16 Datasets	AUC-ROC - 805 Gmean -76.6
2016	[195]	Hybrid Approach	ML Algo & ensemble	Chemical vapor decompositio n process data	AUC – 91 Gmean – 92 F-Score – 69
2016	[96]	Data-level Approach	CNN	Mnist, Cipar, DLSVRC.	AUC ROC – 99.41 Over sampling :99.35. Under sampling : 96.85.
2016	[196]	Data-level Approach	Transfer learning + CNN	WHOI-plan database. 103 classes	-

2016	[197]	Hybrid Approach	LMLE-KNN	CelebA dataset	Accuracy – 84.
2017	[198]	Algorithmic Approach	ANN	Highly Imbalance	Gmean - 98.41 Sensitivity - 96.56 Specificity : 97.59
2017	[199]	Algorithmic Approach	BRKNN,HOMER,MCKNN	Uniprot KB	Apvalue – 85.891
2017	[200]	Algorithmic Approach	Proposed (ResNet- 34) CNN	Emotion Net 2017 challenge Track 1 Dataset	Accuracy – 82.2. F1 – 64.1.
2018	[201]	Hybrid Approach	Data Augmentation + CNN transfer learning Model + Dynamic Sampling Model	Own Dataset: 10000 images. 299*299 pixel.	Accuracy: 80.2 F1-score – 59
2018	[202]	Algorithmic Approach	MLP,CSDNN	Barnes Jewish hospital EMR databases.	Accuracy – 89, Specificity: 89, sensitivity: 26, F1-score- 48
2018	[185]	Hybrid Approach	CifarNet +CRL, ResNet +CRL, DenseNet +CRL,	3 imbalanced datasets were used celebA, X-domain, Deep fashion, 1 balanced dataset CIFAR-100.	Mean Accuracy CRL – 80.42
2019	[203]	Algorithmic Approach	CNN	CNN(multi class SVM)	Accuracy : 90.43
2019	[204]	Algorithmic Approach	ResNet50,VGG16	CBIS-DDSM	Sensitivity:86.7 Specificity:96.1 AUC-ROC – 98
2020	[205]	Hybrid Approach	3DCNN(UNet +DDS pool)	Own Dataset ultrasound images	Sensitivity – 95

				ABVS images	
2020	[206]	Algorithmic Approach	CNN(DenseNet169+max pooling)	IN breast data & CBIS dataset	Accuracy: 93.4 on IN Accuracy: 83.8 on CBIS
2020	[207]	Algorithmic Approach	CNN(SVM)	Own dataset. Total1952 cases. 437 malignant & 1222 benign	FFT: 66-77 DCT: 64-83 SSIM: 63-71
2021	[208]	Hybrid Approach	AlexNet, Ensemble architecture of multiple CNN's	-	Ppv: 98.57 Sensitivity: 98.58 Accuracy: 98.57 Specificity: 98.57
2022	[209]	Data-level Approach	Cross validation using random forest method+ SMOTE+ ML algorithms(Decision trees, KNN, Random Forest, SVM)	Mammography images, Highly imbalance	AUC:99
2023	[210]	Data-level Approach	AoADL-HBCC+SweezeNet	BreakHis Highly imbalances	Accuracy-96.77
2023	[211]	Algorithmic Approach	CNN-LSTM	BreakHis Highly imbalance	Accuracy: 92.5 for binary class and 99 for multi class.
2023	[212]	Algorithmic Approach	GARL-Net	BreakHis Highly imbalanced	Precision:99
2024	[159]	Algorithmic Approach	Residual Unet	BRATS 2020	Accuracy: 99 , Dice Score: 88
2024	[162]	Hybrid Approach	CACBL-Net	HAM-10000	Sensitivity: 90.60.

2.4 Generalisation

Generalization in deep learning refers to a model's ability to perform well on new, unseen data, beyond the examples it was trained on. In other words, a model generalizes effectively if

it can make accurate predictions for data it has never encountered before. This is crucial because, in real-world applications, models are expected to handle a wide range of inputs, and training on every possible scenario is impractical. Achieving strong generalization is challenging due to the complexity and over parameterization of deep learning models, which can sometimes lead to overfitting meaning the phenomenon where a model performs well on training data but poorly on new data. Thus, improving the generalization ability of deep learning models is a key focus of research, aiming to develop models that are both accurate and robust in diverse environments.

2.4.1 Literature survey of Generalisation techniques

In the field of machine learning, generalization refers to a model's ability to perform well on unseen data. In recent years, deep learning has achieved remarkable success in a variety of complex tasks, yet understanding and improving its generalization ability remains a core challenge. This thesis explores key research efforts that have contributed to enhancing the generalization performance of deep learning models, focusing on both theoretical and practical perspectives. The following sections provide a detailed overview of selected works that address various aspects of this challenge.

Generalization is strongly influenced by the learning efficiency of neural networks (DNNs). Lyle et al. [213] investigate how different phases of learning contribute to model generalization, particularly in supervised and reinforcement learning contexts. The study provides insights into how the parameters of deep networks evolve during training and how this evolution leads to generalization in new, unseen scenarios. By understanding the interaction between training data, model architecture, and learning algorithms, Lyle et al. [213] emphasize the importance of dynamic learning processes in enhancing generalization. This approach is critical in real-world applications where it is impossible to enumerate all possible inputs, thus necessitating models that can generalize beyond the training data.

Complementing this, Kawaguchi et al. [214] offer theoretical insights into why deep learning models are able to generalize so well, despite being highly over parameterized and complex. Their work focuses on providing non-vacuous generalization guarantees, which help explain why these models, which often fit training data perfectly, perform exceptionally well on unseen data. The authors argue that the architecture and training algorithms used in deep learning play a crucial role in achieving such generalization, even when traditional theoretical bounds for generalization are not satisfied. This study contributes to understanding why, in practice, deep learning models often outperform expectations derived from classical machine learning theory.

A more empirical approach to understanding generalization is taken by Aleksandar et al. [215], who explore complexity measures for deep learning models. Their research proposes new methods for predicting generalization based on local measures of distortion derived from approximation theory and information theory. These complexity measures, which are applied layer by layer within deep networks, offer an inexpensive way to predict how well a model will generalize to new data. The study finds that these measures correlate strongly with model performance, providing a practical tool for evaluating generalization without requiring access to extensive validation datasets. This research advances the ability to predict generalization, particularly in scenarios where training data is limited.

In the domain of medical imaging, generalization poses a unique set of challenges, as demonstrated by Yagis et al. [216]. The authors explore how convolutional neural networks (CNNs) perform in classifying neurodegenerative diseases, a task that often involves dealing with highly imbalanced and noisy data. The study highlights the impact of different data division strategies on the generalization capabilities of CNNs, showing that careful partitioning of data can significantly improve model accuracy and robustness. The paper underscores the

importance of considering data structure and variability in training, especially when working with medical data where generalization is critical for accurate diagnosis.

A novel approach to understanding generalization in deep learning is provided by Zhang et al. [217], who apply optimal transport theory to the problem. Their research shifts focus from traditional worst-case analysis to average-case scenarios, offering a new framework for evaluating how deep networks generalize. By framing generalization as a transport problem, they present a mathematical framework that explains the success of deep learning models in practical applications. This perspective provides a more refined understanding of how deep networks behave when confronted with unseen data, especially in scenarios where the training data is diverse and complex.

In the context of recommender systems, generalization is addressed through hybrid modelling techniques. Cheng et al. [218] introduce the concept of Wide & Deep Learning, which combines the memorization capabilities of wide linear models with the generalization strengths of deep networks. This approach is particularly useful in recommendation tasks where sparse data and user-item interactions present significant challenges. By leveraging both feature engineering and low-dimensional embedding's, the model enhances performance by capturing both high-level patterns and specific instances. The integration of wide and deep models represents a significant step forward in improving generalization for recommender systems and similar applications that deal with sparse datasets.

Another important contribution to the generalization discourse comes from Wilson and Izmailov et al. [219], who explore Bayesian deep learning as a way to improve model generalization. Bayesian methods, which involve marginalizing over model weights, provide a probabilistic framework for understanding generalization. This allows for more accurate predictions and better-calibrated uncertainty estimates, which are particularly important in applications where the consequences of errors are severe. Their work also addresses the

problem of overfitting by introducing Bayesian model averaging, which mitigates the effects of double descent, a phenomenon where increasing model capacity can initially worsen, but then improve, generalization.

Generalization in deep learning can also be enhanced through human-aided methods, as demonstrated by Boyd et al. [220]. The authors propose a novel approach in which human judgment about salient regions in images is incorporated into the training process. By leveraging human-generated saliency maps, the model is able to achieve higher accuracy and better generalization, particularly in cases where training data is limited. This method reduces error rates by guiding the model's attention to the most relevant parts of the input, which is especially useful in tasks where domain knowledge is crucial.

In the area of deep fake detection, Coccomini et al. [221] investigate how different deep learning architectures generalize when faced with the task of detecting manipulated videos. Their study compares convolutional neural networks (CNNs) and vision transformers (ViTs), revealing that while CNNs perform well with limited datasets, ViTs exhibit superior generalization capabilities when applied to more diverse data. The authors also highlight the promise of attention-based architectures like the Swin Transformer, which show strong generalization across various datasets, making them ideal for applications where dataset diversity is a challenge.

Finally, Mungoli et al. [222] proposes an Adaptive Feature Fusion (AFF) framework designed to improve generalization in deep learning models. By dynamically adapting how feature representations are fused across layers, the AFF framework leads to better performance across a variety of tasks. This approach outperforms traditional feature fusion techniques, offering a flexible method for enhancing generalization. Extensive experiments demonstrate that AFF is particularly effective in tasks requiring robust feature extraction and generalization, making it a valuable addition to the deep learning toolkit.

The generalization of deep learning models has been a topic of significant research, with contributions spanning theoretical, empirical, and practical domains. From understanding the dynamics of learning to applying novel frameworks like optimal transport and Bayesian inference, researchers continue to push the boundaries of how deep networks can generalize beyond their training data. These works highlight the importance of model architecture, training strategies, and data partitioning in achieving strong generalization performance, ensuring that deep learning models are robust and reliable in real-world applications.

2.5 Publically Available Datasets

The following Table 2.2 shows the statistical analysis on the publically available datasets along with their degree of imbalance.

Table 2.2 Statistical analysis on the publically available datasets.

Dataset Name	Description	Number of images	Type of classification	Imbalance ratio
DDSM	DDSM is a database containing scanned mammography images.	10239	Multi-class	Partially
Figshare	It is publically dataset containing raw images obtained from patients.	3064	Multi-class	Highly
Br35h	Br35h dataset is designed for Binary classification	3000	Binary-class	Highly
BrainTumor MRI	This dataset is obtained from Kaggle which is a combination of three datasets namely Figshare, SARTAJ and Br35h.	7021	Multi-class	Partially

Brain MRI	The dataset consists of brain MRI images	256	Binary-class	Highly
BT-small 2c	Contains Brain tumor images	253	Binary-class	Highly
BT-large 2c	Contains Brain tumor images	3000	Binary-class	Highly
BT-large 4c	Contains Brain tumor images	3064	Binary and multi-class	Highly
BreakHis	Contains histopathological images of breast cancer obtained from 82 patients at different magnifications.	9109	Binary-class	Highly
BACH	It is composed of histopathological images related to breast cancer, sourced from Breast cancer 2019 grand challenge	400	Binary-class	Partially

2.6 Performance Evaluation Metrics

Table 2.3 presents the performance evaluation metrics utilized in this study to assess the

robustness of the proposed models.

Table 2.3 Performance Evaluation Metrics.

Metrics	Definition	Formula	Range
Accuracy	Measure how well the model predicts the correct labels.	$Accuracy = \dfrac{TP + TN}{TP + FP + TN + FN}$	0-1

Precision	Measure of how many of the predicted positive labels are actually positive.	$$Precision = \frac{TP}{TP + FP}$$	0-1
Recall	Measure of how many of the actual positive labels are correctly predicted.	$$Recall = \frac{TP}{TP + FN}$$	0-1
F1-Score	The harmonic mean of Recall and Precision.	$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$	0-1
MCC	Used to compare the anticipated labels with actual labels in order to assess how well a classification model is performing.	$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$	0-1

2.7 Research Gaps

The literature survey from Table 2.1 reveals several noteworthy challenges and limitations. These include:

> ### *Computational Resource Requirements*

Hybrid models for image processing, especially deep learning-based systems like CNNs and ViTs, are computationally intensive. Training these models requires substantial processing power, memory, and time, which makes them difficult to deploy in environments with limited infrastructure. For instance, real-time image processing in applications like surveillance or medical diagnostics in rural areas may not have access to high-performance

GPUs or cloud resources. Research into lightweight, efficient models that maintain high accuracy but operate on constrained hardware is still lacking.

➢ *Scalability and Generalization*

Many hybrid models perform well on specific image datasets, but struggle to generalize across various image types, modalities, or applications. For instance, a model trained on satellite images may not perform as well on medical or industrial images without significant retraining. The challenge of scalability is further compounded when attempting to deploy models across different contexts, such as different imaging technologies, environments, or populations. The lack of generalization and scalability across diverse datasets and tasks presents a critical gap that needs to be addressed for broader real-world adoption.

➢ *Synthetic Data Quality*

Techniques like SMOTE and Generative Adversarial Networks (GANs) are often used to generate synthetic image data to balance class distributions in image processing. However, the quality of synthetic images is often insufficient to capture the true complexity and variability of real-world data. Poorly generated synthetic images can lead to overfitting or misclassification, especially in high-dimensional data like satellite imagery or medical scans, where minute details are crucial. Ensuring that synthetic data mimics real data effectively is a challenge that requires further research.

➢ *Handling of Noisy and Artifact-Laden Data*

In many real-world image processing applications, data is often noisy or contains artifacts that can obscure important features. For instance, images captured in low-light environments or medical scans affected by movement artifacts present significant challenges. Many current hybrid models lack robust methods for handling such imperfections, which limits their utility in practical settings. Developing more noise-resilient and artifact-tolerant models

is essential for improving performance in dynamic environments such as real-time surveillance or remote sensing.

> ➤ *Computational Complexity*

The use of hybrid models often introduces additional computational complexity due to adaptive weighting, rebalancing, and combining multiple models or techniques like resampling. For image processing tasks that require real-time or large-scale processing, such complexity can become prohibitive. In applications such as traffic monitoring or large-scale industrial inspections, the additional computational overhead can slow down processing times, making these models unsuitable for time-sensitive tasks. Reducing computational complexity without sacrificing accuracy remains a significant research gap.

To address these gaps, it is crucial to explore innovative approaches capable of overcoming issues such as computational resource constraints, scalability, and generalization across diverse datasets. Our research endeavors to develop hybrid deep learning techniques that effectively balance computational efficiency with performance, enabling accurate and reliable processing of diverse types of image data. By integrating advanced methodologies, we aim to enhance the robustness and adaptability of hybrid models, thus addressing the evolving demands of real-world image processing applications.

Chapter Three: A COMPREHENSIVE EXPLORATION OF HYBRIDIZED ARCHITECTURES TO MITIGATE CLASS IMBALANCE

This chapter provides the overview of comprehensive exploration of how effectively models are handling class imbalance condition when various data-level and algorithmic level modifications are done. Results obtained highlight the importance of hybrid models in addressing class imbalance.

3.1 Introduction

In the era of Industrial Revolution 4.0, deep learning (DL) and machine learning (ML) have gained significant attention for their broad applications across sectors like energy management, social network analysis, medical informatics, and computer vision [46]. ML and DL models are broadly categorized into supervised and unsupervised algorithms, with supervised models further split into classification and segmentation tasks. Despite extensive training, researchers have noted performance declines in real-world applications, attributed to various factors including insufficient data quality, highlighting the need for robust models that can generalize effectively [223].

One of the main challenges in ML/DL models is the issue of class imbalance, where the unequal distribution of classes negatively impacts model performance. This problem is particularly evident in tasks like classification, object detection, and image segmentation. Class imbalance is seen in numerous applications such as fraud detection, medical decision-making, and text classification, and new techniques like data re-sampling and hybrid models have emerged to mitigate this issue. However, handling multi-label classification presents even more complexity, as multiple labels within a class exacerbate the imbalance, requiring approaches like problem transformation and algorithmic adaptations to improve results [224].

In the medical field, the challenge of class imbalance is critical, particularly in breast cancer detection, where early diagnosis can save lives. Breast cancer, one of the most prevalent and deadly cancers globally, is characterized by imbalanced datasets that hinder accurate model predictions. In this context, the study performed a comprehensive analysis of class imbalance and proposed a hybrid model using the BreakHis dataset, which addresses this imbalance in breast cancer classification. This study highlights the limitations of existing state-of-the-art DL models and introduces a solution aimed at improving performance in high-imbalance medical datasets [51].

3.2 Materials and methodology

This section highlights the various performance evaluation metrics employed to assess model effectiveness, accompanied by a detailed description of the datasets used and the experimental setup established for consistency. It also outlines the data-level and algorithmic-level approaches, including hybrid methods that combine these modifications to address class imbalance. Furthermore, we examine the diverse deep learning techniques utilized in our comparative analysis, emphasizing their strengths and suitability for the tasks at hand. This comprehensive overview establishes a solid foundation for understanding the effectiveness of the strategies applied in this research.

3.2.1 Various Data-level and Algorithmic Level Modification

The researchers working on classification methods has shown a fascinating interest in resolving issues that arise while working on imbalanced datasets. The lack of labelled data causes the classification or segmentation algorithms to generate biased results or the skewness problem. Skewness in the dataset prevents the efficient performance of classification problems. Further, various data level and algorithmic approaches are implemented and their performance is recorded.

3.2.1.1 Data Augmentation and Pre-processing

Data augmentation and pre-processing are employed in machine learning, especially when there's a small dataset that could lead to overfitting and imbalanced labels. Data augmentation helps by creating more training examples from existing data, making the model better at generalizing. Pre-processing, on the other hand, cleans and improves the quality of images by removing noise, which helps the model analyse them more accurately. By using both these techniques, we can create a variety of data samples and prepare the images to handle issues like label imbalance, making the model stronger and more reliable.

Random Cropping: This technique involves cropping multiple unique sections from an image by randomly selecting valid corner points. As a result, the cropped images maintain distinct characteristics, promoting diversity in the training set.

Rotation: Images are rotated by a predefined angle to ensure that they remain clear and recognizable, thus allowing the model to learn from different orientations.

Colour Shifting: In this method, numerical adjustments are made to the red, green, and blue (RGB) channels of an image. This creates various color distortions, helping the model learn to recognize objects under different lighting conditions and color variations.

Flipping: Images are flipped either horizontally or vertically, which aids the model in generalizing across different angles and orientations.

Intensity Variation: This technique modifies the brightness of images by adjusting their intensity levels, making them either brighter or darker, thus exposing the model to a range of lighting conditions.

Translation: Image pixels are shifted within a specified range, normalizing the images and ensuring that key features remain consistent across variations.

Resizing: This technique adjusts the dimensions of input images, either reducing or enlarging their sizes. Resizing ensures that images are tailored to meet the specific requirements of

70

machine learning (ML) or deep learning (DL) models, facilitating efficient processing and analysis.

Rebalancing the Classes: Previous research has demonstrated that imbalanced datasets, where one class significantly outnumbers others, can introduce bias and yield inaccurate predictions. To mitigate this issue, rebalancing techniques are applied, which involve removing redundant instances from overrepresented classes to create a more equitable distribution of classes.

Normalization: Normalization involves scaling pixel values to a range between 0 and 1, thereby reducing the variance in colour distribution and intensity. Techniques such as strain normalization with colour contrast and strain normalization without colour contrast are employed to ensure consistent representation across images.

Image Contrast Enhancement: This technique improves the overall brightness of an image by adjusting the mapping of pixel values from the lowest to the highest grayscale levels. By enhancing the contrast, this technique makes important features in the images more distinguishable, thereby aiding in subsequent analysis.

3.2.1.2 Loss function

Employing loss functions as algorithmic-level approaches is essential for optimizing model performance, particularly in addressing class imbalance. Traditional loss functions may not effectively handle imbalanced datasets, leading to biased predictions favouring majority classes. Custom loss functions, such as focal loss or class rectification loss, assign different weights to classes, enabling the model to focus on minority classes and improve overall classification accuracy. Additionally, specialized loss functions can enhance model generalization by penalizing overfitting and accommodating noisy labels in real-world datasets. The following equations (3.1), (3.2), (3.3), (3.4), (3.5), (3.6), (3.7) are employed for performing algorithmic level modifications:

Cross entropy [243]:

$$L_{BCE}(y, \hat{y}) = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y})) \tag{3.1}$$

Here, $\hat{y}$ is the predicted value by the prediction model.

Focal loss [132]:

$$CE = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \tag{3.2}$$

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \tag{3.3}$$

$$CE(p, y) = CE(p_t) = -\log(p_t) \tag{3.4}$$

By applying a modifying factor, focal loss suggests minimising simple instances and concentrate training on difficult negatives, $((1-p)t)^\gamma$ as shown below:

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \tag{3.5}$$

Here, $\gamma > 0$ and when $\gamma = 1$ Focal loss works like Cross-Entropy loss function.

Dice loss:

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1} \tag{3.6}$$

When $y = \hat{p} = 0$

Class Rectification loss:

$$L_{bln} = \alpha L_{crl} + (1-\alpha)L_{ce}, \quad \alpha = \eta\Omega_{imb} \tag{3.7}$$

In this case, the parameter α is intended to have a linear correlation with the training class imbalance metric Ω_{imb}. Considering varying sizes of individual class data samples, Ω_{imb} is determined as the minimum percentage count of data samples that must be present across all classes to achieve an overall uniform or balanced class distribution in the training data. Additionally, the model hyper-parameter η can be estimated through cross-validation, regardless of the imbalance between individual labels, as it is not affected by it.

3.2.1.3 Dataset

The BreakHis dataset, a widely used and publicly available resource, has been employed for this study to address the binary classification of breast tumor tissue into benign and malignant categories. The dataset comprises a total of 9,109 high-resolution microscopic images, each with a resolution of 700×460 pixels. These images are captured under four different magnification levels: 40X, 100X, 200X, and 400X, providing a diverse range of visual details for analysis.

The dataset is organized into two primary classes: benign tumors and malignant tumors. The benign class represents non-cancerous tissue samples, while the malignant class includes cancerous tissue samples. A notable challenge in this dataset is the inherent class imbalance, as the malignant class contains 5,429 images, whereas the benign class is represented by only 2,480 images. This uneven distribution between the two categories introduces challenges in training machine learning models, as it can lead to biased predictions favouring the majority class if not properly addressed.

For experimental purposes, the dataset has been divided into three subsets to facilitate systematic development and evaluation of classification models. The training set is used to learn patterns and features distinguishing benign from malignant samples. The validation set is employed for hyper-parameter tuning and monitoring model performance during training to prevent overfitting. Finally, the testing set is reserved for assessing the model's performance on unseen data, ensuring a fair evaluation of its generalization capabilities.

3.2.1.4 Experimental setup and Data Augmentation

The fastai v2.7.10 library with Python 3.10 was utilised on a personal machine equipped with an Nvidia RTX 3080 GPU to carry out the experiments. Due to the limited VRAM in our GPU, we used a batch size of 16 for our experimental setup. The dataset's images were divided

into three distinct groups: 60% were allocated for training purposes, 20% were reserved for validation and remaining 20% were set aside for testing. The dataset was partitioned into sets, with each set including the same proportion of images from each of the dataset's 32 classes. Because we are employing pre-trained models that require images of size 224x224 pixel, all images were shrunk to 224x224 pixel. During the training process, various data augmentation transforms were applied to the mini-batches, which helped to increase the amount of data by including slightly modified copies of already existing data, resulting in less over fitting. These augmentations included random flips with a probability of 0.5, random affine transforms with a probability of 0.75, and random changes in brightness and contrast with a probability of 0.75. The affine transforms included random rotation of 10 degrees, random zoom, and perspective warping. The models were fine-tuned on the training set for 61 epochs each. Only the last layer is trained during the first epoch, and the rest of the model's pre-trained weights are frozen. All of the model's weights were unfrozen for the next 60 epochs, and the entire model was trained. The model weights which gave the best accuracy on the validation set after each epoch were selected as the final model weights.

3.2.2 Techniques Employed

3.2.2.1 ResNet-50

ResNet-50 is a convolutional neural network (CNN) architecture as shown in Figure 3.1 is designed to tackle the vanishing gradient problem through the innovative use of residual connections. These connections facilitate more effective gradient propagation by allowing gradients to bypass certain layers during backpropagation. This shortcut mechanism enhances the flow of gradients directly from the output to the input of each residual block, thereby preserving the gradient and enabling more efficient training of deeper networks. The architecture's design significantly improves training efficiency, allowing for effective learning in complex tasks.

The formula to calculate the residual block in ResNet is given in equation in (3.8):

$$H(x) = F(x) + x \tag{3.8}$$

Where x is the input to a residual block,

F(x) is the output of a set of convolutional layers and non-linear activation function applied to x and H(x) is the output of the residual block.

The residual connections in the formula allows the gradient to flow directly from the output to the input through the residual block, bypassing the convolutional layers and activation functions. This helps to preserve the gradient and prevent it from vanishing, allowing for deeper networks to be trained more effectively.
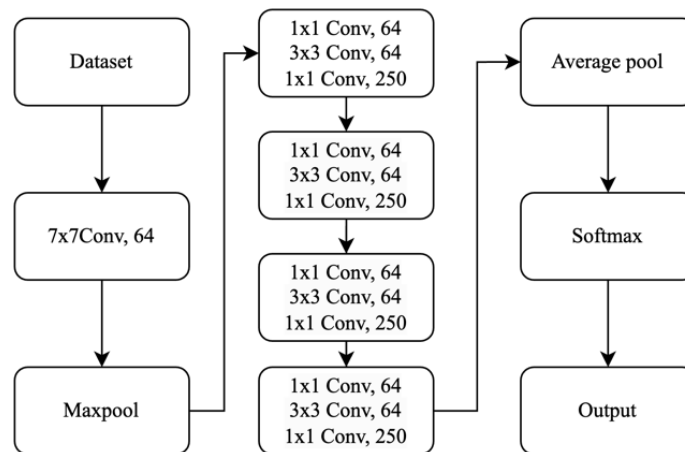
Figure 3.1 Network Architecture diagram of ResNet-50

3.2.2.2 Efficient Net

Efficient Net model as shown in Figure 3.2 is a family of CNN architectures that balances performance with computational efficiency in image classification tasks. It employs a unique scaling methodology that uniformly adjusts the network's depth, width, and resolution using a compound coefficient. Instead of relying on arbitrary scaling, Efficient Net uses a systematic approach to determine scaling coefficients through grid search optimization. This methodology ensures that the network maintains superior performance while adhering to

computational constraints. By carefully scaling these parameters, Efficient Net constructs highly efficient architectures capable of delivering excellent classification results across various applications.

The equations (3.9), (3.10), (3.11) are employed to calculate the scaling coefficients of n different dimensions is given by:

$$Depth = \alpha * \pi \tag{3.9}$$

$$width = \beta * \pi \tag{3.10}$$

$$Resolutions = \delta * \pi \tag{3.11}$$

Where:

π is a user-defined coefficient that controls the overall network size.

α , β and δ are scaling coefficients that control the network depth, width and resolution, respectively. These coefficients are determined through a grid search optimization process, where their values are chosen to maximize the networks performance subject to a constraint on computational cost.

Once the scaling coefficients are determined, the network depth, width and resolution can be calculated using the formulas above. The Efficient Net architecture is then constructed by stacking convolutional layers and other operations in a specific order.
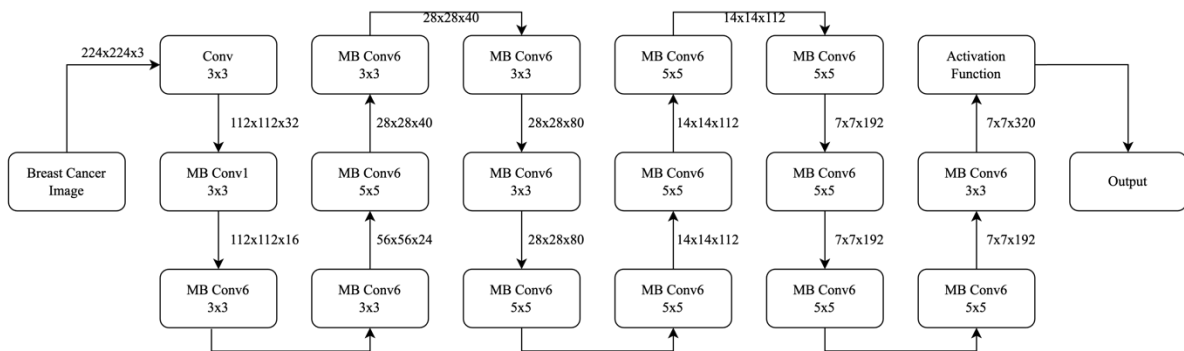
Figure 3.2Network architecture diagram of Efficient Net.

3.2.2.3 LeViT

LeViT is a hybrid architecture that combines the strengths of convolutional neural networks and Vision Transformers. It utilizes a hierarchical structure as shown in Figure 3.3 to enhance performance and efficiency when processing input images for classification tasks. The architecture features a feature extractor composed of convolutional and pooling layers, which generates a rich representation of the input image. The output undergoes processing through multiple layers, including layer normalization and multi-layer perceptron (MLP) modules, before generating class probabilities. This design allows LeViT to effectively capture complex patterns in data, resulting in improved classification accuracy and efficiency.

The equation (3.12) used to calculate the output of a LeViT block is:

Where:

$$LeViT(X) = Softmax(W4 * MLP2(LN2(MLP1(FE(X))) * W3) * W2) * W1) \quad (3.16)$$

Where X is the input image, FE is a feature extractor that applies a set of convolutional and pooling layers to the input image to generate a feature map.

LN1 and LN2 are layer normalization functions W1, W2 and W3 are learnable weight matrices, and softmax is the activation function used to generate class probabilities. The LeViT architecture uses a set of liner projections to transform the feature map generated by the FE function into a sequence of tokens, which are then processed by a set of Transformer layers. The output of the Transformer layers is then processed by an MLP module (MLP1), followed by a layer normalization (LN2), another MLP module (MLP2), and finally a set of fully connection layers with softmax activation to generate class probabilities.
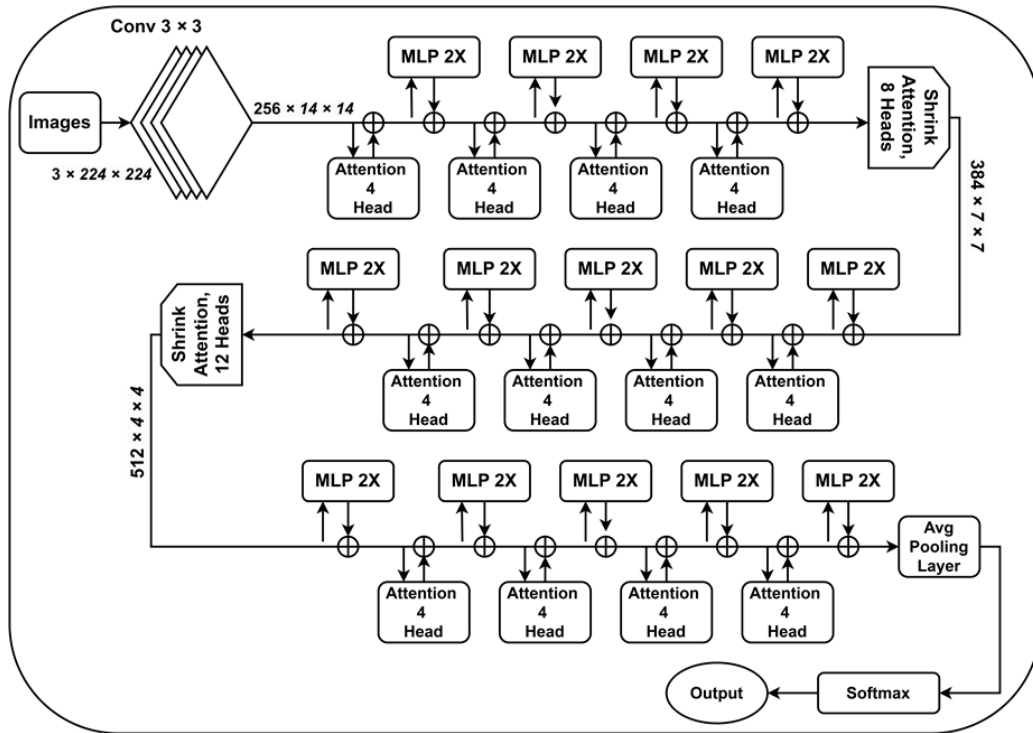
Figure 3.3 LeViT architecture.

3.2.2.4 ConvNeXT

ConvNeXT leverages a combination of grouped convolutions and cardinality to enhance the representational capacity of the network while maintaining memory efficiency. Grouped convolutions divide input channels into distinct groups, enabling the network to capture various features in parallel. The architecture as shown in Figure 3.4 incorporates residual connections that facilitate improved gradient flow and training dynamics. Additionally, ConvNeXT includes a feature recalibration module, which adapts the feature maps based on their channel-wise statistics. This adaptive mechanism further enhances the network's ability to learn intricate representations, allowing it to effectively balance performance and resource utilization.

The formula used in the ConvNeXT architecture is given by:

$$y = F(x, \{w\_i\}) + x \tag{3.17}$$

where:

78

x is the input to the residual block, F is a sequence of convolutional and activation layers' hat transform the input x into a new representation, Wi are the weights of the layers in F. + denotes the element wise addition operations.

Overall, the ConvNeXT architecture is a powerful and memory efficient variant that uses grouped convolutions and cardinality to increase the representational power of the network.
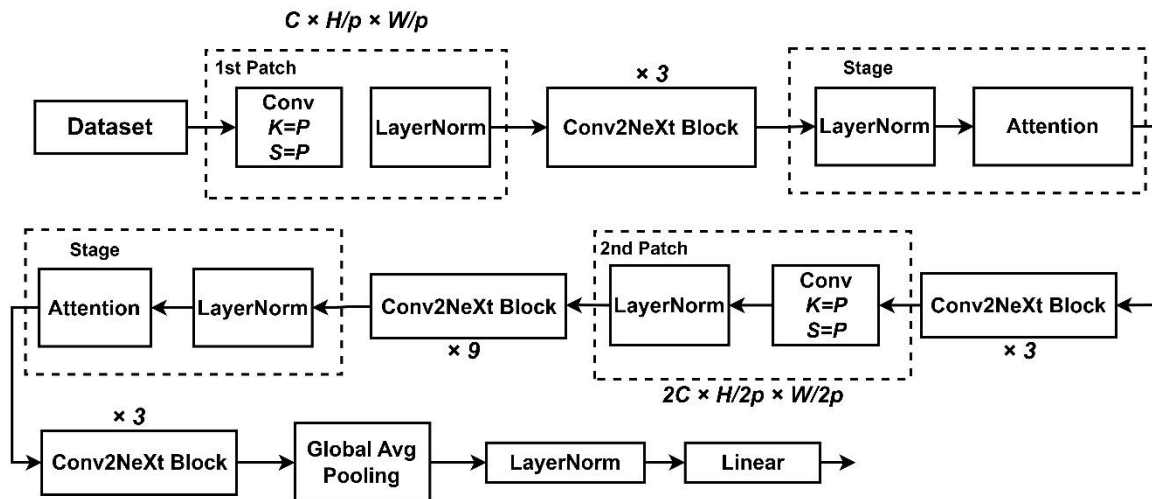
Figure 3.4 Network architecture diagram of ConvNeXT model.

3.2.2.5 Inception-ResNet V2

Inception-ResNet V2, introduced by Google in 2016, is a complex deep convolutional neural network architecture characterized by multiple interconnected components as shown in Figure 3.5. It begins by processing a 299x299 RGB image through a stem, which consists of convolutional layers, pooling layers, and normalization layers. This initial stage reduces the spatial dimensions of the image while increasing the number of channels. The architecture's core consists of Inception-ResNet blocks, which combine the Inception module with residual connections. These blocks allow the network to bypass specific layers, enhancing its learning capabilities. Additionally, reduction blocks are interspersed among the Inception-ResNet blocks to further reduce spatial dimensions while amplifying channel numbers. The

classification head processes the output of the final Inception-ResNet block through a global pooling layer, which averages each feature map's values and generates a vector sent to a fully connected layer for the final classification output. This design exemplifies a powerful approach to deep learning in complex image classification tasks.

The formula for describing the Inception ResNet V2 architecture can be structured as follows:

$$Branch\_i\,(x) = f\_i\,(x) \tag{3.18}$$

Where, $f_i\,(x)$ represents the operations such as convolution with different filter sizes applied in each $Branch_i$.
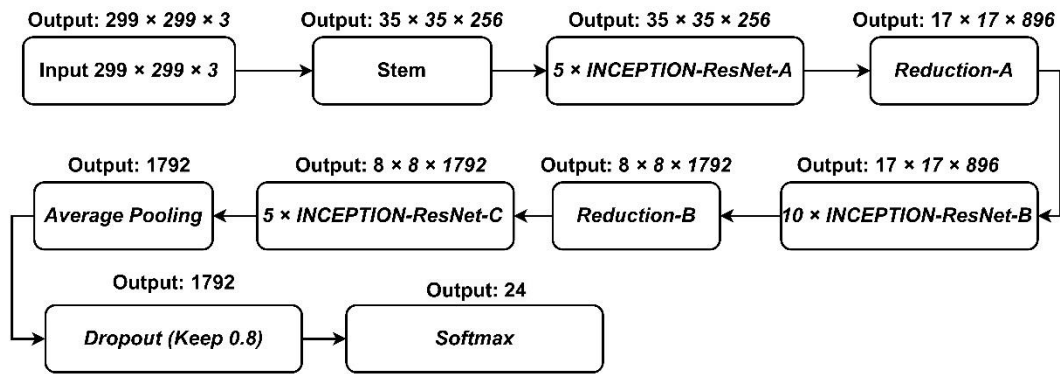
Figure 3.5 Network architecture diagram of Inception ResNet V₂ model.

3.3 Results and Discussion

The primary objective of this research is to analyse class imbalance issues and propose effective solutions to address them. To achieve this, a comprehensive study on class imbalance and management techniques was conducted. A hybrid approach was developed using the BreakHis dataset, accompanied by a comparative analysis involving multiple experiments based on two foundational architectures: Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs).

The CNN variants utilized in this study include ResNet-50, Inception-ResNet V2, and EfficientNet, while LeViT and ConvNeXT were employed as the Vision Transformer models. To highlight the effectiveness of the hybrid models in mitigating class imbalance, two key

experiments were performed, and the results were rigorously compared. In the first experiment, the base models were trained using data-level modifications. Various data augmentation and pre-processing techniques were applied to the dataset, which was then fed into the models using pre-trained weights from ImageNet. The results obtained from different deep learning models after applying these data-level modifications are summarized in Table 3.1.

The findings in Table 3.1 highlight notable differences in model performance on the testing set. ResNet50 and Inception ResNet V2 exhibited poor outcomes, which can be attributed to their large parameter counts and computational complexity. These models are prone to overfitting and require extensive hyper-parameter tuning, making their optimization process time-intensive. Additionally, their deep architectures often face challenges such as gradient vanishing or exploding, hindering convergence, particularly on smaller datasets like BreakHis.

EfficientNet demonstrated superior performance by leveraging uniform scaling of width, depth, and resolution, achieving an excellent balance between accuracy and computational efficiency. Vision transformers, including LeViT and ConvNeXT, have further advanced image classification by integrating convolutional and transformer layers. ConvNeXT, for example, enhances ResNet by introducing a patchify layer and depth-wise convolutions, enabling the model to learn from each channel independently and significantly improving its performance on imbalanced datasets.

The success of EfficientNet, LeViT, and ConvNeXT underscores the effectiveness of hybrid architectures that combine CNN methods with transformer-based innovations. These models excel in addressing challenges associated with testing on imbalanced datasets, highlighting their potential for broader applications.

Table 3.1 The training and testing scores of various DL models with data-level modifications.

Training score					
Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score(%)**	**AUC-ROC**
ConvNeXT	**98.99**	**99.54**	**99.45**	**99.26**	**99.95**
EfficientNet	98.93	99.26	99.17	99.22	99.83
LeViT	98.48	98.99	98.80	98.89	99.69
ResNet 34	97.66	98.34	98.80	98.30	99.63
Inception ResNet V$_2$	97.41	98.07	98.53	98.11	98.89
ResNet 50	95.51	96.09	97.61	96.75	98.41
Testing score					
Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC-ROC**
ConvNeXT	**98.30**	**99.0**	**98.64**	**98.76**	**98.89**
EfficientNet	98.26	98.90	98.67	98.54	98.87
LeViT	98.32	98.12	98.34	98.19	99.09
ResNet 34	97.02	98.64	99.02	98.12	99.08
Inception ResNet V$_2$	97.02	97.78	97.92	97.24	97.98
ResNet 50	95.20	95.68	96.68	96.79	99.00

Figures 3.6, 3.7, 3.8, 3.9, and 3.10 illustrate the performance of various deep learning (DL) models, modified at the data level, based on metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The results reveal that ConvNeXT achieved the highest training accuracy at 98.99%, followed closely by EfficientNet at 98.93%, LeViT at 98.48%, ResNet-34 at 97.66%, Inception ResNet V2 at 97.41%, and ResNet-50 at 95.51%. While ConvNeXT outperformed the other DL models, its performance still fell short when compared to existing state-of-the-art models, prompting further investigation into the underlying reasons for this

82

decline. A thorough analysis indicated that the decreased performance was primarily due to

class imbalance, which contributed to overfitting issues within the model.
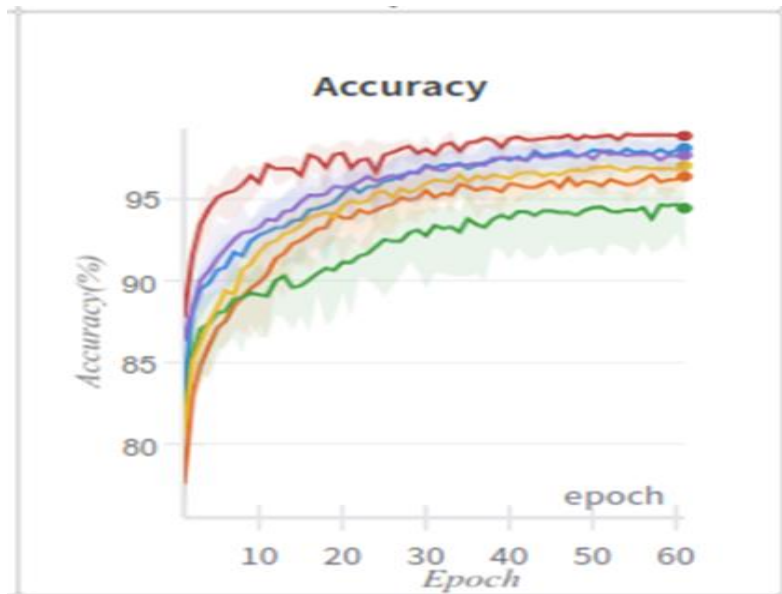
Figure 3.6 Test accuracy of DL models
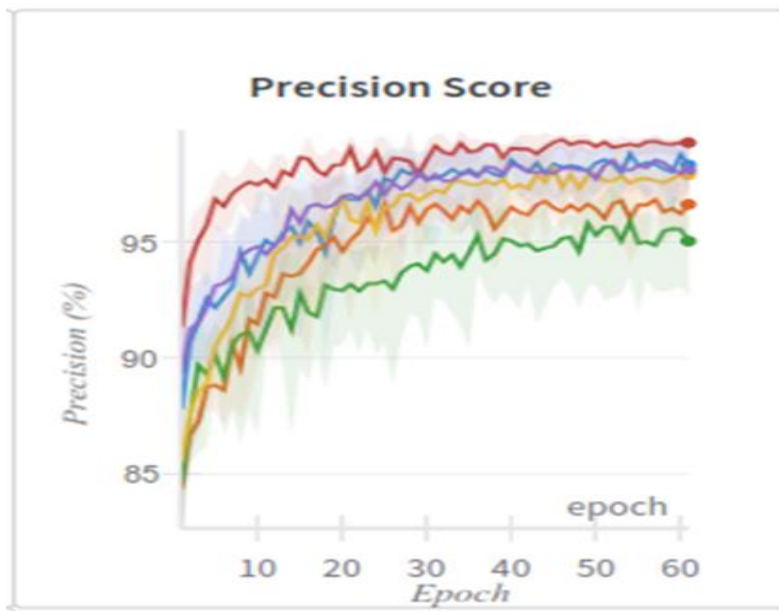
Figure 3.7 Test precision of DL models

Figure 3.8 Test Recall of DL models
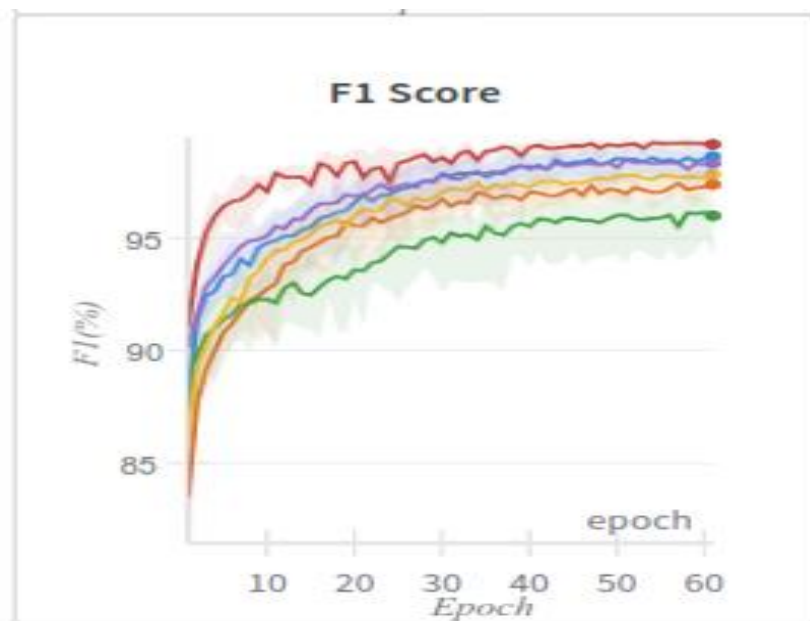
Figure 3.9 Test F1-Score of DL models
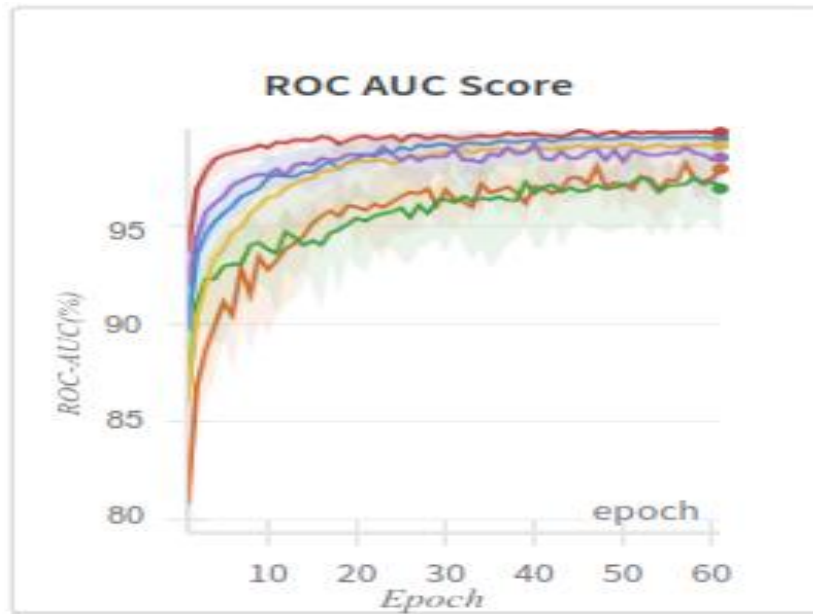
Figure 3.10 Test ROC-AUC of DL models

Table 3.2 summarizes the performance of the proposed model with respect to the different loss functions. The hybrid model developed in this study achieved impressive scores of 99% accuracy and 99.2% precision, significantly outperforming other models. A model is generally considered effective when its validation accuracy exceeds its testing accuracy. As shown in Table 3.2, models such as ConvNeXT, EfficientNet, and LeViT exhibited a reasonable difference between validation and testing accuracy when utilizing the Class Rectification Loss (CRL) function, suggesting that they effectively addressed underfitting or overfitting and yielded reliable classification results.

This section also evaluates how adjustments to the loss function can address multi-class imbalance issues. The selection of an appropriate loss function plays a crucial role in influencing model performance. The analysis highlights that the CRL function outperforms others in this context. While cross-entropy loss is commonly used to quantify differences between probability distributions, it tends to favour majority classes, making it less effective

for imbalanced datasets. Dice loss, which measures similarity, treats all samples equally and leads to a bias towards more frequent classes.

As illustrated in Table 3.2, CRL recalibrates predicted probabilities based on their similarity to true classes, adjusting them to enhance classification accuracy. By reducing the predicted class probabilities for dissimilar classes and increasing those for similar ones, CRL fosters a more balanced distribution. The application of CRL has led to significant improvements in model performance, particularly in predicting minority classes. These findings underscore the importance of selecting appropriate model architectures and loss functions to enhance deep learning performance, especially in critical areas like medical image classification.

Table 3.2. Comparison of result obtained from different models with respective to different loss functions.

Model	Loss Function	Accuracy (%)	Precision (%)
ConVneXT	**CRL**	**99.0**	**99.2**
ConVneXT	Dice	99.04	99.5
ConVneXT	CEL	99.17	99.6
ConVneXT	Focal	99.55	99.9
Efficient NET	**CRL**	**98.92**	**99.2**
Efficient NET	Dice	97.26	97.2
Efficient NET	CEL	98.47	98.9
Efficient NET	Focal	98.41	98.7
Inception ResNet V_2	CRL	97.77	98.1
Inception ResNet V_2	Dice	96.62	96.4
Inception ResNet V_2	CEL	97.26	96.38

Inception ResNet V$_2$	Focal	96.94	97.9
ResNet	CRL	96.9	96.9
ResNet	Dice	93.7	93.9
ResNet	CEL	97.1	97.0
ResNet	Focal	96.3	96.7
LeviT	**CRL**	**98.03**	**98.3**
LeviT	**Dice**	**97.20**	**97.09**
LeviT	CEL	98.47	99.07
LeviT	Focal	97.90	97.90

3.3.1 Comparison with state of art

This section represents comparison of different state of art models proposed in literature with our proposed model. The above mentioned can be analysed from Table 3.3 listed below.

In this section, a comparison between the proposed hybrid model and existing state-of-the-art models for classifying breast cancer using the BreakHis dataset is analysed and is presented in Table 3.3. The proposed hybrid model demonstrates superior accuracy and precision for multi-class imbalanced classification, outperforming other models. Among the various deep learning hybrid models, ConvNexT exhibited the best performance. The use of the CRL loss function further improved the performance of the ConvNexT model. In the ConvNexT model, depth-wise convolutions were employed, where the convolutional layers were grouped, allowing the model to learn from each group. The inclusion of a down sampling layer between the stages enabled the extraction of spatial information from every group, without losing useful information. The CRL loss function addressed the imbalance issue by scaling the probabilities of the classes based on their similarity to the positive classes. This allowed for the identification of classes that are more similar to positive classes. The proposed

hybrid model utilizes the ConvNexT model to extract spatial information and the CRL loss function to reduce overfitting and accurately predict values on the BreakHis dataset.

Table 3.3 comparison of proposed model with different state of art models using BreakHis dataset.

Year	Dataset	Model	Accuracy(%)
2018	BreakHis	MVP-Net	92.2
2018	BreakHis	AlexNet	91.05
2019	BreakHis	SE-ResNet	99.06
2019	BreakHis	Inception-V$_3$	98
2022	BreakHis	ResNet	97.81
2022	BreakHis	AlexNet-BC	98.48
2023	BreakHis	Self-stacked Random Forest+AdaBoost	99.0
2023	BreakHis	Hybrid CNN	94.49
Our Approach	**BreakHis**	**Proposed ConvNexT-CRL**	**99.2**

3.4 Chapter Summary

This chapter provides a detailed look at the class imbalance problem, explaining its impact on model performance and exploring methods to address it. Through a survey, we identified two main approaches to tackle class imbalance: data-level methods (which focus on modifying the dataset itself) and algorithmic-level methods (which adjust how the model learns). While each approach has its strengths, our analysis suggested that a hybrid approach, combining both data-level and algorithmic-level techniques, might offer better results. To test this, we conducted three experiments. In the first experiment, we applied only data-level modifications, like resampling and data augmentation. However, the results showed that these changes alone didn't fully solve the imbalance problem. In the second experiment, we used only algorithmic-level adjustments, such as adjusting the cost of misclassification for different classes. Although these changes improved results slightly, they still weren't enough on their

own. Finally, in the third experiment, we combined both data-level and algorithmic-level techniques. This hybrid approach produced significantly better results, showing that it handled class imbalance more effectively than using either method alone. The findings from this study suggest that a hybrid technique is a more powerful solution for managing class imbalance, offering more reliable and robust model performance when working with imbalanced data.

Chapter Four: DATA AUGMENTATION AND AUXILIARY NEURAL NETWORKS FOR ADDRESSING CLASS IMBALANCE

This study highlights the potential of GANs in generating additional data and using auxiliary network in improving the accuracy and efficiency of the model while working on imbalanced datasets.

4.1 Introduction

The aberrant proliferation of cells in and around the brain is what distinguishes a brain tumor (BT) from a normal brain. The cells can be formed inside the brain tissue or near the tissue. Brain tumor is a life threatening condition and cause several medical issues. Sensation loss, hearing loss, visual impairment, migraines, nausea, and seizures are only some of the medical issues brought on by BT. Not all tumor formations in the brain are critical. They can be noncancerous or benign. Meningioma, the most common benign brain tumor, develops from the membrane that covers the brain. Malignant brain tumors are relatively uncommon, but the most common types are gliomas, which originate in glial cells and the brain stem, along with glioblastomas, which originate in the brain itself [225].

Brain Tumors are commonly diagnosed using Magnetic Resonance imaging (MRI). MRI images show the tumors with different pixel intensity which contrasts from nearby normal tissues. MRI images are then used by neurosurgeons to correctly diagnose the condition. Detecting tumors from MRI images provides an opportunity for automation of diagnosis using artificial intelligence methods. Image classification and detection algorithms in artificial intelligence technologies have surpassed the human ability to process and analyze images. Several studies have attempted automated diagnosis of brain tumors using MRI scans, often starting with traditional machine learning classifiers like support vector machines (SVMs), k-

nearest neighbors (KNNs), and random forests (RFs) trained on manually-crafted properties of MRI slices [226].

Traditional machine learning classifiers have several limitations when it comes to image classification and detection. They can only utilize limited features that are manually extracted for preforming classification. This greatly constricts their power for advanced analysis and prediction. Neural Network architectures especially convolutional neural networks (CNN) overcome the problems faced by traditional machine learning models. They have more representational power and can automatically extract complex features with minimal preprocessing. Recent advance in CNN architectures is the development of Vision Transformers (ViT). The ViT, an image-specific adaptation of the transformer, was presented in and has shown to outperform convolutional neural network (CNN) models in the enormous data domain by improving on the JFT dataset's 300 million image training set. ViTs are a type of neural network that can learn long-range dependencies in images, which is important for tasks like brain tumor classification. The suggested ViT models may acquire more suitable inductive biases that are requirement-specific, and their reduced inductive biases are a result of their global patch-based learning. In medical imaging diagnostics, the adoption of ViT models is still in its infancy because of the newness of ViTs and the intensive training needed to utilize massive amounts of data and advanced computer resources. Within the computational framework of the ViT architecture, model training, hyper parameter tuning, and testing may all take place. The procedures for building the ViT models and gauging their efficacy are also provided in [179]. This work utilizes the power of ViT model for classification of brain tumors.

The proposed work mainly focuses on two issues faced in developing automatic brain tumor classification. First issue for creating accurate brain tumor models is lack of training data. Second issue is creating a model that utilizes all the available features to accurately predict tumors without false positives. To overcome the issue of lack of data this work proposes use

91

of data augmentation along with a more advanced methodology, Generative Adversarial Networks (GAN). By incorporating GAN-based data augmentation procedures and a structural similarity loss function, the proposed approach addresses issues such as visual blurriness and model collapse, resulting in improved efficiency. The issue of incorporating more features is overcome by proposing an auxiliary neural network attached (ViT) model through which additional extracted features are presented into the network for classification. Furthermore, the study compares the effectiveness of the ViT network model with other methodologies, including data augmentation (DA). The results demonstrate that the proposed GAN-based approach produces superior results compared to DA. Overall, this work contributes to advancing the field of image synthesis by introducing an innovative method that improves upon existing techniques [227].

4.2 Dataset Employed

For performing experiments and comparison, a publically available dataset is considered. The dataset was taken from Figshare repository and it contains 3064 MRI image slices. These slices were taken from around 233 people. These slices were labelled among classes no-tumor, meningioma, glioma, or pituitary tumor. Samples from the dataset is given in Figure 4.1. The MRI scans were captured from different positions for each patient. The MRI slices were processed in to grayscale images across three channels and the total number of images are given in Table 4.1. As the amount of images in the dataset are limited. They are not sufficient to train a model that generalizes well on unseen data. To overcome this, additional images across each class are generated using multiple image generation techniques and the resulted dataset images are shown in Table 4.2. These techniques are detailed in the following sections. Some clinical information regarding the class labels specifying different cancers is given below [228].
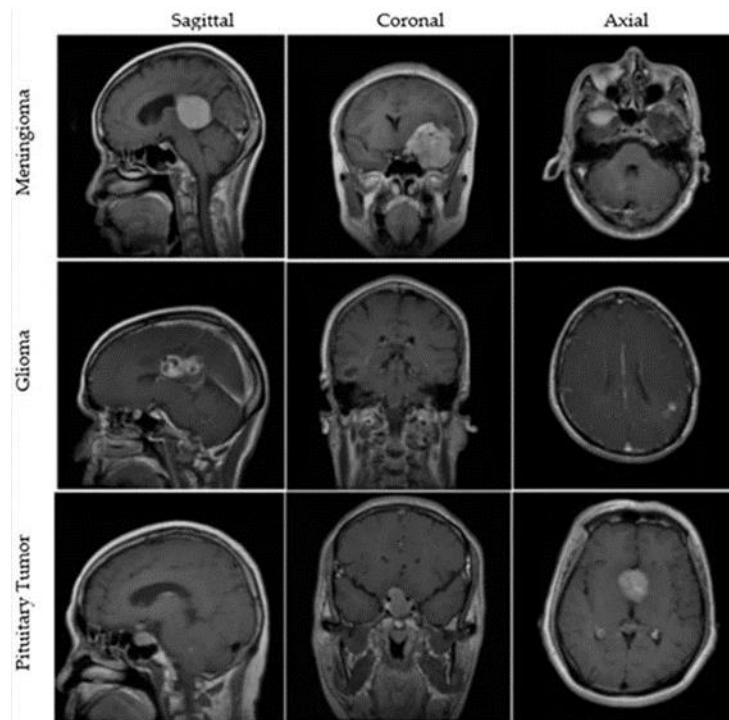
Figure 4.1 Sample Brain MRI image

- Meningioma: this is a type of non-cancerous brain tumor that originates from the arachnoid cap cells. It predominantly affects women and individuals aged 60 and above. Brain tumors account for approximately 13 to 26 percent of all tumors found within the skull.

- Glioma: the most frequently observed primary brain tumors are malignant gliomas, which constitute the majority of malignant brain tumors, accounting for approximately 81% of all instances.

- Pituitary Tumor: pituitary tumors, which arise within the pituitary gland, are typically benign and pose minimal threat. However, due to the gland's crucial role in regulating numerous hormones, tumors in this area can have wide-ranging impacts. They account for approximately 10% to 15% of all cases of brain cancer.

Table 4.1: Brain tumour, MRI, pictures.

BT Type	Total image	Training	Validation	Testing
Meningioma	708	502	75	131
Glioma	1426	988	148	290
Pituitary Tumour	930	647	91	192
Total(N)	3064	2137	314	613

Table 4.2: Additional images generated using data augmentation and GAN.

BT Type	Original	Data Augmentation	GAN	Total
Meningioma	502	1500	200	2202
Glioma	988	2600	250	3838
Pituitary Tumour	647	1900	230	2777

4.3 Techniques used

These are the following techniques employed in our existing work:

4.3.1 GAN

Generative Adversarial Networks (GANs), also referred to as GANs [259], are a type of data generation toll designed to produce new data samples. GANs possess a robust methodology for data generation, which, when applied appropriately, can lead to resilient models. By utilizing random noise extracted from a latent space, GANs generate unique images that accurately capture the feature distribution of the training dataset. Notably, a GAN consists of two separate networks rather than just one namely discriminator network and the generator network.

In GAN, the "Discriminator" plays the role of distinguishing between genuine and counterfeit samples, accurately labeling them as "real" or "fake". It receives actual sample

batches from the source dataset, while the Generator supplies the counterfeit samples. On the other hand, the "Generator" lakes random noise from a latent space as input and generates fabricated data, which the discriminator assesses as "fake". The primary objective of the Generator is to generate visuals that deceive the discriminator into perceiving them as authentic. Through the utilization of backpropagation, the Generator progressively acquires the ability to produce samples that closely mimic the physical and mathematical distribution of the original dataset. This is achieved by iteratively adjusting the weights and biases of these models.

The GAN algorithm is formulated as a minimax game between the generator G and the discriminator D. The minimax game can be mathematically formulated in equation (4.1) as follows:

$$\min_{G} \max_{D} V(G, D) = E_{x \sim P_{data}}[log D(x)] + E_{z \sim p(z)}\left[log\left(1 - D\big(G(z)\big)\right)\right] \tag{4.1}$$

Where input noise z is drawn from a previous distribution p(z), and G(z) is the sample of data created by the generator network. D(x) is the discriminator network's estimate of x's data sample probability. Training data empirical distribution is data(x). The logarithmic loss function estimates the discriminator's output error for a real data sample x. The logarithmic loss function estimates the discriminator's output error using a generated data sample G(z).

➤ *GAN loss function*

The standard GAN loss function could be calculated using the equation (4.2) is given below:

$$\min_{G} \max_{D} V(G, D) = E_{x}[log D(x)] + E_{z}\left[log\left(1 - D\big(G(z)\big)\right)\right] \tag{4.2}$$

The default loss function in a GAN is known as the min-max loss. In this setup, the discriminator aims to maximize the value of this loss function, while the Generator strives to minimize it to the greatest extent possible. From a min-max perspective, this loss formulation proves advantageous. However, there is a limit for the Generator when it fails to keep up with

the discriminator, resulting in frequent training cessation. The standard GAN's loss function can be divided into two parts: Generator loss and Discriminator loss.

➢ *Discriminator loss*

During the training process, the discriminator is responsible for classifying both actual and fake data produced by the Generator. It optimizes a function that penalizes itself when it incorrectly labels a fake instance as real and a real instance as fake is done by maximizing the function given as follows in equation (4.3).

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[logD(x^{(i)}) + \log\left(1 - D\left(G(z^{(i)})\right)\right) \right] \tag{4.3}$$

log(D(x)) = probability generator classifies real image, increasing log(1-D(G(z))) leads to the labeling of the fake image correctly.

➢ *Generator loss*

During the training process, the Generator takes input from an unexpected source and creates an output based on that source. Once the output reaches the discriminator, it is sorted into "Real" or "Fake" categories based on the discriminator's accuracy. Next, the discriminator's categorization determines the loss; the Generator receives a reward if it can trick the discriminator but a penalty otherwise.

This equation (4.4) is minimized to train the Generator is given as follows:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G(z^{(i)})\right)\right) \tag{4.4}$$

4.3.2 Gabor filter

Gabor filters are widely utilized linear filters in image processing that are specifically designed for edge detection. These filters have shown effectiveness in texture analysis and discrimination due to their frequencies and orientation representations, which closely resemble those employed by the human visual system. Two-dimensional Gabor filters possess a Gaussian kernel function that undergoes sinusoidal modulation within the spatial domain.

These filters can be created from a single mother wavelet through stretching and rotation, exhibiting a self-similar nature. The impulse response of a linear Gabor filter is determined by amplifying a harmonic function with a Gaussian function. When performing the Fourier transform of the impulses response, the harmonic function and Gaussian function are convolved together.

When applying the Gabor feature extraction method to process an image, the corresponding Gabor Features are obtained. By employing customized Gabor filters based on statistical data related to character structures, features can be directly extracted from grayscale character images. To enhance performance when dealing with low-quality images, the outputs of Gabor filters are adaptively subjected to sigmoid transformation. Histogram features are then constructed by independently utilizing the positive and negative real sections of the Gabor filter outputs. This approach aims to improve the discriminative capacity of the extracted features. Here is the mathematical equation (4.5) for a Gabor filter:

$$g(x,y) = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \tag{4.5}$$

Where, $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$

x and y are the filter's spatial coordinates; γ is the filter's aspect ratio; σ is the Gaussian envelope's standard deviation. The wavelength of the sinusoid component is denoted by λ. The sinusoidal component's phase shift is denoted by ψ.

4.3.3 Vision Transformers Networks

Although Transformers have been used in vision tasks before, their popularity in computer vision can be traced back to when the authors. effectively adapted the encoder of [220] for the visual classification problem. The underlying processes are compressed into network layers and returned to them later to understand this adaptation. Figure 4.2 shows the

basic layout of the "Vision Transformers" (ViT). The ViT model implemented mainly global attention, but instead of focusing on the full picture, it split the image into 16x16 pixel patches.
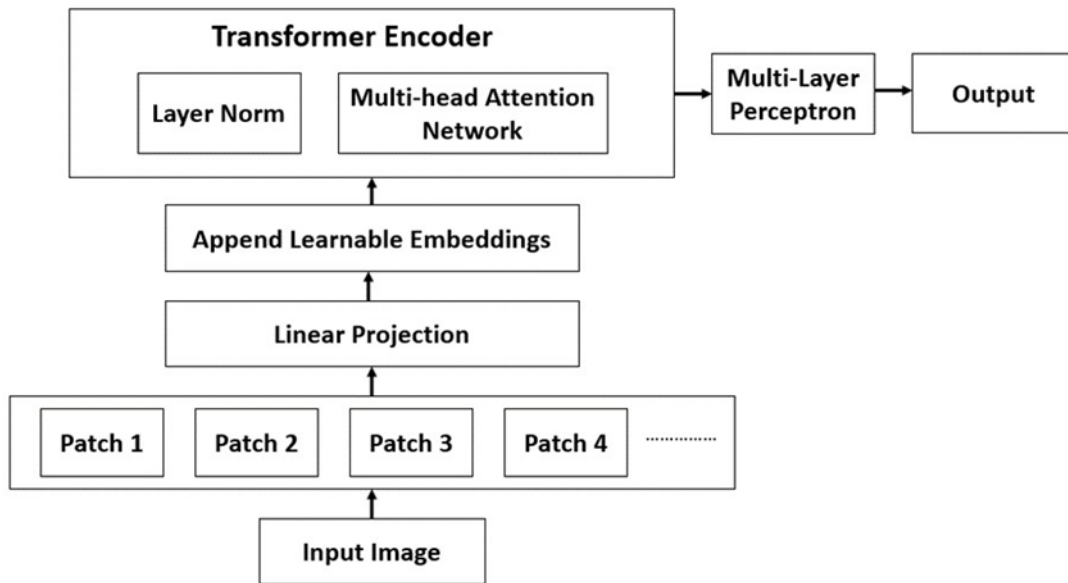
Figure 4.2 Structural representation of Vision Transformers

ViTs are network models that have achieved better results than the CNN models. ViTs are based on Google's transformers networks, which were used in state of the art language models. The ViTs architecture as follows. The first step in a vision transformer network is to spilt the input image into patches. The ViT network cannot process full input image as a whole. Generally, image is divided into 16x16 patches. Mathematical equations employed are given in equation (4.6), (4.7) and (4.8)

$$\text{Original Image: } x \in R^{H \times W \times C} \tag{4.6}$$

H, W-resolutions of the original image; C- number of channels.

$$\text{2D Converted Patches: } x_p \in R^{N \times (p^2.C)} \tag{4.7}$$

(p, c) is the resolution of each image patch.

$$N = {HW}/{P^2} \tag{4.8}$$

N-serves as the effective input sequence for the transformer.

After the input is divided into patches, linear projection is performed in those patches. Since the patches are in array format, linear projection converts those arrays into vector representation. To these vector representations, learnable embedding's are added by producing linear embedding's and positional embedding's. equation employed are given in (4.9) and (4.10).

$$Z_0 = [X_{class}; X_{p^1}E; X_{P^2}E; \dots X_{p^N}E] + E_{pos} \qquad (4.9)$$

$$E \in R^{(P^2.C) \times D}, E_{pos} \in R^{(N+1) \times D} \qquad (4.10)$$

This embedding's specify the extra class embedding's. The vectors with added embedding's are given as input to the standard encoder of a transformer. This encoder contains three elements: Layer norm, Multi-Head Attention Network (MSP) and Multi-Layer Perceptron (MLP).

The Layer normalization element helps the transformer model adapt to the subtle differences and variations in the input images by employing the following equations (4.11) and (4.12).

$$Z_{l'} = MSA(LN(Z_{l-1})) + Z_{l-1}, \quad l = 1 \dots L \qquad (4.11)$$

$$Z_l = MLP(LN(Z_{l'})) + Z_{l'}, \quad l = 1 \dots L \qquad (4.12)$$

The MSP module generates self-attention maps for the patches. Attention maps are crucial component of a transformer and they help the network to focus on the most important features of the input required for classification. MLP is the final module in the transformer. It produces the required output. It is a two-layer network that contains GELU (Gaussian Error Linear Unit) at the end of the network.

4.4 Proposed methodology

This section outlines the proposed methodology. The MRI images are pre-processed to make them suitable for further processing. Pre-processing includes converting the pixel values of images to grayscale range of 0 to 255 and resizing the images to constant height and width.

First images are processed using different data augmentation techniques such as random rotate, random flip, padding and color augmentation. This is done to generate more images from original images. The augmentation generated images along with original non-augmented images are used to train Generative Adversarial Network (GAN). For each of the image class GAN model is trained separately to ensure it is able to generate images within that class. On total, GAN network is trained 4 times each with images belonging to classes no tumor, glioma, meningioma and pituitary tumor. Trained GAN is used to generate additional images of each class. Since each class has separate GAN network, it is easy to generate images with known labels.

In the other path, all the images including the original images, images generated using data augmentation and images generated using GAN are passed through different image processing techniques to extract different features that aid in classification. The images are passed through Gaussian filter and Canny edge detection algorithm. First processing technique applied on the images is the Gabor filter. Gabor filter bank is applied on the images to obtain set of filtered images. On each filtered image, magnitude and phase of the complex response is computed. From the magnitude and phase, mean and standard deviation over the edge map created from Canny edge detector. These mean and standard deviation are stored as spatial features.

Second processing technique is to compute the area, centroid of the edge maps from canny edge detector. Computed values are stored as geometric features. Finally, third processing technique includes computing the mean and standard deviation of the intensities of the pixels of the edge map along with histogram of oriented gradients (HOG). These features are stored as appearance features. All three types of extracted features are stored as feature vector.

Images generated through data augmentation and images generated through GAN along with original pre-processed images are given as input to the (ViT) model. The ViT model has an auxiliary Artificial Neural Network (ANN) that takes the above extracted feature vector of each image as input. The features extracted by ViT are concatenated with the features extracted by auxiliary network as follows in equation (4.13).

Here the features extracted by ViT are in the size of 1xp and features extracted by auxiliary neural network are of size 1xq.

So, the resulting concatenated features will be of size:

$$1 * (p + q) \tag{4.13}$$

Finally, the concatenated features are passed through artificial neural network layers for final classification. The performance of the network is evaluated using different evaluation metrics such as precision, recall and f1-score. Figure 4.3 illustrates the flowchart of the proposed methodology. Algorithm 1 shows the steps involved in the proposed methodology.
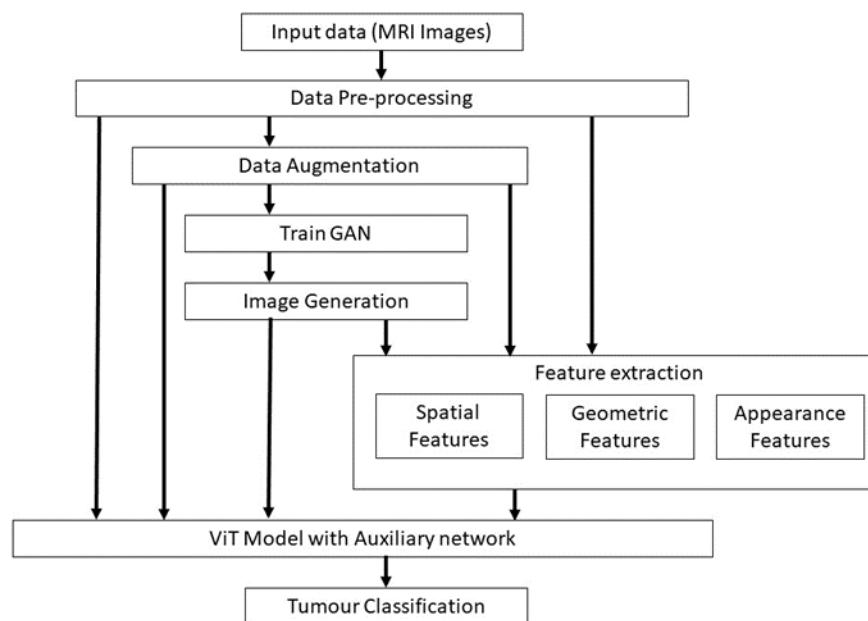
Figure 4.3 Proposed methodology

Algorithm 1: Working of Proposed Model

Input: an image I

Output: Class label

Start

For each image I

        Convert the image I into grayscale range

        Resize the image I

End for

For each image I

        Randomly apply data augmentation techniques such as random rotate, random flip, padding and color augmentation

End for

For images in each class c_i do

Define the GAN architecture and training parameters

        Define the generator network G and the discriminator network D

        Define the loss function for training the GAN (e.g. binary cross-entropy)

        Define the optimization algorithm (e.g. Adam)

        Set the number of epochs for training the GAN

        Set the batch size for training the GAN

Train the GAN on the original dataset D along with augmentation generated images

        Train the GAN for the specified number of epochs using batches of images

        Save the trained generator G

Augment the dataset D by generating new images using the trained generators G_i

For each image I in dataset D

        Apply a Gaussian filter to smooth the image

        Apply an edge detector (e.g. Canny) to obtain a binary edge map E

        Apply Gabor Filter bank to the image to obtain a set of filtered images {G1, G2, ..., G_n}

        for each filtered image G_i do

                Compute the magnitude M and phase P of the complex response

                Compute the mean and standard deviation of M and P over the edge map E

                Store the mean and standard deviation as spatial features

        End for

        Compute Geometric Features:

                Compute the area and centroid of the binary edge map E

                Compute the orientation and aspect ratio of the bounding box of E

                Store these features

        Compute Appearance Features:

                Compute the mean and standard deviation of the intensities of the pixels inside E

                Compute the histogram of oriented gradients (HOG) of the image inside the E

Store these features

End for

Concatenate all the computed features into a feature vector F

Disease Classification using ViT model with auxiliary network from Step X → X.

Input:

A dataset D of images to classify and Feature Vectors F

A set of labels L indicating the type of disease for each image in D

A pre-trained ViT model M for image classification

An artificial neural network N

A batch size B for processing images in batches

Several epochs E for training

Output: A trained model MN' for disease classification

Define hyper parameters

        Define the learning rate alpha (e.g. 0.001)

        Define the loss function (e.g. cross-entropy)

        Define the optimizer (e.g. Adam)

Split data into training and validation sets

Split D, F and L into training and validation sets with a specified ratio (e.g. 80:20)

Prepare data for training

Create a data loader for the training set with batch size B

Create a data loader for the validation set with batch size B

Train the ViT model along with ANN model

for each epoch in E do

a.   Set the models to training mode

b.   For each batch in the training data loader:

→   Forward pass the batch through the models

→   Concatenate the features of both models

→   Pass the concatenated features through Dense layers

→   Compute the loss between the predicted labels and the true labels

→   Compute the gradients of the loss concerning the model parameters

→   Update the model parameters using the optimizer and gradients

c.   Set the models to evaluation mode

d.   For each batch in the validation data loader:

→   Forward pass the batch through the models

→   Compute the loss between the predicted labels and the true labels

e.   Compute the validation accuracy using the predicted and true labels

f.   Print the epoch number and validation accuracy

**Output the trained model**

End

4.5 Experimental Setup and Results Analysis

The following section provides a detailed explanation of the experimental setup and an in-depth analysis of the results.

4.5.1 Experimental Setup

Every experimental test is conducted concurrently on two NVIDIA GPUs, specifically A5000s with 24 GB of RAM each. RAM for the system is 128 GB. Numerous data analysis frameworks, such as Pandas, Numpy, Seaborn, 496 Matplotlib, and Scikit-learn, were employed in the study. The entire framework is executed for ten epochs. 20% of the data were used for testing, while the remaining 80% were used for training. The proposed model was trained on the training data. The suggested model was fine-tuned utilizing a variety of parameters, including hyper parameters, to improve classification accuracy and prevent overfitting. Different learning rates were applied, and it was found that the default learning rate of 0.001 gave better results.

The proposed approach utilizes both the features extracted by the ViT model and the features extracted through different image processing techniques to predict the tumor type. One of the major issues with creating tumor prediction algorithms is lack of data. This is certainly true for the original dataset considered. The dataset does not have enough samples for the model to learn and generalize well for unseen data. To overcome this limitation, two image generation techniques are employed. One is randomly applying data augmentation to the images to create more images and second is to train a GAN model with the images so that the model can generate images from noise with same distribution. The GAN model is trained separately for images that belong to same class label. This is done to distinguish the generated images per class. The GAN network had optimize well for each class. The generator and discriminator had good balance between them and both the losses were optimized efficiently. The loss per epochs for no tumor class image training is given in Figure 4.4.
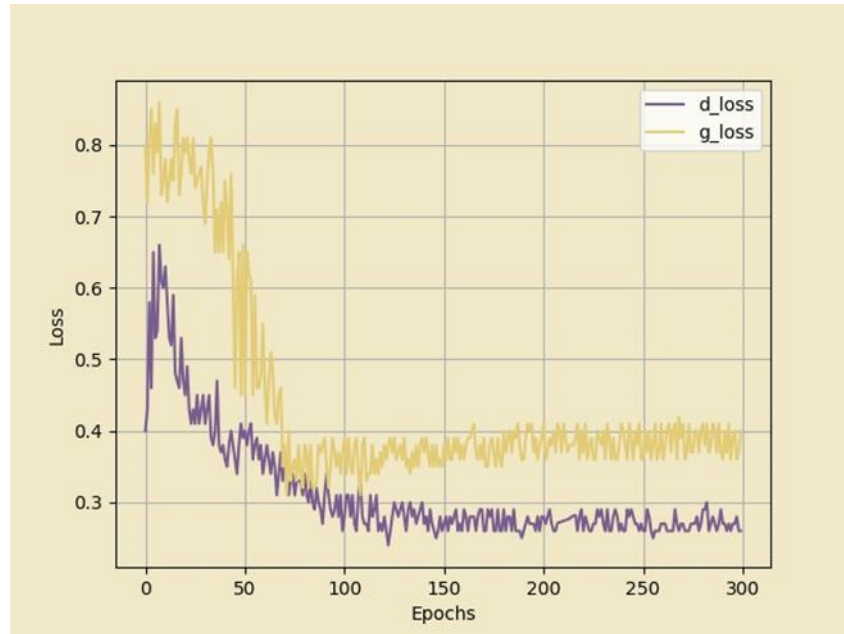
Figure 4.4 GAN loss per epochs for no tumour class

4.5.2 Result Analysis

Generating images through augmentation and GAN has certainly helped the proposed network in generalizing well on unseen data. Proposed model trained with generated images has achieved higher validation accuracy and test accuracy when compared to same model trained without the generated images. Model trained without the generated images was overfitting after initial epochs. To make sure that the model does not over fit, dropout was added to some of the final layers and training was stopped early before the model started overfitting. This issue was not faced when the proposed model was trained with the generate images. Training, validation and test accuracy of the proposed model when trained with and without the generated data is given in Figure 4.5. From Figure 4.5, it is evident that the data generation techniques helped the network avoid overfitting and perform effectively on the unseen data. The proposed data generation techniques can be used in applications where data available is limited especially in biomedical applications.
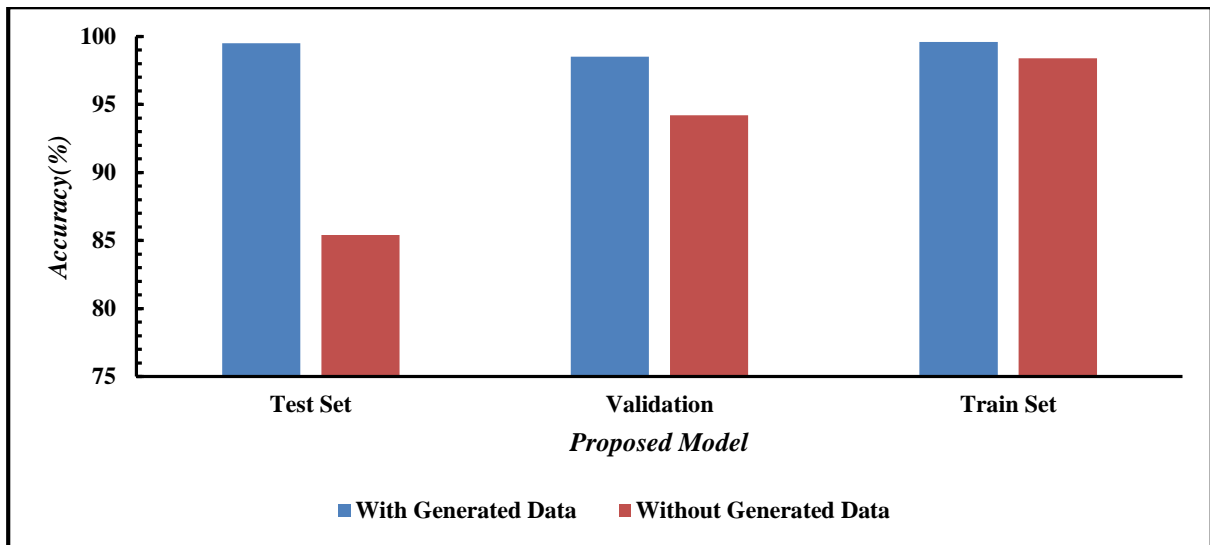
Figure 4.5 Accuracy of proposed model when trained with and with generated data

The proposed model has an auxiliary network attach to the main ViT network. Most of the DL model have an issue called vanishing gradient problem. Due to the complexity of the model some of the signal may not pass throughout the network. This causes some of the features to vanish and fail to propagate to the network. Another issue is that the model may fail to recognize some of the features important for classification. The proposed auxiliary network helps overcome these problems. Features extracted through different image processing techniques are given to the auxiliary network and the image that these features are extracted from is given to the ViT model. Both the network features are concatenated to produce the class label. The auxiliary network helps introduce some of the vanished features back to the network, which are useful for correct classification. It can also help force the network to focus on features that are important in producing correct results. Experiments performed on the model when compared to normal ViT model and other state-of-the-art models have shown the effectiveness of the proposed ViT model with auxiliary network.
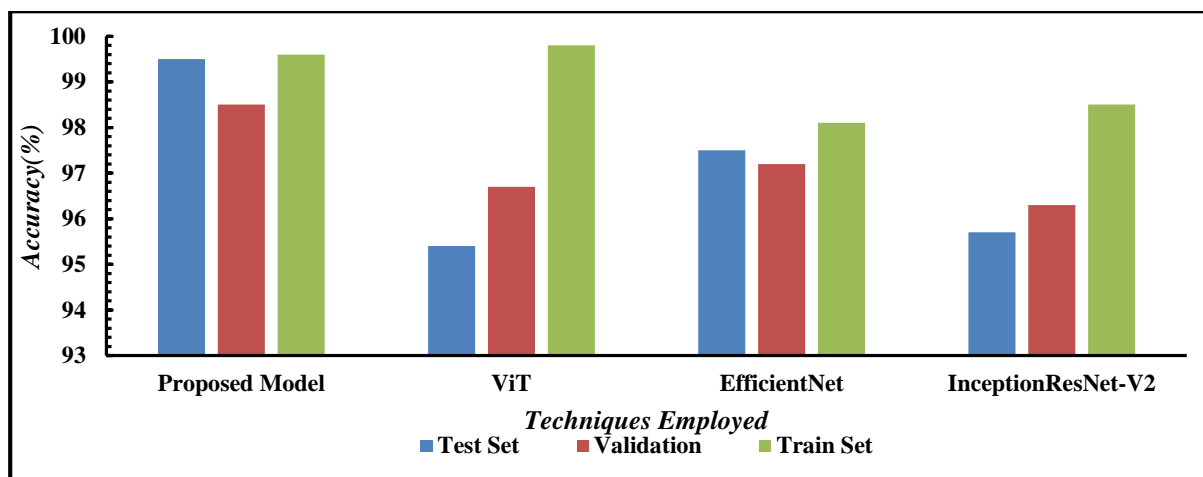
Figure 4.6 Accuracy of proposed model when compared to other state-of-art models

The accuracy of the proposed model, in comparison to state-of-the-art techniques, is illustrated in Figure 4.6. When compared with other state-of-the-art methods, the proposed network has achieved overall good results. The auxiliary network has helped balance the distinction between the classes which is evident from higher precision and recall. It can be seen from Table 4.3 that proposed model had consistent results with less false positive and false negatives across all classes. The proposed model has achieved the highest F1-Score across all classes, demonstrating a strong balance between precision and recall. The proposed model was able to correctly predict across the classes where other model has performed well in some classes and struggled in other classes.

Table 4.3. Evaluation Metrics of proposed vs State-of-art models

Metric	Class	Proposed	ViT	EfficientNet	InceptionResnet-v2
Precision(%)	*No Tumour*	**98.3**	98.4	99.1	98.5
	Pituitary	**98.4**	98.6	97.6	98.4
	Meningioma	**100**	96.8	98.5	95.1
	Glioma	**99.6**	100	96.7	98.1
Recall(%)	*No Tumour*	**99.1**	98.2	99.6	96.4
	Pituitary	**100**	95.8	96.8	97.2
	Meningioma	**98.4**	98.7	97.4	99.1
	Glioma	**98.6**	97.5	97.3	97.3
F1-Score(%)	*No Tumour*	**98.7**	97.4	99.7	96.9
	Pituitary	**99.2**	97.1	97.3	97.4

	Meningioma	99.2	97.2	97.9	96.95
	Glioma	99.1	98.7	98.32	97.4

Confusion matrix for the predictions on test data of the proposed model is given in Figure 4.7. It shows that the model had only 8 false classifications. Out of those false classifications, glioma class had the highest with 4 false classifications. The results obtained from the proposed model demonstrate its effectiveness in achieving the highest accuracy of 99.5% among all the compared models on unseen test data. This indicates that the proposed model has generalized well to new data and successfully distinguished between different classes by incorporating additional features into the network. In applications such as brain tumor detection, where minimizing false positives is critical, the proposed model excels by achieving zero false positives.
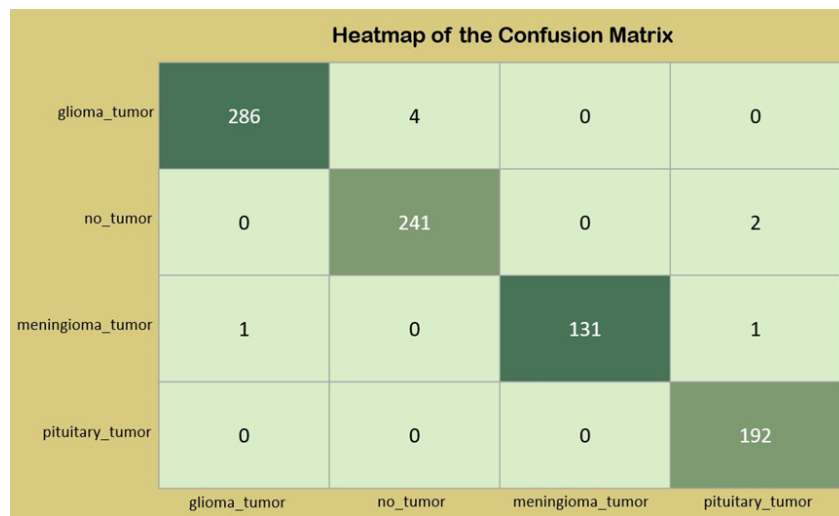
Figure 4.7 Confusion matrix on test data for the proposed model

4.6 Chapter summary

This chapter focuses on developing a hybrid approach for accurately classifying brain cancers from imbalanced data. The proposed hybrid method integrates data processing, augmentation techniques, and an enhanced network model to improve classification performance. When training data is limited, the proposed approach can generate additional data

samples that maintain the original data distribution, supporting more effective model training. The improved network model incorporates custom auxiliary features alongside those extracted by Vision Transformers (ViTs), enabling the classification network to leverage a wider range of relevant features necessary for accurate predictions. Experimental results have highlighted the efficacy of this hybrid approach, demonstrating that it outperforms state-of-the-art classification models. The proposed model, with its custom auxiliary network, achieved significantly lower false positives and false negatives. Overall, the model's performance metrics consistently surpassed those of other compared models, underscoring its potential for reliable and precise brain cancer classification in medical applications.

Chapter Five: A LINEAR TIME SHRINKING SL(t)-ViT APPROACH FOR HANDLING CLASS IMBALANCE

This chapter provides the experimental analysis of leveraging a Structural Similarity *Loss function, GANs generate high-quality annotated images, enriching the training dataset and improving model robustness.*

5.1 Introduction

The current limitation faced by DL algorithms stems from an inadequate availability of magnetic resonance (MRI) imaging datasets, impeding their ability to achieve superior performance. This predicament arises due to the intrinsic resemblance between the distribution of segmented images generated by classical data augmentation techniques and the original images. Consequently, the model's capacity to generalize is hampered. Another issue is that recent DL advancements require large computational resources to complete the given task [229].

To surmount the challenge of inadequate availability of MRI images, a novel approach is proposed, which involves employing Generative Adversarial Networks (GANs) for data augmentation. The primary objective of this approach is to mitigate problems associated with inadequate data, poor quality of data, picture blurriness, and model collapse. Moreover, to enhance the efficiency of the suggested model, in addition structural similarity loss function is integrated, which is used to calculate the similarity score of the new images vs the original image [230]. Furthermore, the study incorporated the utilization of the SL(t)-ViT model to classify diseases within the data. The key concept of SL(t)-ViT is to improve the computation efficiency of the model by utilizing linear time attention layers, which capture multi-scale features including channel-wise, spatial, local, and global features [228]. The attention mechanisms within each Transformer layer enable the model to capture both global and local dependencies between patches, allowing it to attend to different parts of the input sequence.

The proposed linear attention mechanisms within each transformer layer enable the model to selectively focus on relevant patches while aggregating information from nearby and distant patches, thus enhancing its overall performance. The attention mechanism achieves these computations in a linear time thus enhancing the overall efficiency of the model. The literature shows that the SL(t)-ViT model showed improved usage of computational efficiency compared to traditional ViT and CNN models. In addition to the linear attention layers in the model, to further improve the model efficiency in terms of performance metrics and computational efficiency, a shrinking nature is added to the model, making it more computationally efficient. The resulting model is thoroughly evaluated, assessing its effectiveness in identifying and categorizing brain tumors [226].

5.1.1 Key contributions

Following are the key contributions of this chapter:

(i) Development of Novel SL(t)-ViT architecture: In this work a novel concept of the Shrinking Linear Time Vision Transformer (SL(t)-ViT) is proposed, marking the first occurrence of a transformer backbone designed exclusively for a range of pixel-level dense classification tasks, aimed at enhancing computational efficiency.

(ii) Architectural Versatility: The suggested SL(t)-ViT structure achieved notable adaptability, generalization, and efficient performance in various scenarios, including both, multi-classification and binary classification.

(iii) Resource-Efficient Innovation: We developed advanced reduction methods and incorporated multi-scale Nystrom attention mechanisms into the SL(t)-ViT model, strategically mitigating resource usage. This improvement empowers the model to accurately capture multi-scale and high-resolution features while utilizing fewer resources within a linear timeframe.

(iv) Superior Performance: A thorough comparative examination with DL models documented in the current literature is performed. The proposed SL(t)-ViT model exhibited

superior accuracy in both multi-classification and binary classification scenarios. Particularly noteworthy were the exceptional accuracy rates of 0.993 for multi-classification and 0.997 for binary classification, surpassing established benchmarks.

5.2 Dataset

This section describes the techniques used and the proposed methodology. Techniques employed are GAN, Gabor filters, SL(t)-ViT. The GAN is used for data augmentation, while the Gabor filter selects the salient edge features and SL(t)-ViT for image classification. The following datasets were used for experimentation:

(i) Figshare [231]*:* There are 233 people with meningioma's, gliomas, or pituitary tumors included in a publicly available dataset on Figshare, including 3064 T1w CE MRI slices. Additionally, some brief clinical information about the aforementioned triad of malignancies is also provided below.

(ii) Br35h [131]*:* The BR35H dataset is designed for binary classification and comprises 3,000 MRI images of the brain. Among these, 1,500 images depict brain tumors, while the remaining 1,500 represent normal brain MRI scans. Each image is available in JPEG format and includes both T1 weighted and T2-weighted contrasted MRI scans, offering a comprehensive dataset for analysis. The images vary in grayscale sizes, providing a diverse set for training and evaluation purposes.

(iii) Brain tumor MRI KAGGLE [232]*:* This dataset was obtained by combining three datasets: Figshare, SARTAJ, and Br35H. There are four classes in total in the dataset: brain MRI images from individuals with glioma, meningioma, pituitary tumors, and healthy individuals. There are 1,623 images for glioma, 1,627 images for meningioma, 1,769 images for pituitary tumors, and 2,002 images for healthy individuals. A total of 7,021 MRI images were used. The dataset is open-sourced on the Kaggle platform. Each file is a 512×512 JPEG

with a label indicating the type of brain tumor. This dataset was used as input data for each model.

(iv) Brain MRI Images [233]*:* The dataset used in this study consists of brain MRI scans, comprising approximately 256 raw images with various dimensions measured in pixels. These images were sourced from a Kaggle dataset and are stored in Joint Photographic Experts Group (JPEG) format. The dataset is divided into two categories: "Yes," indicating the presence of a tumor, and "No," indicating the absence of a tumor. Among these, there are about 158 images depicting benign tumors and 98 images depicting malignant tumors.

BT-small 2C, BT- large 2C, BT-large 4c[234]: BT-small-2c consists of 253 brain MRI images, with 155 images depicting tumors and the remaining 98 images depicting healthy brains. The BT-large-2c dataset encompasses a larger collection of 3,000 images, with 1,500 images showcasing brain tumors and the remaining 1,500 images depicting healthy brain scans without any tumors. The BT-large-4c dataset comprises 3,064 T1-weighted images, each depicting one of three distinct types of brain tumors: gliomas, meningioma's, and pituitary tumors.

5.3 Proposed methodology

This section proposes an approach for processing medical images using GAN. The model's input is an image dataset, which is further pre-processed for feature extraction. The flowchart of the proposed model is shown in Figure 5.2.

5.3.1 Input Image

The first step involves loading the dataset into memory or accessing it from storage. In dataset D, containing Z examples, the goal is to create a learning algorithm that generates a classifier output denoted as 'T.' This classifier, represented by a hypothesis function $f(X_i) = Y_i$, aims to predict new values of Y_i for any input X_i provided.
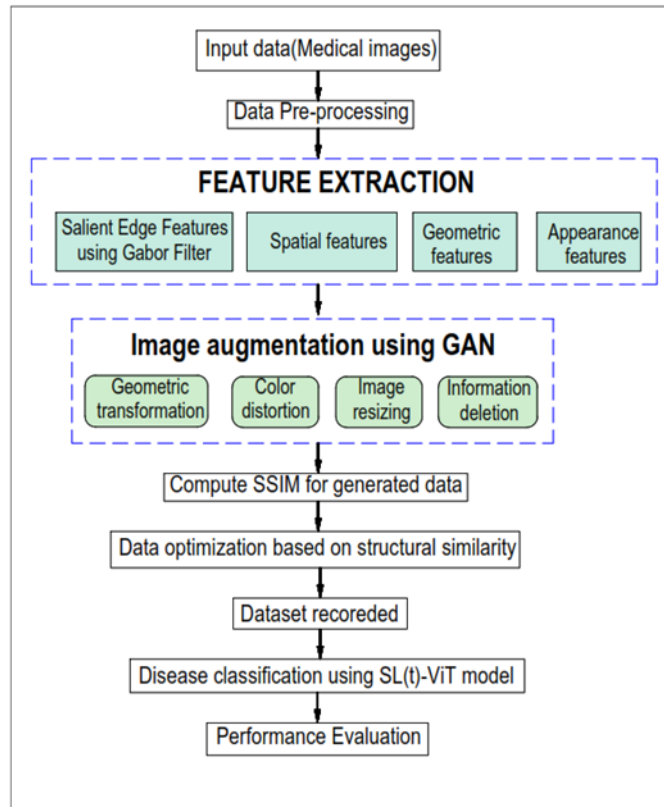
Figure 5.1 The flowchart of the proposed method.

5.3.2 Data Pre-processing and Pre-processing

Once the dataset is loaded, it undergoes pre-processing to ensure it is in a suitable format for the model. This involves converting images to grayscale, parsing the data to extract relevant features, handling missing values, scaling or normalizing the features, and encoding categorical variables if necessary. Among various pre-processing techniques available salient edge features extraction using Gabor filters, Spatial features, Geometric feature extraction technique and Appearance feature extraction technique are the techniques employed.

Grayscale conversion simplifies the images by reducing them to a single channel representing the intensity of light, as opposed to the three channels (red, green, blue) in color images. This simplification reduces computational complexity while retaining essential information relevant to tumor detection. Next, in this process, a 3x3 Gaussian filter is first applied to images for noise reduction and feature enhancement. This is followed by Canny

114

Edge Detection with a threshold of 100 and a 3:1 ratio, ensuring that only strong edges are detected. Gabor filters, known for capturing edges and textures, are then utilized for feature extraction, employing a 5x5 kernel size and 8 orientations to analyse complex spatial patterns in brain scans. Additionally, Histogram of Oriented Gradients (HOG) is used to extract features by computing histograms of gradient orientations within 8x8 pixel cells and normalizing them over 2x2 cell blocks. This combined approach of Gabor filters and HOG enhances the detection of tumour-related features by capturing detailed edge orientations, textures, and spatial arrangements in the images.

5.3.3 Data Augmentation

The main objective of a GAN is to generate realistic synthetic data through the collaboration of its generator and discriminator components. Although GANs are commonly used for data generation, their performance can be hindered by limited generalizability. To overcome this issue, the proposed model leverages GANs for image augmentation, as shown in Figure 5.2, which includes geometric transformations, colour distortions, image resizing, and information deletion. The GAN is trained to produce images with specific geometric properties for geometric transformations. For colour distortions, it generates images with varied colour tones and contrasts. Image resizing and information deletion involve training the GAN on smaller images, resulting in images with certain regions masked or pixels deleted. To ensure label accuracy, separate GAN models are trained for each class. The discriminator then assesses the generator's output, calculating the loss for both real and synthetic images, and assigns a probability score to indicate the likelihood that the images belong to the real data distribution.

5.3.4 SSIM calculation

From Figure 5.2, the subsequent step following image augmentation involves computing the structural similarity index (SSIM) score for the generated data. The SSIM metric is an evaluation measure for assessing the quality of the newly generated images. In the SSIM

phase, a comparison is made between the structural information, luminance, contrast, and structure of an original and generated images using the following equation (5.1).

$$SSIM(I, R) = \frac{(2\mu_I\mu_R + K_1)(2\sigma_{IR} + K2)}{(\mu_I^2 + \mu_R^2 + K1)(\sigma_I^2 + \sigma_R^2 + K2)}$$ (5.1)

SSIM values range from -1 to 1, with values closer to 1 indicating a higher similarity score. Until the SSIM score reaches near to 1 the above process repeats. Once the SSIM module output score near to 1 then next phase starts execution. To compute the SSIM for each image (I) in the dataset (D), and the reference image (R) with a specified window size, and window size employed is 11*11, SSIM constants K1 and K2 are used to stabilize the division K1 and K2 values are 0.01and 0.03 respectively. Likewise, L is the dynamic range of pixel values in the image and SSIM constant L value employed is 255 because as image is in the form of 8-bit grayscale, each pixel is represented by 8 bits, which allows for 256 different intensity levels.

5.3.5 Data optimization based on SSIM score

Following the SSIM score calculation, the next phase involves optimizing the data based on the obtained SSIM score. For optimizing the dataset is validated by applying it to a model and evaluating the model's performance. From the achieved score adjust the SSIM threshold value T generate additional examples through transformations such as rotation, scaling, cropping, or noise addition. Diversifying the dataset helps the model learn robustness to different image content and distortion types, potentially improving SSIM scores on unseen data by avoiding overfitting.

5.3.6 Dataset recorded

Once the dataset has been optimized and validated, it is recorded for use in training models.

5.3.7 Classification using SL(t)-ViT model and performance evaluation

Once the data is optimized, the final dataset is recorded and provided to the SL(t)-ViT model for further classification.

Although Transformers have been used in image processing tasks before, their popularity in computer vision can be traced back to when encoder was effectively adapted [39] for the visual classification problem. Instead of opting for the conventional ViT, a Shrinking Linear Time Vision Transformer(SL(t)-ViT) was designed due to its capability to generate multi-scale feature maps suitable for dense classification tasks by effectively using the resources. The schematic overview of our proposed SL(t)-ViT model is depicted in Figure 5.1
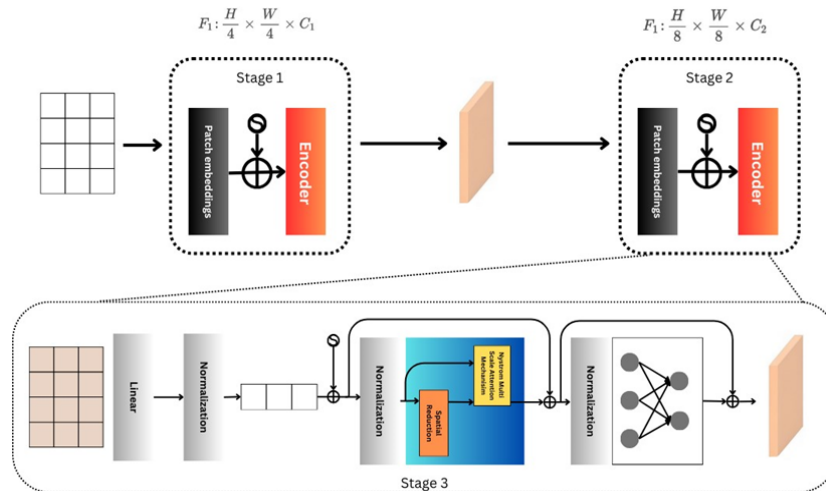
Figure 5.2 Structural representation of SL(t)-ViT.

The model architecture comprises three stages, each responsible for producing feature maps of distinct scales. A uniform architecture, encompassing a patch embedding layer and an encoder layer, is shared among all stages. Diverging from traditional CNN networks, which utilize different convolutional strides for obtaining multi-scale feature maps, SL(t)-ViT adopts

a progressive shrinking strategy. This strategy adjusts the scale of feature maps through the patch embedding layer.

For the first stage, if input images of size H*W*3 are provided to the model, they are divided into $\frac{H*W}{4}$ patches, each measuring 4*4*3. After obtaining these patches, they are flattened before being fed into the input layer. The patch size of the i-th stage, denoted as P_i, is introduced, allowing flexibility in adjusting the scale of feature maps at each stage. At the commencement of stage I, the input feature maps F_{i-1} with dimensions $H_{i-1} * W_{i-1} * C_{i-1}$ where, C_{i-1} describes the channel features extracted at $i - 1$ stage, H_{i-1} describes the height wise features extracted at $i - 1$ stage, W_{i-1} describes the features extracted across the width at $i - 1$ stage are evenly divided into patches.

$$\frac{H_{i-1}*W_{i-1}}{P_i^2} \tag{5.2}$$

Each patch obtained from equation (5.2) is then flattened and projected to a C_i dimensional embedding through a linear projection. After this linear projection, the shape of the embedded patches can be perceived as $\frac{H_{i-1}}{P_i} * \frac{W_{i-1}}{P_i} * \frac{C_{i-1}}{P_i}$. After that, the whole positional embedding is sent via the transformer encoder. In the transformer encoder block, these embedded patches are first sent through the attention and feed forward layers. Instead of utilizing the default self-attention mechanism the multi scale Nystrom attention mechanism is employed.

Initially a subset of landmarks is selected to approximate the full attention matrix from equation (5.3). Select m landmark points from E_i to approximate the attention matrix. Assume these are the first m patches for simplicity.

$$Em = Ei[:m] \tag{5.3}$$

The attention matrices are calculated for both local and global features. By applying the Nystrom attention mechanism, as described in equation (5.4), to individual patches, fine details indicating local features are derived.

$$Local\ Features = NystromAttention(Esmall) \qquad (5.4)$$

Aggregate the attention outputs from the small-scale patches to form the local feature representations.

Then, by employing equation (5.5) Nystrom attention mechanism is applied to individual patches to capture the overall structure and context defining global features.

$$Global\ Features = NystromAttention(Elarge) \qquad (5.5)$$

Aggregate the attention outputs from the large-scale patches to form the global feature representations.

By employing equations (5.6), (5.7) and (5.8) fuse the local and global features obtained from different scales. This is done through concatenation followed by a linear layer and another attention mechanism to combine the features effectively.

$$Flocal\ = concat\ (Local\ Features) \qquad (5.6)$$

$$Fglobal\ = concat\ (Global\ Features) \qquad (5.7)$$

$$F = Wf[Flocal; Fglobal] + bf \qquad (5.8)$$

where $[F_{local}; F_{global}]$ denotes concatenation of local and global features, and $W_f \in R^{D \times 2D}$ and $b_f \in R^D$ are the weights and biases of the fusion layer.

The final output is a combination of local and global features, which can be fed into subsequent layers for classification.

The multi scale Nystrom attention mechanism is applied at multiple scales to capture different levels of details. The self-attention mechanism is not computationally efficient when compared with the proposed attention mechanism. This is because self-attention takes large

computational resources while working on large sequences resulting in quadratic time, whereas the proposed attention module allows for efficient parallelization by operating on a smaller set of reference points in linear time, making it more suitable for parallel processing architecture.

Further, the output from the attention layer is passed to the feed-forward layer. In subsequent stages, the obtained feature map F_1 serves as input. Each successive stage follows a similar process, generating feature maps F_2 and F_3 with strides of 8 and 16 pixels, respectively, concerning the input image. Employing multiple stages with increasing strides forms a feature pyramid $\{F_1, F_2, F_3\}$ capturing linear architectural representations of the input at different scales. This pyramidal structure proved beneficial for tasks requiring understanding fine-grained details and global context. In the last part of this analysis, the medical image dataset is considered the source material for vision converters. The SL(t)-ViT model is utilized to classify diseases such as glioma, meningioma, and pituitary tumours from the recorded data.

After training the model on the training set, the model is evaluated in terms of accuracy, and complexity of the model using the test set using performance evaluation metrics for assessing the 'model's performance on new unseen data. Batch size chosen is 32, the model is run for 50 epochs, learning rate employed is 0.01. The workflow of the proposed model for tumor type detection is explained in detail in algorithm 1.

Algorithm 1: Work flow for Tumor Type Detection
Start
Phase I: Dataset Description.
Step 1: The Dataset used is "brain_tumor" obtained from 233 patients.
Step 2: Input: an image I
Step 3:Output: a feature vector F
Phase II: Preprocessing.
Step 4:Convert the image to grayscale
Step 5:Apply a Gaussian filter to smooth the image
Step 6:Apply an edge detector (e.g. Canny) to obtain a binary edge map E
Phase III: Feature Extraction.

Step 7:Apply Gabor Filter bank to the image to obtain a set of filtered images {G1, G2, ..., Gn}

Step 8:Concatenate all the computed features into a feature vector F

Phase III: Image Augmentation using GAN.

Step 9:Input: a dataset of images D, Output: an augmented dataset D_aug

Step 10:Define the GAN architecture and training parameters

a. Set the batch size for training the GAN

Step 11. Train the GAN on the original dataset D to generate new images

a. Preprocess the images in D (e.g., normalize pixel values to [-1, 1])

b. Train the GAN for the specified number of epochs using batches of images from the D

c. Save the trained generator G

Step 12: Augment the dataset D by generating new images using the trained generator G

Step 13: for each image, I in D do

a. Apply random geometric transformations (e.g. rotate, translate, scale) to

b. Apply random colour distortions (e.g. brightness, contrast, hue) to I

c. Resize I to a random size

d. Randomly delete information from I (e.g. pixels, regions)

e. Use the trained generator G to generate a new image I_aug based on the I

f. Add I_aug to the augmented Dataset D_aug

Step 14:Output the augmented Dataset D_aug

Phase IV: Compute Structural Similarity Index Measure for Data Generation.

Step 15:Compute SSIM

- Compute the SSIM numerator:

The numerator is computed by multiplying two terms:

The first term, $(2_{\mu_I \mu_R} + K_1)$, represents the product of the means of the two images μ_I and μR along with the stabilization constant k_1.

The second term, $(2Cov_{IR} + K2)$ represents the product of the product of the covariance of the two images covIR along with the stabilization constant k_2.

num $\leftarrow (2_{\mu_I \mu_R} + K_1)\ (2Cov_{IR} + K2)$

- Compute the SSIM denominator:

The first term, $(\mu_I^2 + \mu_R^2 + K1)$, represents the sum of squares of the means of the images along with the stabilization constant k_1.

The second term, $(Var_I^2 + Var_R^2 + K2)$, represents the sum of variances of the images along with another stabilization constant k_2.

den $\leftarrow ((\mu_I^2 + \mu_R^2 + K1)\ (Var_I^2 + Var_R^2 + K2)$

- Compute the SSIM value:

ssim = num/(den)

step 16: Return the SSIM value.

Step 17: Output: An optimized dataset D_opt.

Step 18: Define parameters.

a.   Define the window size w (e.g. 11)

b.   Define the constants (k1, k2) and dynamic range of pixel values (L) for SSIM calculation (e.g. k1 = 0.01, k2 = 0.03, L = 255)

Step 19: Iterate through images in D and optimize them.

Step 20: for each image I in D do

a.   Compute the SSIM between original image I and reference image R using the window size w and constants k1,k2,L

b.   If the SSIM is greater than or equal to the threshold T,

skip to the next image

c.   Otherwise, optimize I using a specified optimization algorithm (e.g., gradient descent, genetic) to maximize the SSIM with R.

d.   Add the optimized image I_opt to the optimized Dataset D_opt

Step 21: Output the optimized Dataset D_opt.

Step 22: odel.save('data_structural_tumor.h5')

Phase V: Disease Classification using ViT model.

Step 23: Define hyper parameters.

a.   Define the learning rate alpha (e.g., 0.001)

b.   Define the loss function (e.g., min-max)

c.   Define the optimizer (e.g., Adam)

Step 24: Split data into training and validation sets.

- Split D and L into training and validation sets with a specified ratio (e.g., 80:20)

Step 25: Prepare data for training.

- Create a data loader for the training set with batch size B
- Create a data loader for the validation set with batch size B

Step 26: Train the SL(t)-ViT model.

Step 27: for each epoch in E do

- Set the model to evaluation mode
- Compute the validation accuracy using the predicted and true labels
- Print the epoch number and validation accuracy.

Step 28: Output the trained SL(t)-ViT model.

End

5.4 Experimental Setup and Results Analysis

The experimental setup and detailed results analysis are presented in the following section:

5.4.1 Experimental Setup

Every experimental test is conducted concurrently on two NVIDIA GPUs, specifically A5000s with 24 GB of RAM each. RAM for the system is 128 GB. Numerous data analysis frameworks, such as Pandas, Numpy, Seaborn, 496 Matplotlib, and Scikit-learn, were employed in the study. The entire framework is executed for ten epochs. 20% of the data were used for testing, while the remaining 80% were used for training. The proposed model was trained on the training data. The suggested model was fine-tuned utilizing a variety of parameters, including hyper parameters, to improve classification accuracy and prevent overfitting. Different learning rates were applied, and it was found that the default learning rate of 0.001 gave better results.

5.4.2 Quantitative Analysis:

Every experimental test is conducted concurrently on two NVIDIA GPUs, specifically A5000s with 24 GB of RAM each. RAM for the system is 128 GB. Numerous data analysis frameworks, such as Pandas, Numpy, Seaborn, 496 Matplotlib, and Scikit-learn, were employed in the study. The entire framework is executed for ten epochs. 20% of the data were used for testing, while the remaining 80% were used for training. The proposed model was trained on the training data. The suggested model was fine-tuned utilising a variety of parameters, including hyper parameters, to improve classification accuracy and prevent overfitting. Different learning rates were applied, and it was found that the default learning rate of 0.001 gave better results.

The proposed algorithm not only detects Brain tumors but also detects which type of tumor exists. The proposed algorithm explicitly performs Binary Classification, and if "Yes,"; the Multiclass Classification is performed based on the 'Yes' value.

The proposed algorithm first detects "the existence of tumours". Then Gabor filters are used to evaluate visual texture using the mathematical function described in the methodology section. Gabor filters are Gaussian envelope-modulated sinusoidal waves. Sinusoidal wave frequency and Gaussian envelope standard deviation define the filter to recover brain texture from MRI data. These traits are used to identify tissue types or brain tumours. The picture texture strength is measured by evaluating its energy density across various frequencies and orientations, and based on this, the magnitude of the filtered pictures is calculated. We can understand response intensity at each picture point by measuring the magnitude using Gabor filter. Also, by analysing Gabor filter magnitudes picture areas with interesting texture properties are revealed. Later, PCA is used to reduce data complexity by finding patterns and relationships. PCA can detect Gabor filter-extracted texture patterns in brain tumour MRI images.

In the next step, using MRI tumor data, a GAN produces synthetic tumor pictures that look like actual ones. GANs have generators and discriminators with respective loss functions. As the Generator tries to fool the discriminator with realistic images, its loss function is optimizing the images based on feedback from the discriminator. It optimizes the loss between the generated images and a vector of ones. The discriminator's loss function optimizes the ability to distinguish between real tumor images, a vector of ones, and generated images, a vector of zeros. The generator and discriminator are combined and use a loss function class min-max loss function. The discriminator minimizes this loss function to differentiate actual from produced tumor pictures. The generator is trying to generate realistic tumor images by minimizing the loss based on the discriminator's classification of that image. The discriminator tries to distinguish between real and generated tumor images by minimizing them with zeros and ones.

It depicts that the loss degrades as the number of epochs increases, as depicted in Figure 5.3 It can be seen from Figure 5.3 that the generator and discriminator try to outperform each other. Their models get optimized based on each 'other's output. The discriminator and generator should have a balanced loss optimization per epoch. In a discriminator or generator, if one model loss decreases and the other one decreases. That means the GAN model is not optimizing well. The trained model is not able to generate good images in that scenario. From Figure 5.3 We can observe that both generator and discriminator loss is decreased, ensuring good performance of the GAN.
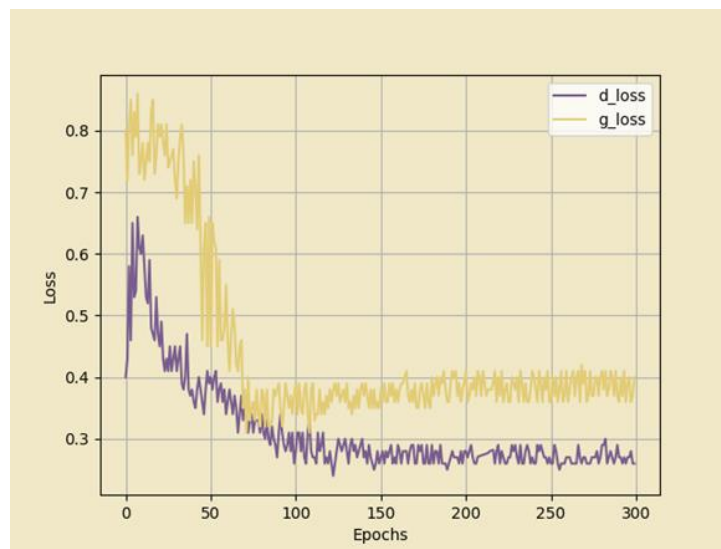
Figure 5.3 Generator and Discriminator loss optimization per epoch.

The sequential model produced good results when trying to classify images with tumors and without for binary classification. Based on the valuation, the proposed sequential model comes with a loss of 0.4 and an accuracy of 99.7. It shows that training loss is extremely low; that's why accuracy is exceptionally high because loss and accuracy are inversely proportionate, as depicted in Table 5.1.

Table 5.1 Summary of binary Classification

| Class Type	Class Category	Accuracy(%)			Loss	Model	Layer
		Validation	Train	Test			
Binary Classification	Yes, No	99.78	99.89	**99.7**	0.4	Sequential Model	Conv2D, MaxPooling, Flatten, Dense, Dropout

To predict the existence of a tumor, the proposed model is used where the test accuracy is 99.7. Figure 5.4 visualizes the comparison of accuracy on the Y-axis between the proposed model and the state-of-the-art technique, plotted on the X-axis. From Figure 5.4 it can be depicted that when compared between the proposed sequential model with EfficientNet gives a higher accuracy of 99.7 compared to the EfficientNet model, 97.8.
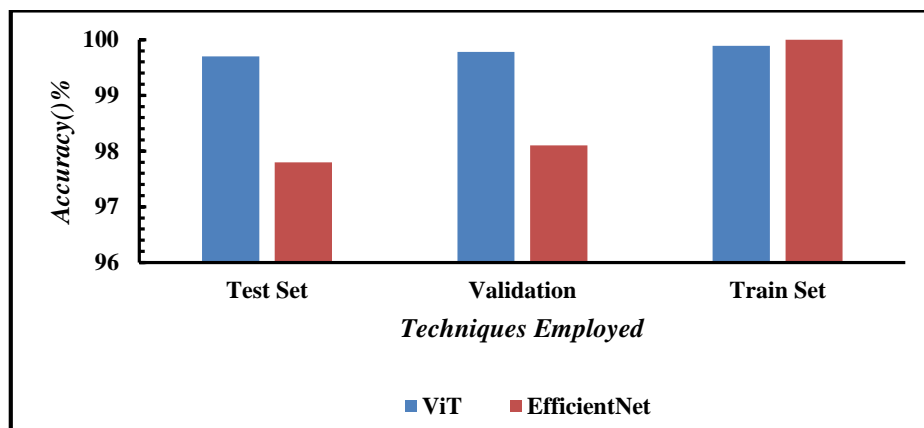
Figure 5.4 Results comparison between SL(t)-ViT and EfficientNet for Binary Classification

The novelty of the proposed algorithm lies in its hierarchical classification approach, which not only detects the tumor but also categorizes it into types such as Glioma, Meningioma, and Pituitary tumors. The shrinking Linear Time Vision Transformer (SL(t)-ViT) model is used to segment the image into a 14 * 14 matrix to detect the multiclass tumor at the initial phase. Using SL(t)-ViT models the given input image is divided into several non-overlapping sections called patches. These patches are then used as input to the encoder to the transformer.

126

All these patches give a global representation, and these global representations are used for Classification.
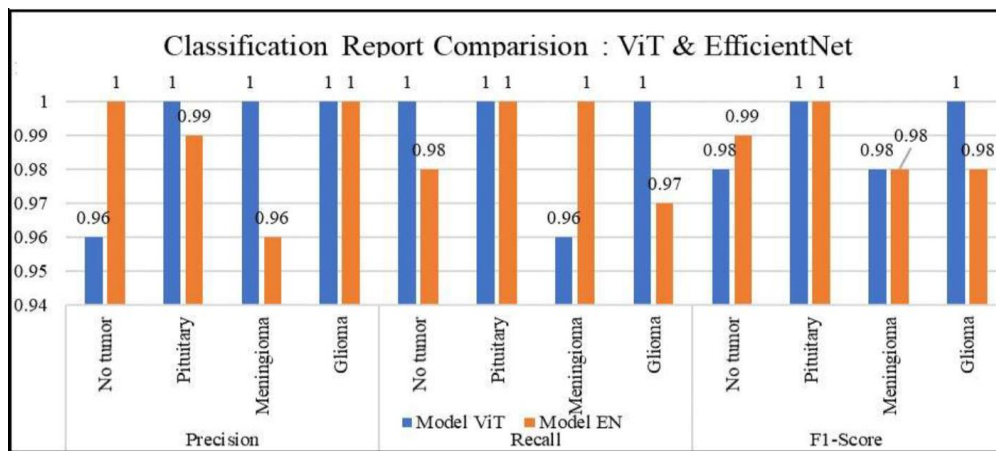
Figure 5.5 Comparison of results between SL(t)-ViT and EfficientNet for multi class classification

Figure 5.5 depicted via visualization comparison between the proposed SL(t)-ViT model with Efficient Net(EN) for the following performance metrics Precision, Recall and F1-Score for each class, namely No-tumor, Pituitary, Meningioma, Glioma, and it can be inferred from the graph the proposed model showed good result compared to EN on Figshare dataset. The effectiveness of a classification model on data where the true values are already known can be summarized in a table called a confusion matrix. It's useful for seeing how well an algorithm performs and gauging how well a model does. In the case of imaging classification for brain tumors, a confusion matrix can be used to describe the model's performance in identifying the different types of tumors, such as glioma, meningioma, and pituitary tumors. In a task involving many classes to be classified, the confusion matrix has the form of a table with rows and columns that the actual and anticipated class labels.
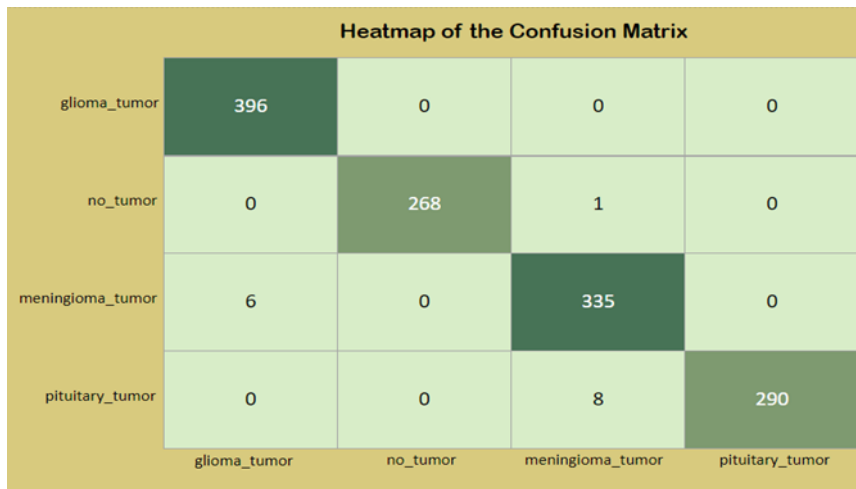
Figure 5.6 Confusion matrix

The confusion matrix for a multiclass classification problem can be represented in Figure 5.6 Performance metrics like recall, precision, accuracy, and F1-score can be computed from the confusion matrix and used to assess the model's effectiveness concerning each class and as a whole.

Table 5.2 Accuracy of Proposed vs SOTA on Binary classification

Dataset	Type of classification	Accuracy achieved by SOTA architecture (Deep CNN) (%)	Accuracy of Proposed model(%)
Br35H	Binary classification	99	**99.5**
Kaggle	Binary classification	88.26	**99.6**
BT-small 2C	Binary classification	92	**99.54**
BT- large 2C	Binary classification	98	**99.8**

This section evaluates the performance of the proposed model in classifying brain tumors compared to the latest state-of-the-art (Deep CNN) techniques. Table 5.2 provides a comparison of the proposed model with several recent SOTA brain tumor classification methods across various publicly available binary classification datasets. With superior accuracy scores, the proposed model surpasses the most advanced techniques.
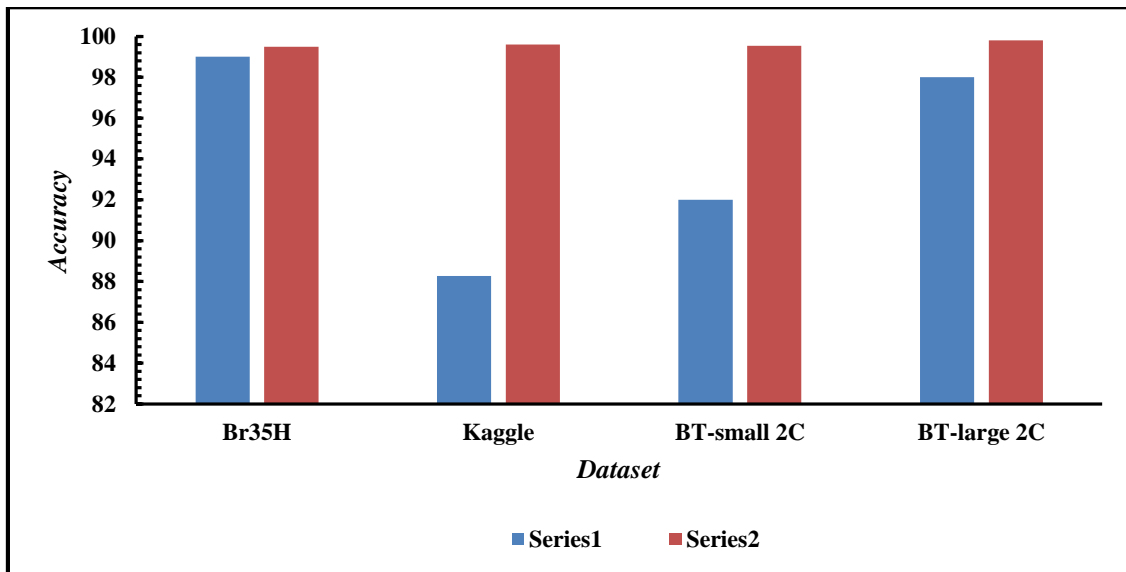
Figure 5.7 Classification report comparison SL(t)-ViT vs SOTA on various publically available binary classification datasets
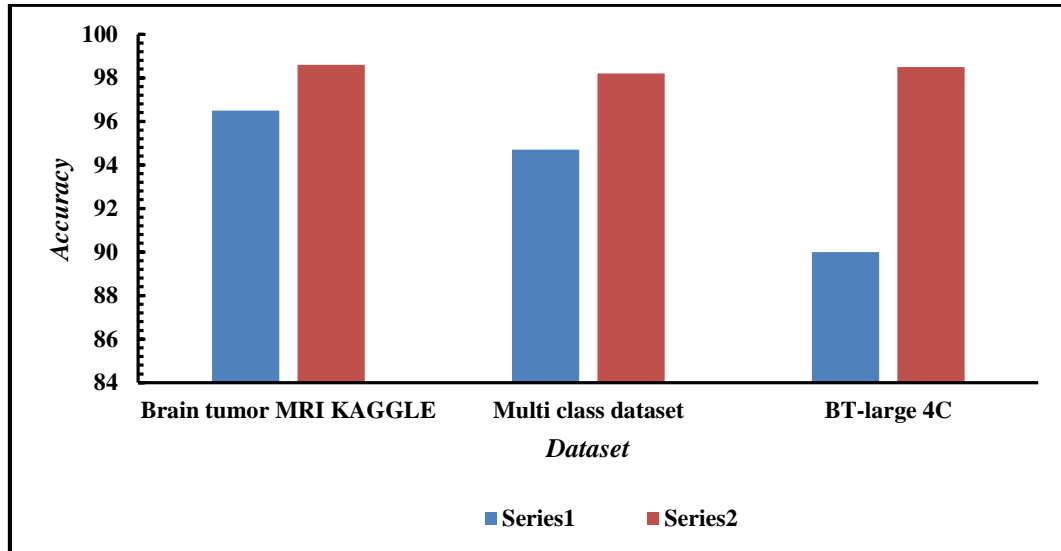
Figure 5.7, depict the graphical representation of SOTA techniques Deep CNN performance versus the proposed model's performance on different datasets. These visualizations demonstrate the effectiveness of the proposed hybrid model over recent SOTA techniques when applied in the current environment. From graph it can be understood that by employing the hybrid model the performance of the model increased between 8% to 1% on various datasets proving the superiority of the proposed model.

Table 5.3 Accuracy of Proposed vs SOTA on Multi classification.

Dataset	Type of classification	Accuracy achieved by SOTA architecture (DL-MaiVot) (%)	Accuracy of Proposed model(%)
Brain tumor MRI KAGGLE	Multi class classification	96.50	**98.6**
Multi class dataset	Multi class classification	94.70	**98.2**
BT-large 4c	Multi class classification	90	**98.5**

In this section, we assess the proposed model's performance in classifying brain tumors relative to the latest state-of-the-art (SOTA) techniques. Table 5.3 offers a comparison between

the proposed model and various recent SOTA brain tumor classification methods across diverse

publicly available multi-class classification datasets. With higher accuracy scores, the proposed

model outperforms the most advanced techniques.

Figure 5.8 Classification report comparison SL(t)-ViT vs SOTA on various publically
available multi-class classification datasets

Figure 5.8, illustrate the graphical representation of various SOTA technique, (DL-

MaiVot) performance compared to the proposed model's performance across different datasets.

These visualizations highlight the efficacy of the proposed hybrid model over recent SOTA

techniques when implemented in the current environment. From the same graph it can be

inferred that by employing the proposed model the effectiveness in classifying tumour

increased from 40% to 20% on minimum making it the most effective technique.

5.4.3 Complexity analysis

Our framework is designed to enhance both the effectiveness and computational

efficiency of the proposed model. When existing techniques were implemented on the same

platform, they required 5.62, 4.98, 5.01, 5.25, and 5.12 seconds per image, highlighting their

relatively higher computational inefficiency. Figure 5.9 graphically presents this comparison,

with the x-axis representing the time taken per image and the y-axis indicating the classifiers

or techniques used. As shown in Figure 5.9, the proposed model consistently outperforms state-of-the-art models by requiring significantly less processing time under the same conditions. These results underscore the proposed model's effectiveness and superior computational efficiency, making it an excellent choice for practical applications.
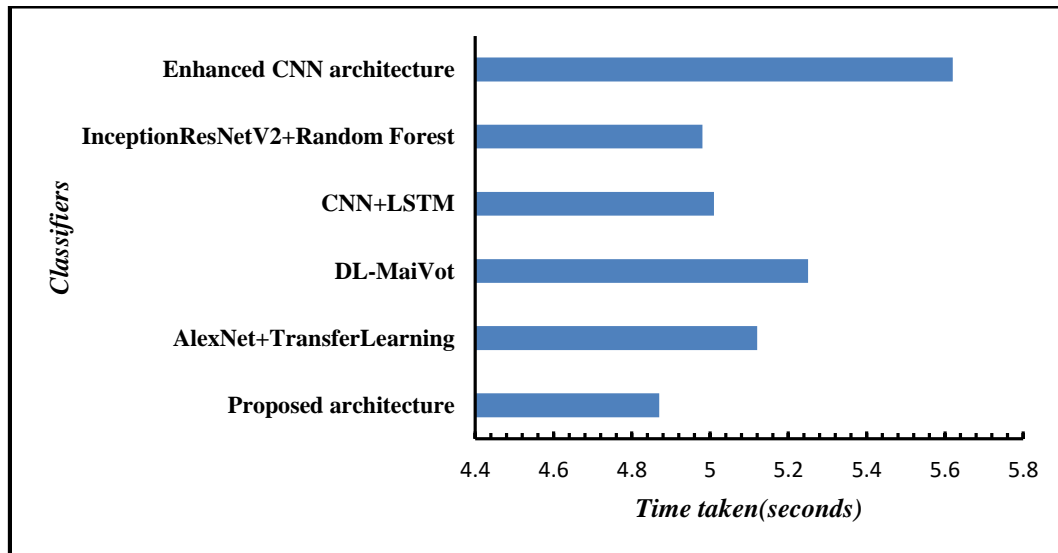
Figure 5.9 Comparison of time taken per image in seconds for SOTA vs proposed

5.5 Chapter Summary

This chapter presents SL(t)ViT, an innovative model for brain tumor classification, designed to overcome key limitations in deep learning due to the scarcity of diverse MRI datasets. Traditional data augmentation (DA) methods often fall short, as the artificially generated images tend to share the same distribution as the original dataset, limiting their ability to enhance model generalization. To address these challenges, this study leverages Generative Adversarial Networks (GANs) to improve data diversity and quality, resolving issues such as image blurriness and model collapse. The SL(t)ViT model employs a GAN-based data augmentation strategy, further reinforced by a structural similarity loss function, which ensures the generation of more realistic and varied training data. A key innovation of SL(t)ViT is its shrinking mechanism, which significantly enhances both binary and multiclass classification

tasks by reducing computational complexity without sacrificing accuracy. Departing from the conventional self-attention mechanism used in Vision Transformers (ViT), SL(t)ViT incorporates a multi-scale Nystrom attention mechanism. This novel approach allows the model to efficiently capture multi-scale features while operating in linear time, providing both computational speed and scalability. SL(t)ViT, redefines the approach to brain tumor classification, offering a ground breaking combination of GAN-augmented data generation, efficient attention mechanisms, and a resource-conserving architecture that sets a new benchmark in accuracy, speed, and model efficiency in the field of medical imaging.

Chapter Six: A FRAMEWORK FOR IMAGE CLASSIFICATION USING HETEROGENEOUS ATTENTION MECHANISIM AND OPTIMIZED FEATURE SELECTION

In this chapter a novel Deep Learning (DL) framework that addresses optimization problem in imbalance dataset, a common challenge affecting model performance is addressed. The proposed novel Attention Mechanism, for improved focus on critical features, alongside a Resource-Efficient Optimization model that streamlines computational demands was developed and results were evaluated.

6.1 Introduction

Cancer is a leading global cause of death, surpassing many other diseases. In 2020, it was reported that over 600,000 individuals died out of nearly 1.9 million diagnosed cases, with predictions suggesting that this number may exceed 28 million by 2030 [123]. Among various forms, certain types contribute significantly to mortality rates. The development of abnormal cell growth leads to various tumor classifications, which differ in aggressiveness and treatment strategies. Factors such as hormonal changes, lifestyle choices, and environmental influences are known to increase the risk of developing these conditions. Tumors can be categorized into benign and malignant based on their invasive capabilities, with malignant tumors posing significant challenges to surrounding tissues. Early diagnosis is crucial for successful treatment and has been linked to improved survival rates.

Diagnostic techniques play a vital role in identifying these growths at early stages, enabling more effective interventions [89]. However, these methods can face challenges, such as variations in tissue composition and the difficulty of distinguishing between different types of lesions. This complexity often results in variability in diagnostic accuracy, highlighting the need for advancements in technology to support professionals. To enhance accuracy, integrating computer-aided detection (CAD) systems and machine learning (ML) technologies

has shown promise. While traditional ML approaches have limitations, advancements in deep learning (DL) demonstrate substantial potential across various recognition tasks. For instance, Vision Transformers (ViTs) have been developed to address challenges associated with conventional models by utilizing self-attention for improved performance. However, their complexity can lead to inefficiencies. To overcome these challenges, a novel attention mechanism has been introduced, optimizing focus on critical features and enabling real-time applications. The proposed multi-model fusion framework in this chapter enhances accuracy and resource efficiency, ensuring greater generalizability across diverse datasets while addressing limitations in existing methodologies.

6.2 Key contributions

The key contributions of the proposed work are outlined as follows:

(i) Innovative DL framework for Breast Cancer Image classification: We introduced a pioneering and highly efficient DL framework tailored for image classification. This framework stands out for its novel design, providing a powerful tool for accurate and generalized classification in medical imaging datasets.

(ii) Hybrid Approach for Improved Generalization Across Varying Dataset Sizes: Addressing the challenge of generalization across varying dataset sizes, we presented a hybrid approach that merges the features extracted from two classifiers. This synergistic combination enhances the model's ability to adapt to diverse dataset scales, ensuring robust performance across a spectrum of data complexities.

(iii) Efficient Attention mechanism for improved performance and computation efficiency: A novel attention mechanism is proposed in one of the classifier employed in the framework, contributing to both enhanced accuracy and computational efficiency. This mechanism optimizes the model's focus on critical features, leading to improved classification

outcomes while maintaining computational efficiency. This approach represents a significant advancement in attention mechanism within DL classifiers.

(iv) Resource-Efficient Optimization through feature selection: To address the critical issue of computation resource efficiency, a sophisticated feature selection algorithm is proposed for optimal hybridization the classifiers. This algorithm ensures that only the most pertinent high-dimensional features that contribute to the classification process are selected, streamlining computational usage without compromising accuracy. Proposed approach stands as a pioneering effort in achieving a judicious balance between computational efficiency and model performance.

(v) Consideration of Heterogeneity for Accurate Tumor class prediction: In contrast to conventional approaches, proposed framework takes into account the inherent heterogeneity within each classifier. By fully integrating high-dimensional features from both classifiers we achieve a more comprehensive understanding of the intricate variations present in the data. This consideration of heterogeneity is essential in enhancing the accuracy of tumor class predictions, marking a substantial leap forward in the precision of breast cancer diagnosis.

(vi) Extensive Analysis for High Accuracy, Efficiency Resource usage and scalability: Our study presents a comprehensive quantitative analysis using BreakHis and BACH datasets, evaluating the performance of our proposed model. The results indicate our framework's potential for enhancing breast cancer prediction in clinical and research domains. Methodological innovations encompassing framework and hybrid model design, sophisticated feature selection, and addressing classifier heterogeneity, collectively propel breast cancer classification to the forefront, optimizing accuracy and methodology.

6.3 Dataset Description and Pre-processing

The dataset utilized and the preprocessing techniques applied are discussed in detail below:

### 6.3.1 Dataset

BreakHis Dataset: BreakHis dataset contains Histopathological images of 9,109 microscope images of breast tumour tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, 400X). BreakHis dataset is divided into two main groups: benign tumors and malignant tumours. It has 5,429 malignant and 2,480 benign 700*460 pixel PNG samples within 3-RGB channels with depth of 8-bit 8-bit in each channel in PNG format. Figure 6.1 shows the distribution of images among two classes malignant and benign. It is evident from Figure 6.1 that the dataset is imbalanced. Figure 6.1, 6.2 and 6.3 shows the sample malignant and benign images utilized.

BACH Dataset: The BACH 2013 dataset, is composed of histological images related to breast cancer, sourced from the Breast cancer 2019 grand challenge. These images originate from biopsy slides of breast tissue, stained with hematoxylin and eosin. Uniformly acquired using a Leica DM 20000 LED microscope and a Leica ICC50 high-definition camera. Each image measures 2048*1536 pixels with a pixel scale of 0.42µm*0.42µm. The dataset comprises 400 images obtained from diverse patients in Covilh and Porto which were labelled by two medical experts. Figure 6.4 shows the sample malignant and benign images of the BACH dataset.
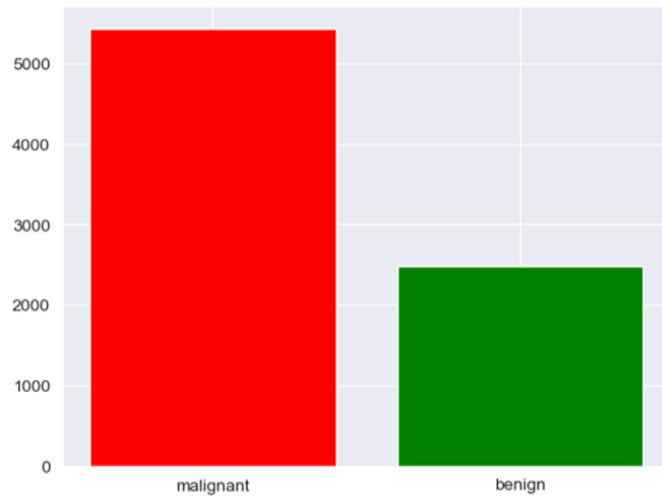
Figure 6.1 Distribution of images among the classes
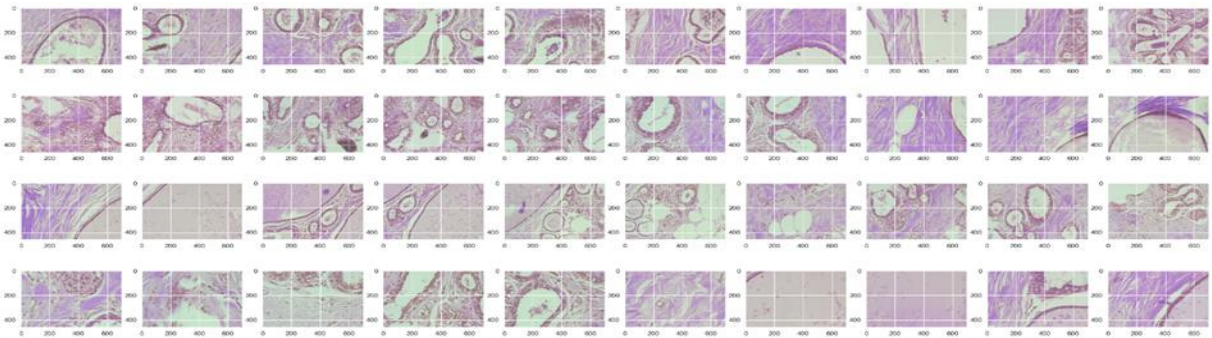
Figure 6.2 Sample benign images from the BreakHis dataset
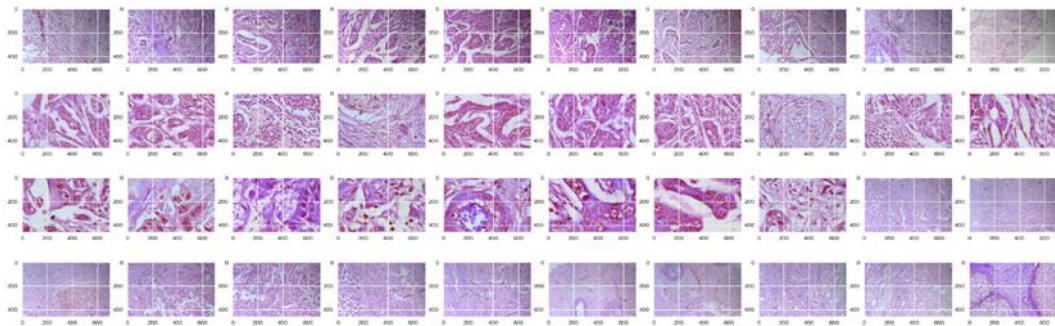
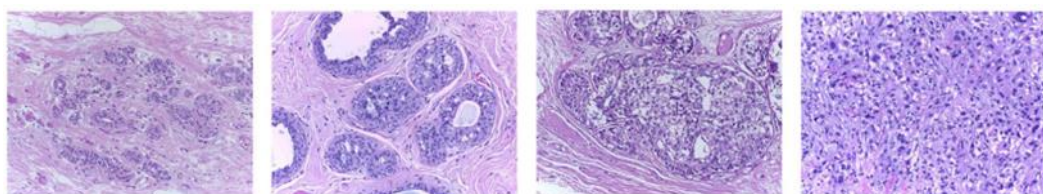**Figure 6.3 Sample malignant images from the BreakHis dataset**

Figure 6.4 Sample images from the dataset BACH

137

6.3.2 Data Preparation

To improve the efficiency of deep learning (DL) models, enhancing their generalization capacity is essential. Generalizability refers to the difference in performance when a model is evaluated on familiar versus unseen data. Poor generalization often results from overfitting, where models adapt too closely to the training data. Effective DL models should show a consistent decrease in validation error alongside training error, with data augmentation being a key technique to achieve this. Models trained on augmented datasets generally outperform those trained on the original dataset, with cropping, rotation, and flipping proving to be particularly effective. For this study, images from the BreakHis dataset are divided into training, validation, and testing sets. Data augmentation techniques employed include horizontal and vertical flipping, generating new images by rotating the original at 90-degree intervals, and cropping, which extracts and resizes portions of images while maintaining spatial dimensions. Rotating images by 90 degrees avoids introducing background noise, allowing the network to focus on relevant features.
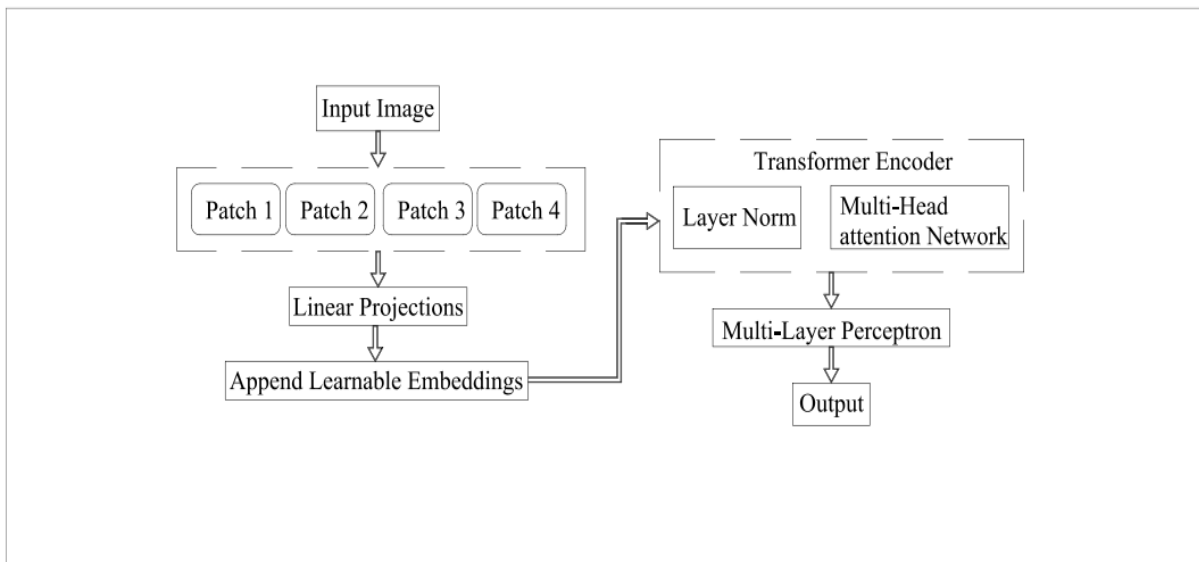
Figure 6.5 Flowchart of the proposed methodology

6.4 Proposed Methodology

The flowchart of the proposed methodology is visualized in Figure 6.5. Initially, the dataset (BreakHis/BACH) is provided as input to the model. In the subsequent phase, pre-processing is performed on the input data because both the datasets are imbalanced and irregular, which can lead to overfitting or under fitting. To address this issue, the dataset is pre-processed using various techniques such as flipping, rotating, and cropping. This enhances the model's efficiency and its ability to generalize to unseen data.

In the next step, the pre-processed data, which includes both original and newly generated images, is fed into the framework. Following this, the dataset is input into classifier-1, and the accuracy obtained by classifier-1 is calculated. The accuracy is then compared with a threshold value. If the achieved accuracy is greater than the threshold, the model outputs the performance efficiency and the classification process ends. If the obtained accuracy is less than the threshold, the features extracted by classifier-1 are stored in a vector, and the initial input images are then provided to classifier-2.

The procedure of comparing the accuracy obtained by classifier-2 with the threshold is repeated. If the achieved accuracy exceeds the threshold, breast cancer classification is performed. If the accuracy is still below the threshold, the features obtained by classifier-2 are also stored in the vector. If both classifiers fail to achieve accuracy greater than the threshold, the features obtained by classifier-1 and classifier-2 are fused using the proposed feature fusion technique and stored in a vector. These fused features are then given as input to the hybrid model for the final classification of breast cancer.

Due to the imbalance condition of the dataset models often show bias towards any of the class labels. This occurs due to the model's disability to learn important features from the database by concentrating only on one class. In such conditions to improve the classification efficiency building a hybrid model by taking the benefit of features obtained from both the

139

classifiers is always appereciatable. At the same time, hybrid models often are time, space-consuming making them computationally inefficient and are often problem-specific leading to a lack of generalizability. To overcome this problem in the following work a framework is proposed that is computationally more efficient by improving the performance and generalization. Figure 6.6 visualizes the diagrammatic representation of the framework proposed.

Q1 is classifier 1 and Q2 is classifier 2 employed in this framework. The model's ability to determine the weight factor is crucial for obtaining superior performance with increased computational efficiency. The main advantage of this framework is that it initially performs on each classifier individually and when individual classifiers do not perform well then it takes advantage of the hybrid model. So to evaluate whether a classifier is performing well or not we need to analyse their performance. For this in our model, we integrated a threshold value where the accuracy acquired from the classifiers is compared with the threshold value. The threshold value is obtained by calculating the average accuracy values obtained by hybrid models reviewed in the literature. Let the threshold value be δ and the predicted accuracy be Q. If the acquired accuracy value of Q1 is $Q > \delta$ it means the classifier 1 is trained well on the dataset and does not require the activation of the second classifier. If the calculated value of Q1 is $Q < \delta$ then this implies that the classifier is not trained well on the dataset and the second classifier is activated. Likewise, for the second classifier also the value of Q2 is evaluated comparing it to the threshold value.
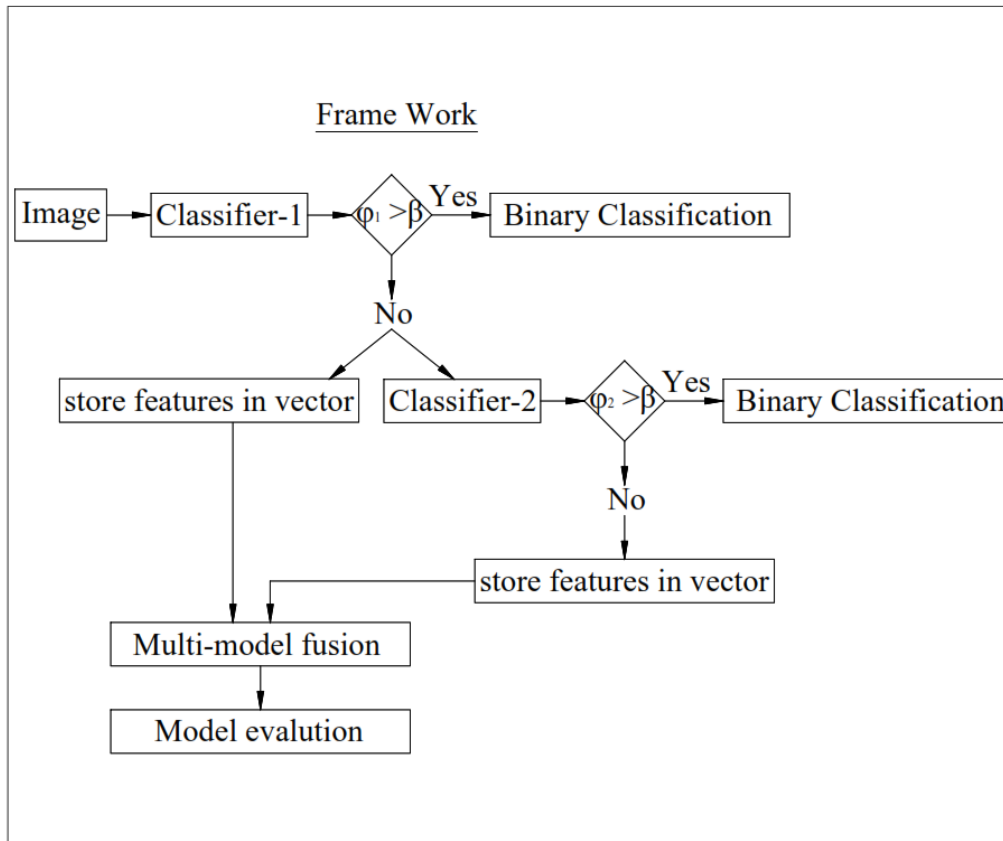
Figure 6.6 Framework proposed

Algorithm 1: Work flow for proposed framework

Input: $Dataset(\{x_i, y_{i,}3\}_{i=1}^{n}$
Initialisation: Trainable parameters P
Size of Batches B
No. of Batches B_{total}
No. of epochs E
Initial Learning Rate
Output:
Trained Framework
Step 1:
Classifier 1
       Initialize (p, β)
       for epochs 1,2,3,…… E do
       for B= 1,2,3,…… B_{total} do
       $x_{l+1} = x_l + F(x_l)$
       $x_{l+1} < -- Q_l$
       $Q_{l>} β$
End for
Else if
       $Q_{l<} β$
       Store Q_l in vector $model_l$
Step 2:
Classifier 2
       Initialize (p, β)
       for epochs 1,2, 3,…… E do
       for B= 1,2, 3,… B_{total} do

$$Y_{c,i} = \Sigma_i \, Norm(A_{c,i,ij} x_{g,j,c,i})$$
$$PE_{pos,2i} = Sin(pos|10000^{2i}|dmode)$$
$$PE_{pos,2i+1} = Cos(pos|10000^{2i}|dmode)$$
$$Q_{2>}\beta$$
End for
Else if
$$Q_{2<}\beta$$
Store Q_2 in vector model$_2$
Step 3:
Fuse the feature
$F_{fused =}\alpha. \, Model_1 + \mu. \, Model_2$
$\alpha+\mu=1$
End for

In the proposed work, InceptionResNetV$_2$ is taken as an initial classifier. InceptionResNetV$_2$ model takes input images with a shape of (75,75,3) where the size of the image is 75*75 with 3 RGB colour channels. The model is trained using pre-trained weights obtained from ImageNet. By initializing the model with weights obtained from ImageNet our model benefits from pre-learned weights that can be fine-tuned for a specific task. After initializing the weights, the input image is then passed through the InceptionResNetV$_2$, which is a hybrid architecture that combines the Inception module along with residual connections. The Inception modules are composed of parallel branches of different convolutions between which residual skip connections are used for allowing the gradient flow more directly mitigate the vanishing gradient problem. From here we got an output tensor with shape (1,1,1536) where image features have been reduced spatially to a 1*1 size while increasing the number of channels to 1536. Equations (6.1), (6.2), (6.3), (6.4), (6.5), (6.6) are formulated to explain the working of InceptionResNetV$_2$. Figure 6.7 shows the visual representation of the proposed classifier 1.
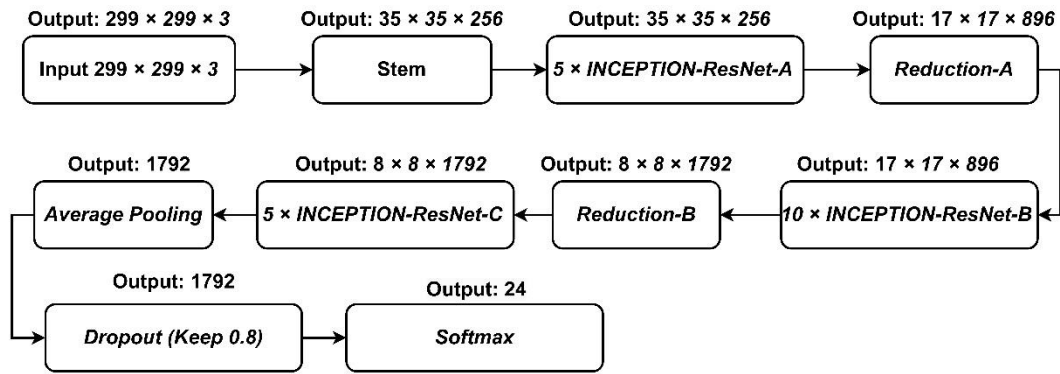
Figure 6.7 InceptionResNetV₂ architecture

The achieved output tensor is then passed through an average pooling 2D layer. Here, the spatial dimensions are further reduced by calculating average values for a window of values. These values help in capturing global information from the features. The output obtained from the average pooling 2D layer is flattened, using the flattened layer. In convolutional layers, the input data is typically represented as a 3D or 4D tensor, where each dimension corresponds to the spatial dimension and the number of channels. However, a fully connected layer only can work on the input of a 1D vector. For this reason, flattened layers are used to convert multidimensional input data into a 1D vector when input is transiting from convolutional layers to fully connected layers. These layers flatten the input tensor, preserving the total number of elements while removing the spatial dimensions. By using this layer our model learned relationships between different features regardless of their spatial arrangement. As the dataset is imbalanced to create a balance Dropout layer is added. A Dropout layer is applied with a rate of 0.5, which means that each element of the flattened vector has a probability of 0.5 of being zeroed out which means on average, half of the connections are set to zero.

The dropout layer disabled their contribution to the subsequent computations by making them not affect the forward pass or backward pass during training. This made the

143

proposed model more regularized and generalised. Now the remaining active connections from the flattened output are passed through a fully connected Dense layer. Here the fully connected Dense layers map these high-dimension features to the desired output enabling the network to learn complex patterns and the relationships in the data. The obtained output (i.e.) features extracted are passed on to the convolutional layers which classify the given images based on which class it belongs to. Once the accuracy in classifying an image is calculated the obtained accuracy calculated is compared with the threshold value. If the obtained accuracy is lesser than the threshold value, then the features obtained by the first classifier 1 are stored in the form of a vector and classifier 2 is activated for execution.

Equation (6.1) is the basic linear transformation step often found in convolutional neural networks. The equation describes the process of applying a convolutional operation to the input x_l using filters W, and adding a bias term α to produce an output feature map $P(x_l)$.

$$P(x_l) = W * x_l + \alpha \tag{6.1}$$

Where x_l represents input to the convolutional layer at layer l. This typically is a multi-dimensional tensor representing the feature map of an image. W are the convolutional filters or kernels to detect various features such as edges, textures, and patterns in the input image. α represents the offset or bias added to each element of the output feature map. It helps in shifting the activation function and is learned during the training process. $P(x_l)$ is the output feature map after applying the convolution operation and adding the bias.

Inception-ResNet employs residual connections within the Inception modules. The residual connections help with the flow of information through direct addition of the input to the output modules. Equation (6.2) provides the formula for a residual connection.

$$x_{l+1} = x_l + F(x_l) \tag{6.2}$$

x_l is the input to the residual block, $F(x_l)$ is the transformation function applied to the residual block involving a series of operations such as convolutions, batch normalization and activation

144

functions and x_{l+1} is the output of the residual block. It is the sum of the input x_l and the transformer input F (x_l).

Further, equation (6.3) is utilized to combine the information from two sources into T_1 to form the final output of the Inception-ResNet module.

$$T_l = S(P) + h(x_l) \tag{6.3}$$

T_1 gives the sum of two branches, S(P) is the information captured through the convolutional operation when the activation function is applied and h(x_l) is used for simple input transformation that provides an alternative path for information flow.

Equation (6.4) gives the final output of the residual module obtained by applying the activation function to the combined tensor T_l.

$$x_{l+1} = S(T_l) \tag{6.4}$$

x_{l+1} gives the final output of the residual module at layer $l + 1$,

T_l is the tensor obtained from the previous equation (8),

S is the activation function used and applied element-wise to T_l.

Equation (6.5) is the activation function used in the proposed architecture.

Equation (6.5) defines the ReLU activation function, which is widely used in the neural networks.

$$ReLu - R(x) = max(0, x) \tag{6.5}$$

x is the input value, max(0,x) is the output. The output is x if it is positive, otherwise, it is zero.

Equation (6.6) is used to represent a relationship between the input and outputs of different residual units, taking into account the residual function and scaling factor α. This equation illustrates how the outputs of different residual units are related to each other through scaling and the residual function.

$$\frac{\alpha X_n}{\alpha X_i} = \frac{\alpha X_i + P(X_i, W_i, \alpha_i)}{\alpha X_i} = \frac{1 + \alpha P(X_n, W_n, \alpha_n)}{\alpha X(n)} \tag{6.6}$$

X_i represents the iput of the $i - th$ residual usint, X_n is the input of the $n -$

th unit and P is the residual function eapplied to X_i with weights and bias and α

is the scaling factor

In the proposed methodology ViT is taken as classifier 2. Pre-processed images are given as input to the ViT model. Unlike models containing convolutional layers as their base ViT model solely relies on attention layers and feed-forward neural networks within the transformer encoder block to process the input image. ViT divides the input images into patches and processes them using the Attention mechanism. Instead of using the default self-attention mechanism of the transformer, we employed the Multi-Scale linear time Nystrom attention mechanism which aids in better computational efficiency comparatively. This attention mechanism can produce linear computations efficiency while still extracting the required global dependencies. Here the taken image size is (1,1536). Image size is a tuple of (height, and width). Figure 6.8 Visualizes the steps involved in the proposed classifier 2.
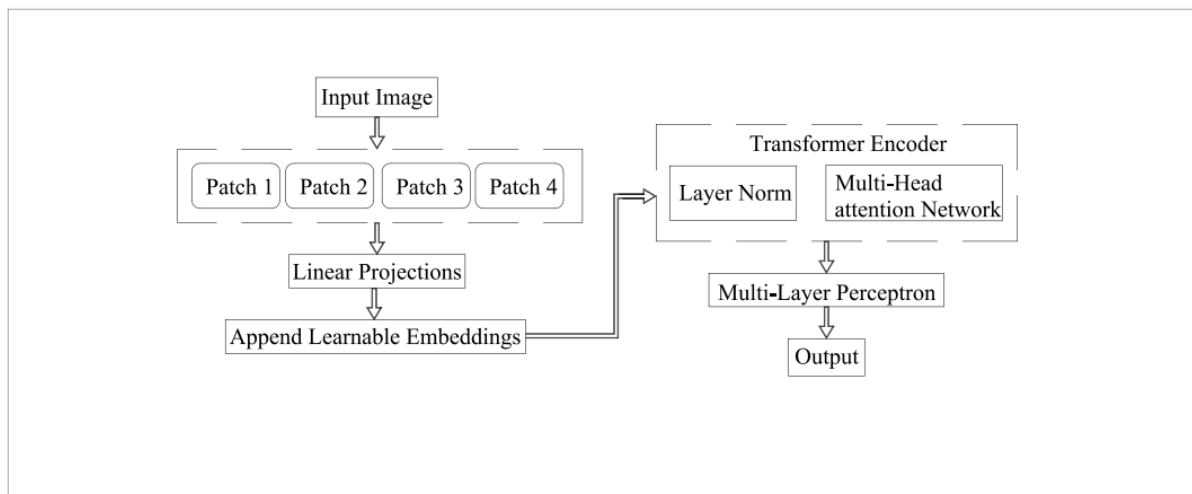
Figure 6.8 Vision Transformer architecture.

Further, the input images are divided into smaller patches of equal size where each patch belongs to a spatially localized region of the input. (1,16) is the patch size taken where patch size is a tuple of (height, width). After specifying the patch size, we initialized the

146

num_classes parameter that represented the number of output classes which is 2. Then the dimensionality of the transformers model is specified which represents the hidden size or the number of features in the model. Here the dimensionality of the ViT model is 256. Then the depth of the ViT model is specified which is used to calculate the capacity or complexity of the model. Here, the depth of our model is 9. The ViT model's attention heads are initialised, enabling the model to concentrate on various input segments simultaneously. The aim of our research work is to reduce the complexity of the model which was done by employing multi-time linear time Nystrom attention mechanism that regularized model depth. Several attention heads employed in our ViT is 8.

Initialized the MLP dimension (mlp_dim) parameter tospecifies the dimensionality of the feed-forward neural network used in the transformer model. The MLP layer is a component of the transformer model. Here, the layer has a hidden size of 256. Next, we initialized a dropout layer this layer is a regularization layer that randomly nullifies a random set of fractions of input units at each training step. Here, the dropout rate is set to 0.3 which implies that 30% of the patch embeddings are nullified. Then number of input channels used in ViT is specified. In this there is only one input channel indicating that the input image is in grayscale, positional encoding is added to the ViT model to incorporate the spatial information of the patches. Now the transformer encoder block processes the attention patches. This mechanism allows the model to perform with linear complexity by allowing the model to attend to relevant patches when encoding the information. Now, the classification heads are fed with output achieved from the transformer encoder block where the representation to the desired number of out classes is 2 enabling image classification. Equations (6.7), (6.8), (6.9), (6.10), (6.11), (6.12), (6.13), (6.14), (6.15) are formulated to explain the working of ViT.

Equation (6.7) represents the orientation of the original image. This representation indicates that the image is a 3-dimensional tensor with height H, width W, and C colour channels.

$$Original\ Image : x \in R^{H \times W \times C} \qquad (6.7)$$

x is the input image taken, R denotes that x is a tensor in real-valued space, H is the height of the image, W is the width of the image taken and C is the number of channels in the image, which is typically 3 for RGB images.

Equation (6.8) explains the converted patches and equation (14) gives the final number of patches images are divided into.

$$2D\ Converted\ Patches : x_p \in R^{N \times (p^2.C)} \qquad (6.8)$$

x_p is the tensor representing the patches of the space, N is the total no. of patches, p is the resolution of each patch.

Equation (6.9) is used to calculate number of patches. H, W are the resolutions of the image taken, C is the no.of channels, P is the resolution of each image patch.

N indicated the final number of patches images are divided into. N is then given as the input sequence for the transformer.

$$N = \frac{HW}{P^2} \qquad (6.9)$$

Equation (6.10) is the equation that describes the initial input sequence for the transformer

$$Z_0 = \left[X_{class}; X_{p^1}E; X_{P^2}E; \dots X_{p^N}E \right] + E_{pos} \qquad (6.10)$$

Z_0 is the initial sequence of embedding's. X_{class} is a learnable classification token, $X_{P^2}E$ The embedded patches, where E is the embedding matrix applied to each patch, and E_{pos} is the positional encoding added to each patch to retain positional information.

The equation (6.11) specifies the dimensions of the embedding:

$$E \in R^{(P^2.C) \times D}, E_{pos} \in R^{(N+1) \times D} \qquad (6.11)$$

E is the embedding matrix, $P^2.C$ is the dimensionality of each patch, D is the dimensionality of the embedding space, E_{pos} are the positional encoding matrices, $N+1$ is the number of positional encodings, including the classification token, and D is the dimensionality of the positional encoding.

Equations (6.12) and (6.13) show the series of multi-head linear time Nystrom attention and multi-head perceptron blocks present in the transformer encoder block.

$$Z_{l'} = MSA\big(LN(Z_{l-1})\big) + Z_{l-1}, \quad l = 1 \dots L \tag{6.12}$$

$Z_{l'}$ *is t*he output of the self-attention layer, MSA is the Multi-head self-attention mechanism employed, LN is the Layer normalization applied to the input sequence, Z_{l-1} *are the* input to the current layer, which is the output from the previous layer, and l is the layer index.

$$Z_{l} = MLP\big(LN(Z_{l'})\big) + Z_{l'}, \quad l = 1 \dots L \tag{6.13}$$

In equation (6.13) Z_l *is t*he output of the MLP block, MLP is the Multi-layer perceptron, LN is the Layer normalization applied to the input sequence, $Z_{l'}$ is the output from the previous self-attention layer, and L is the layer index employed.

Equations (6.14) and (6.15) display the positional encodings containing the absolute and relative positions of the tokens that are added to the input sequence embedded at the bottom of the encoder and decoder stacks.

$$PE_{pos,2i} = Sin(pos|10000^{2i}|dmode) \tag{6.14}$$

$$PE_{pos,2i+1} = Cos(pos|10000^{2i}|dmode) \tag{6.15}$$

$PE_{pos,2i}$ *is the* the positional encoding employed for even indices, $PE_{pos,2i+1}$ are the positional encodings employed for odd indices, pos is the position index, and d is the dimensionality of the model.

Now the obtained accuracy from the ViT model is compared with the threshold value, if the obtained classifier 2 accuracy is greater than that of the threshold value then the execution

terminates at that point and the probability of the image belonging to a particular class is returned if not then the features extracted are stored in a vector for future use.

ViT usually has an inbuilt self-attention mechanism which helps in identifying the region of interest while training the model. However, the self-attention mechanism present in the transformer often utilizes more computational resources, which results in the quadratic computational efficiency of the model. To overcome this problem, we modified the self-attention module in the transformer module and replaced it with multi-scale linear time Nystrom attention mechanism that takes linear or sublinear time to complete the task.

Our proposed technique employing the Multi-Scale linear time Nystrom approximation method can help address the memory and computational limitations associated with self-attention in transformer models. It aims to provide an efficient approximation of the attention mechanism, particularly for large-scale input sequences. It leverages a subset of randomly sampled tokens or patches from the input sequence to construct a low-rank approximation of the full attention matrix. This reduces the computational complexity and memory requirements from quadratic to linear or sublinear, making it more scalable for large inputs. But by employing this technique there is a chance that the model may miss out on some valuable features. So to overcome this situation a multi-scale weight adaptive nature is added to Nystrom's attention. It means an individual subset of landmarks namely local, global, spatial and channel-wise features are selected to attend to rather than attending all tokens in the sequence.

Here's a high-level overview of how the proposed attention mechanism works:

(A) Selecting local landmarks Channel-wise and spatially: In this step using equation (6.16) local landmarks are selected channel-wise and spatially using the Nystrom attention technique. For each channel c and spatial position i, sample local landmarks are given by:

$$X_{l,c,i} = x_{1,l,c,i}, x_{l,2,c,i}, \ldots \ldots x_{l,k,c,i} \tag{6.16}$$

Where k is the number of local landmarks, and $X_{l,k,c,i}$ represents the K-th local landmark in channel c and spatial position i.

(B) Selecting global landmarks channel-wise and spatially: In this step by employing equation (6.17) global landmarks are selected channel-wise and spatially. For each channel c and spatial position, I use a deterministic strategy to select global tokens as:

$$X_{g,c,i} = x_{g,1,c,i}, x_{g,2,c,i}, \ldots \ldots x_{g,m,c,i} \tag{6.17}$$

where m is the number of global tokens, and $x_{g,m,c,i}$ represents the m-th global token in channel c and spatial position i.

(C) Compute Affinity Matrix channel-wise and spatially: Computing the Affinity matrix is a very crucial and novel step in the proposed architecture because after extracting the local and features both channel-wise and spatially these extracted points should be analysed to find the pairwise relationships or similarities between elements in the input sequence, determining the attention weights assigned to each element when computing the weighted sum in the attention mechanism. For each channel c and spatial position i, the affinity matrix $A_{c,i}$ is computed between local and global landmarks using the equation (6.18):

$$A_{c,j,ij} = \text{sim}(x_{l,i,c,i}, x_{g,j,c,i}) \tag{6.18}$$

In equation (23) i is the index for local landmarks and j is the index for global landmarks.

(D) Normalize Affinity matrix: normalizing the affinity matrix ensures that the attention weights assigned to each element are interpretable and comparable. It transforms the raw similarities into a distribution, where each element's attention weight is proportional to its relative importance in the context of the entire sequence. For each channel c and spatial position I, β is the parameter used to control the sharpness by effectively scaling the affinities before

normalizing the affinity matrix $A_{c,i}$ to make it a valid attention distribution by using the equation (6.19):

$$\text{Norm}(A_{c,i}) = \text{softmax}(\beta A_{c,i}) \qquad (6.19)$$

The attention scores $Y_{c,i}$ are calculated channel-wise and spatially by employing equation (6.20) for each channel c and spatial position I, combine the local and global features using the normalized affinity matrix:

$$Y_{c,i} = \Sigma_i \text{Norm}(A_{c,i,ij}x_{g,j,c,i}) \qquad (6.20)$$

In the next step, all the obtained features channel-wise- local, global and spatial- local, global all are concatenated together and are given as input to the next layer. The main characteristic of the proposed multi-scale adaptive Nystrom attention mechanism is to automatically adjust the parameters namely the number of local landmarks (k), the number of global tokens (m), and the scaling factor (β) by adapting to the specific task and dataset characteristics by following the above formulas.

Now when both classifier 1 and classifier 2 showed declined performance, then the feature vectors extracted from classifier 1 and classifier 2 are taken and are given as input to the neural network. Before giving them as input to the neural network the features extracted from both the classifiers should be concatenated. In our framework, a multi-fusion model is utilized for concatenating the features extracted from classifier 1 and classifier 2. This multi-fusion model's primary objective is to combine the features of multiple models in order to identify the unique inherent features, preventing the hybrid model from overfitting due to unbalanced data. The features extracted by classifier 1 are represented as $model_1$ and the features extracted by classifier 2 are represented as $model_2$. Figure 6.9 visualizes the step-by-step procedure followed while creating a hybrid model.
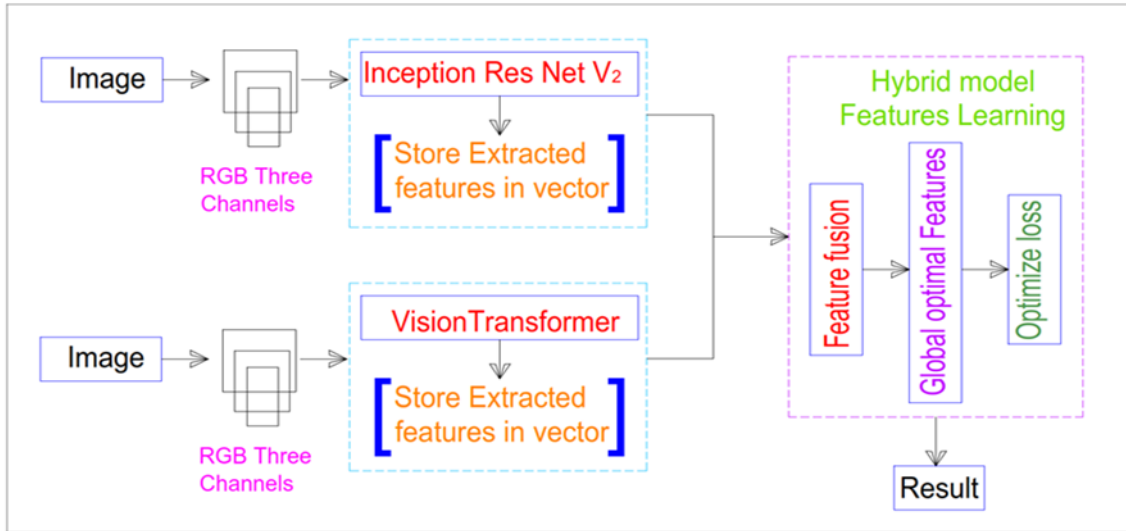
Figure 6.9 visualizes the steps involved in multi-model fusion architecture

To concatenate the features obtained from both models we used the following technique specified in Equations (6.21) and (6.22):

$$\text{Result}= \alpha. \text{ classifier}_1 + \mu. \text{ classifier}_2 \qquad (6.21)$$

$$\alpha+\mu=1 \qquad (6.22)$$

In equation (6.21) and (6.22) α represents the weights of classifier 1 and μ represents the weights of classifier 2.

In the next step, the best combination of alpha and beta values is found on the validation set. In simpler words from the extracted features, only those features whose alpha and beta values result in less minimum loss when calculated using cross-entropy loss are taken. This step is almost as crucial as extracting valuable features from the dataset. The mathematical formulation of this step is shown in equation (6.23), (6.24). To find out the valuable features from the concatenated features the following distance-based optimal search algorithm is utilized:

$$\text{Min (Loss (Result, Label))} \qquad (6.23)$$

$$\alpha+\mu=1 \qquad (6.24)$$

The reason for using a distance based optimal search algorithm is because of its ability to solve combinational optimization problems. As the main aim of our research is to optimize the model's performance the distance-based optimal search algorithms worked well in achieving this. Many distance metrics can be employed for calculating the optimal features among which the Euclidean distance vector is most famously employed. But Euclidean distance works in a specific space and it cannot handle random real values properly. As our problem is associated with working on features extracted from the real world instead of frequently used Euclidean distance we employed chebyshev distance-based optimal search approach.

The efficacy of the proposed CDBA method relies on the definition of the optimal value of the objective function that identifies the ideal values of the features involved. These optimal values related to the respective feature set are defined over a vector BST (v1, v2, v3, …, vn). The BST vector is the set of ideal values of attributes present within the range of the provided dataset. This vector set BST acts as a point of reference that is used to evaluate the effectiveness of the proposed alternative to obtain optimal objective function. This reduces the problem statement to identify an alternate solution as close as the defined ideal BST vector. This can be given by the objective function as formulated in equation (6.25):

$$\text{MIN } \delta \text{ \{DIF(x), BST\}} \tag{6.25}$$

Here, DIF describes an alternative value in n-dimensional space and δ describes the distance from the optimal point. After the final features are selected from the concatenated features the selected features are given as input to the model network and the network is trained and then tested for its performance parameters using its testing dataset. In specific our proposed framework optimized the time taken at each step and provided better results in terms of both accuracy and optimization.

6.5 Experimental setup and Results Analysis

The experimental setup and detailed results analysis are presented in the following section:

6.5.1 Experimental setup

Every experimental test is conducted concurrently on two NVIDIA GPUs, specifically A5000s with 24 GB of RAM each. 128 GB RAM system is deployed. Numerous data analysis frameworks, such as Pandas, Numpy, Seaborn, 496 Matplotlib, and Scikit-learn, were employed in the study. The entire framework is executed for ten epochs. 20% of the data were used for testing, while the remaining 80% were used for training. The proposed model was trained on the training data. The suggested model was fine-tuned utilising a variety of parameters, including hyper parameters, to improve classification accuracy and prevent overfitting. Different learning rates were applied, and it was found 506 that the default learning rate of 0.001 gave better results.

6.5.2 Result analysis

In detail results analysis is given in the following section:

6.5.2.1 Quantitate Analysis

DL algorithms have come a long way in the last several years, and some of their versions have been effectively used to address breast cancer classification issues. In this section performance of the models used in the proposed framework is evaluated and analysed. We take into account a broad range of research instances with different measurement indicators, such as F1 score, area under the ROC curve (AUC), accuracy, precision, recall, and receiver operating characteristics (ROC). Firstly, we evaluated the performance of the classifier 1 model using the quantitative evaluation metrics. Table 6.1 shows the quantitative summary of the classifier 1 model on two publically available datasets namely, BreakHis and BACH. The proposed classifier 1 achieved a testing accuracy and precision of 95 and 94, whereas, the recall, F1 score and MCC value are 91, 92 and 88 respectively on BreakHis dataset. The

proposed classifier 1 obtained accuracy, precision, recall, F1-score, and MCC values of 93, 94.87, 92.5, 93.66, and 87.55 on the BACH the data set respectively, according to the same table. Since the achieved accuracy value is low relative to the threshold value of 0.96, we examined the causes of the declining performance. After specific analysis on declined performance of classifier 1 it is understood that due to its very deep nature and dataset's limited size model failed to explore various patterns or characteristic's in the images this reason consistently leads to misclassification resulting in overfitting that lead to poor generalization.

Further second classifier is activated and analysis is performed by evaluating the performance of the classifier 2 model using the quantitative evaluation metrics. Table 6.2 shows the quantitate summary of the classifier 2 models on two publically available datasets namely BreakHis and BACH. The proposed models achieved a testing accuracy and precision of 96 and 96, whereas, recall, F1 score and MCC value of 93, 94 and 92 respectively on the BreakHis dataset. The BACH dataset findings are also summarised from the same table, with the following scores: 96.7, 95.8, 95, 95, and 93 for accuracy, precision, recall, F1 score, and MCC

Table 6.1 Performance analysis of classifier 1 on two publically available breast cancer datasets namely BreakHis and BACH.

Datasets	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	MCC(%)
BreakHis	95	94	91	92	88
BACH	93	94.87	92.5	93.66	87.55

From Tables 6.1 and 6.2, it is evident that compared to the performance of classifier 1 and classifier 2 ViT showed the best result. In comparison to the standard attention mechanism in ViT, we chose to use the multi-scale linear time Nystrom Attention mechanism since it performs well with limited computational resources. This mechanism approximates the full attention matrix with a low-rank approximation, which leads to more efficient computations

156

and faster training. After specific analysis on declined performance of classifier 2 it is understood that due to limited dataset size model suffered from overfitting which lead to poor generalization on unseen data. From the same tables it can be inferred that MCC scores are relatively less providing insights that model suffered overfitting due to ineffective handling of imbalance in the dataset.

Table 6.2 Performance analysis of classifier 2 on two publically available breast cancer datasets namely BreakHis and BACH.

Datasets	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	MCC(%)
BreakHis	96	96	93	94	92
BACH	96.7	95.8	95	95	93

As the achieved accuracy values of both classifier 1 and classifier 2 are less than the threshold value, features extracted from both the classifiers are fused using the novel feature fusion algorithm "CDBA" resulting in a hybrid model. A hybrid model is then trained on selected features obtained from concatenating selected features from classifier 1 and classifier 2 models. Later the test set images are given to the hybrid model, and model is evaluated. From Table 6.3 it can be inferred that the accuracy of the model is improved when compared to the threshold value indicating the superior performance of the hybrid model effectively captured all the local and global features irrespective of the imbalance issue in the dataset. Furthermore, Table 6.3 illustrates that when additional evaluation metrics are examined, it can be observed that there is an increase in the hybrid model's evaluation metric scores when compared to the results obtained by classifiers 1 and 2.

Table 6.3 Performance analysis of hybrid model on two publically available breast cancer datasets namely BreakHis and BACH.

Datasets	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	MCC(%)
BreakHis	99.36	99.19	98.79	98.98	98.52
BACH	99.4	98.7	99.89	99.3	98.8

From Table 6.3, it is evident that the accuracy and precision of the proposed hybrid model are 99.36, 99.19 for BreakHis and 99.4, 98.7 on the BACH dataset respectively which indicates that the model has a high level of precision for both classes. This means that when the model predicts instances as a specific type of tumour, it is highly likely to be correct. This indicates that the model can make accurate predictions for each tumour class. As can be seen the proposed hybrid model performed exceptionally well in categorising various tumour subtypes when compared to classifiers 1 and 2. It demonstrated exceptional recall, precision, and F1 scores in both tumour classes, striking a notable balance between precision and the capacity to recognise pertinent cases. Further insights into the model's performance were made accessible by evaluating the model performance using other metrics which further assisted with optimisation efforts.

Similarly, from Table 6.3 it can be observed that model achieved, recall values of 98.79 and 100% on BreakHis and BACH respectively, which imply that the model exhibits a substantial level of recall across all tumour classes. This indicates the model's proficiency in accurately identifying and capturing a significant majority of instances belonging to each specific tumour class. Hence, it can be inferred that the model demonstrates a remarkable ability to effectively detect instances of binary tumour classes. Furthermore, from the same table, it can be inferred that on both the datasets F1 score approaching 100% denotes that the model attained a commendable equilibrium between precision and recall for both tumor

classes. This shows that the model achieves a heightened level of accuracy while adeptly capturing instances from each tumour class. This implies that the model effectively strikes a balance between precision, represents the correctness of predictions and recall, denoting the ability to identify relevant instances of attacks. Further, MCC score is calculated to compare how effectively models handles class imbalance situation. Compared to classifier 1 and classifier 2 MCC score's, hybrid model MCC score is reported as 98.52 and 98.8 on BreakHis and BACH dataset's respectively providing a strong evidence about how well the hybrid model handled class imbalance situation.
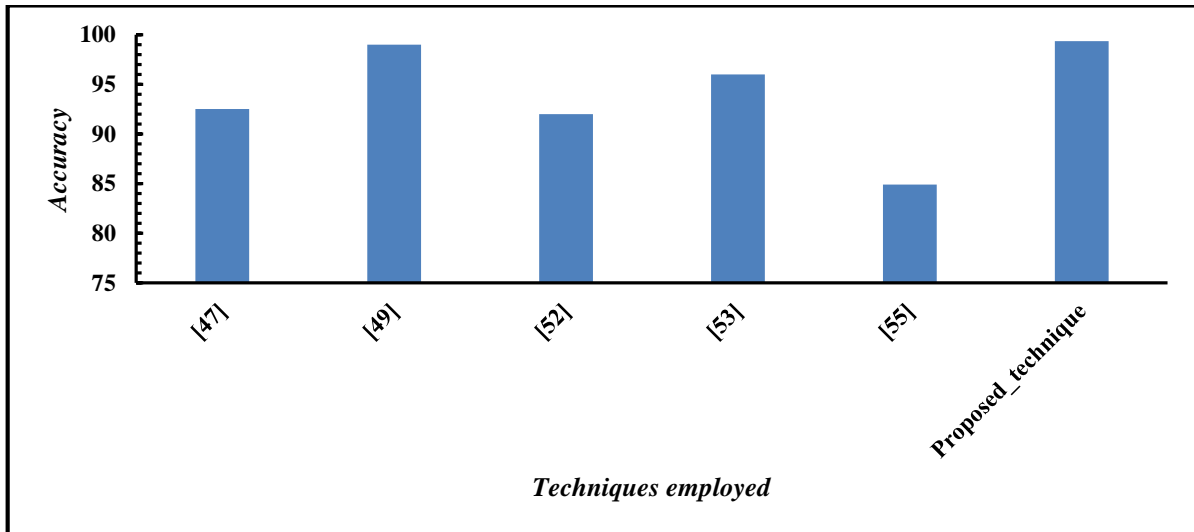

Figure 6.10 Graphical representation of SOTA techniques performance VS proposed model performance on BACH datasets

In this section, the suggested hybrid model's performance for classifying breast cancer is evaluated against the most recent state-of-the-arts. Figure 6.10 and 6.11 Visualizes the graphical representation of SOTA techniques performance VS proposed model performance on BreakHis and BACH datasets.

Figure 6.10 visualizes a comparison of the proposed hybrid model against several recent state-of-the-arts breast cancer classifications on the BACH dataset. With the best

accuracy scores, the proposed hybrid model outperforms the most advanced techniques. Figure 6.11 visualizes and compares the performance of the proposed hybrid model with the state-of-the-art methods using the BreakHis dataset. From Figure 6.11 the superiority of the proposed hybrid model on the BreakHis dataset can be inferred. It performs far better than any of the most recent SOTA approaches to classifying breast cancer.

Figure 6.11 Graphical representation of SOTA techniques performance VS proposed model performance on BreakHis datasets
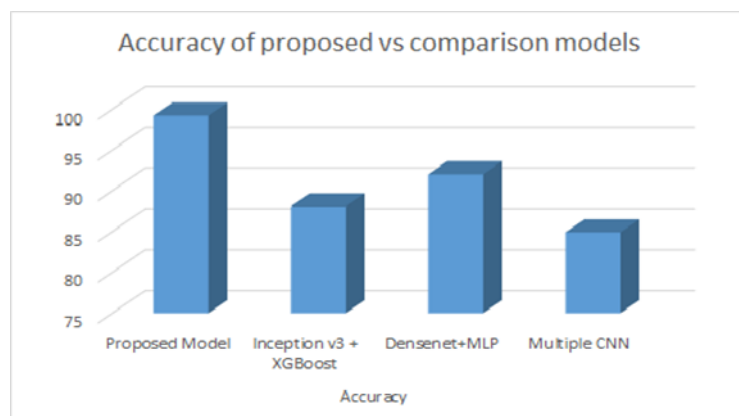
Figure 6.12 Proposed hybrid model accuracy VS state-of-the-art methods in the current environment

Figure 6.12 visualizes the accuracies achieved by state-of-art hybrid techniques employed in the simulation environment same as the environment used for the proposed

framework. From the graph, it is evident that the proposed hybrid model bragged better performance compared to that achieved by the state-of-the-art techniques. The proposed hybrid model achieved an accuracy of 99.36 on BreakHis dataset.

6.6 Complexity Analysis

Overall, the time and space complexity of the model can be quite high, especially for large input data and a large number of trainable parameters. However, these problems can be minimised with the aid of effective algorithms and hardware accelerators like GPUs. Furthermore, methods like early stopping and weight regularisation can assist in reducing the amount of trainable parameters as well as decreasing overfitting, both of these can lower the model's time and space complexity.
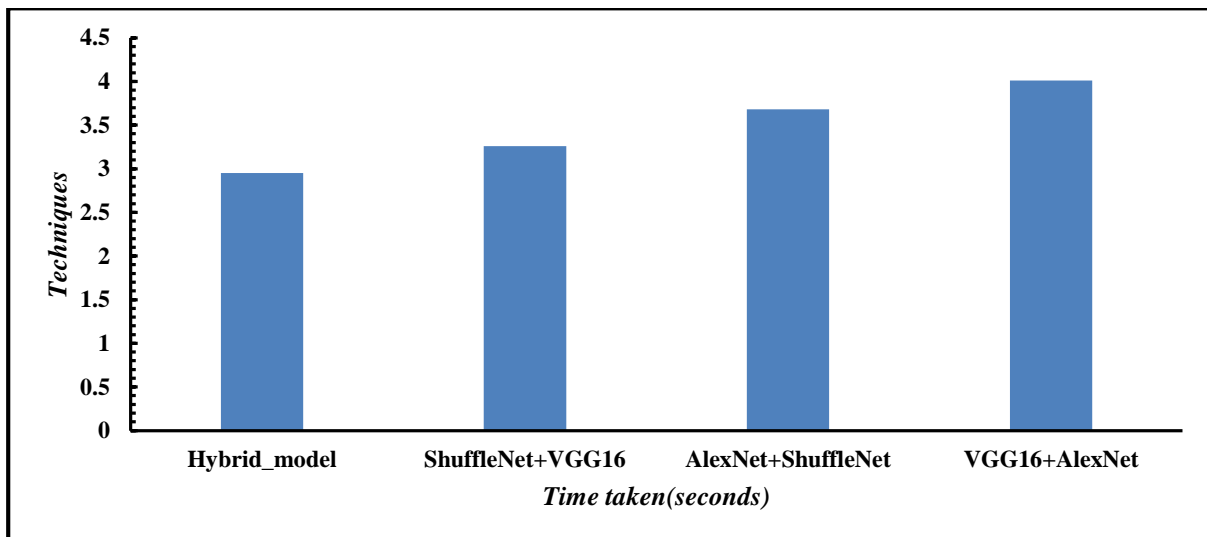
6.6.1 Time Complexity

The primary objective of our framework is not only to integrate the efficiency of the proposed model but also to enhance the computational efficiency of the model. The computational efficiency in terms of time complexity of state-of-the-art methods is compared with the proposed framework in Figure 6.13 on BreakHis dataset. On the x-axis, we have the time taken in seconds and on the y-axis we have the classifiers on the techniques used. The figure provides insights into the efficiency of our proposed hybrid model, which took 3.19 seconds of time per image time, making it the most computationally efficient option when compared with state-of-the-art hybrid techniques. Meanwhile, the state-of-the-art techniques took 4.82,4.23 and 5.96 seconds per image making them computationally inefficient.
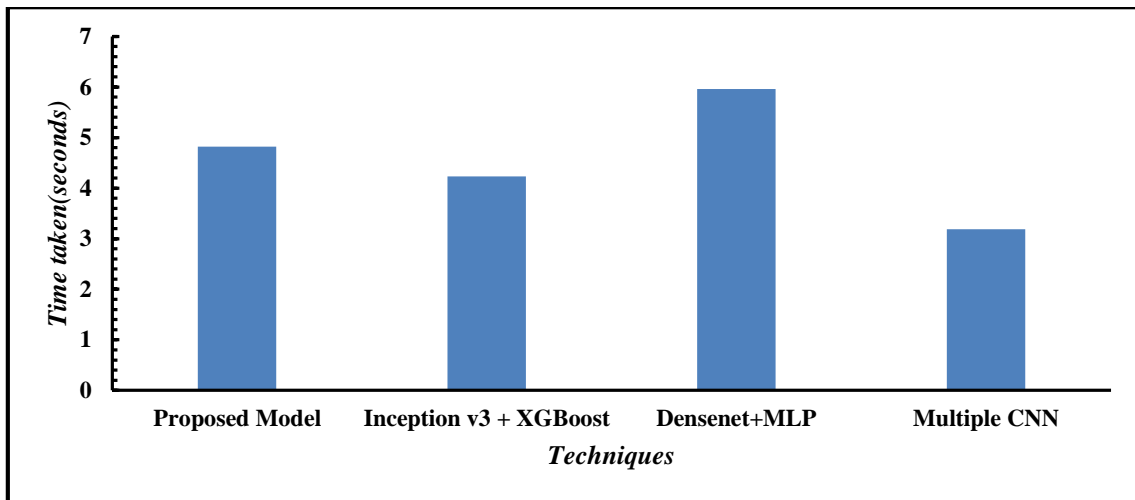
Figure 6.13 Time taken per image in seconds for the hybrid model.

Similarly, Figure 6.14 provides insights into the efficiency of our proposed hybrid model, which required 2.95 seconds per image time, making it the most computationally efficient option when compared with state-of-the-art hybrid techniques on BACH dataset. On the x-axis we have the classifiers or the techniques used and the time taken per image in seconds on y-axis. Whereas, the state-of-the-art techniques took 4.012, 3.68 and 3.26 seconds per image making them computationally inefficient.

Figure 6.14 Time taken per image in seconds for the hybrid model.

6.6.2 Space Complexity

The space complexity of the proposed model is proportional to the quantity of trainable parameters within the proposed model. This is because each trainable parameter requires a certain amount of memory to store its value, and the total memory required to store all the trainable parameters is proportional to their number. In addition to the trainable parameters, the model also requires memory to store the input data, intermediate activations, and gradients during training. The memory required for these operations is proportional to the size of the input data, the number of units in the model, and the number of training steps required to train the model. Any DL model takes millions of trainable parameters due to which space complexity is increased for large-scale models and datasets. The main objective of the proposed model is to improve the computational efficiency of the model which is achieved by employing two techniques in the model that are:

Usage of Multi scale Nystrom attention mechanism in the Transformer.: By employing the multi-scale Nystrom attention mechanism instead of the default self-attention module the overall computational efficiency of the model is improved.

Employment of CDBA for multi-model fusion: Instead of taking all the features extracted by both the classifiers and fusing them, the novel CDBA technique fuses only those features that are contributing to the model efficiency.

6.6.3 Scalability Performance

Finally, we have observed enhanced scalability characteristics in our proposed model. Remarkably, as we increased the epoch number from 10 to 50 and it is represented on x-axis in the range 1-5, the accuracy of our proposed model remained nearly unchanged, indicating its scalability. Figure 6.15 visualizes the scalability performance of our suggested model. Figure 6.15 it can be observed that even by changing the epochs also 0.99 is the constant accuracy of the model showing its efficiency.
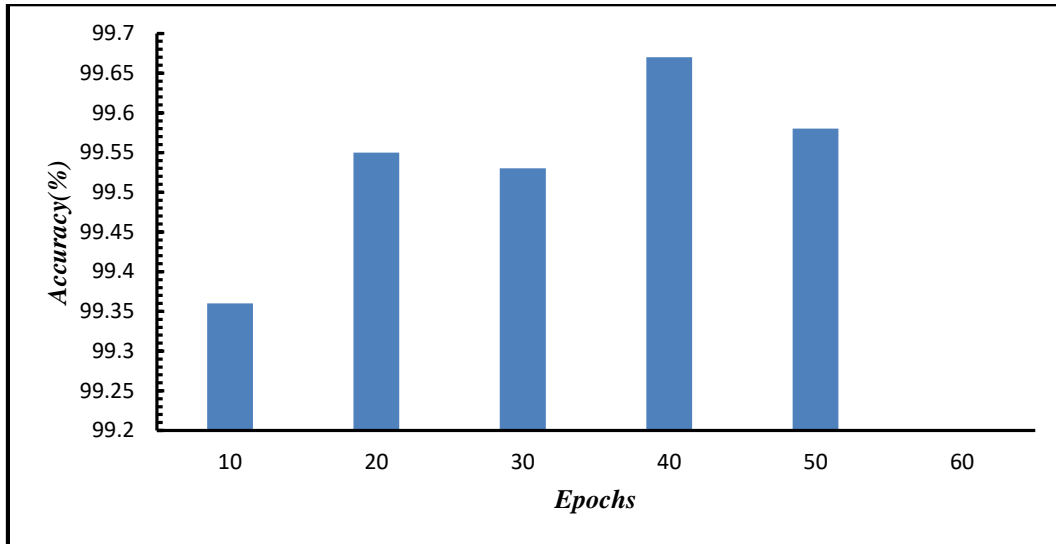
**Figure 6.15 Scalability analysis of our proposed model**

6.7 Ablation study

This section includes an ablation investigation to validate the efficiency of the proposed framework. Specifically, the framework is trained on the BreakHis and BACH datasets. The following cases are evaluated:

• **Case A:** The model is trained exclusively on the InceptioResNetV2 baseline architecture.

• **Case B**: The model is trained only on the baseline architecture of ViT.

• **Case C:** The model is trained with the proposed architectures of InceptioResNetV2.

• **Case D:** The model is trained only proposed architecture of ViT.

• **Case E:** The model is trained without framework by training it on the hybrid model.

• **Case F:** The model is trained with the framework without utilizing the CDBA technique for fusing the features.

• **Case G:** (proposed model): The model is trained with a proposed framework by integrating proposed architectures of classifier, classifier 2, multi-head linear time attention layer, and CDBA for feature fusion to the framework.
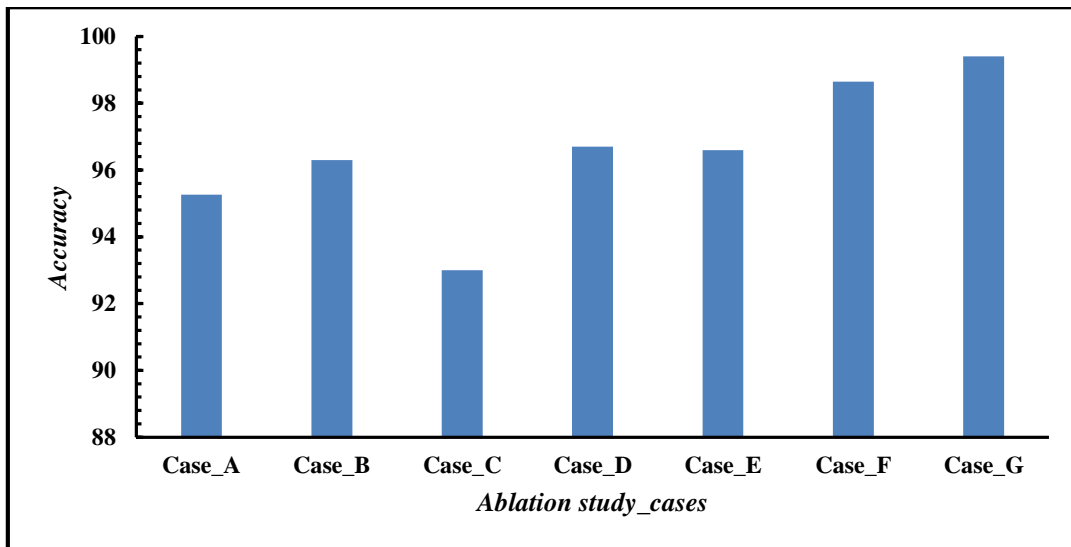
Figure 6.16 Comparison of accuracies of Ablation study cases on BreakHis dataset.
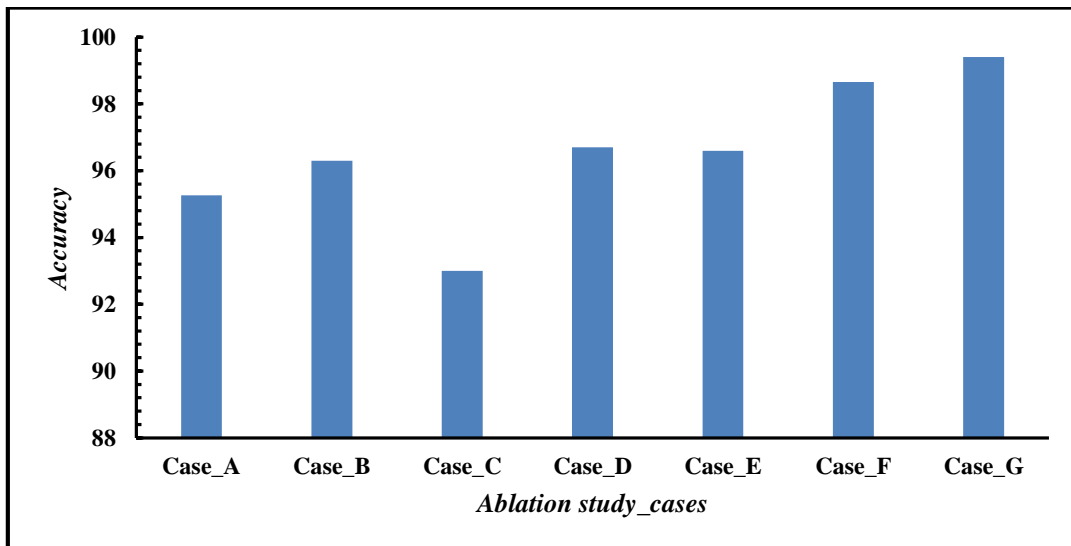
Figure 6.17 Comparison of accuracies of Ablation study cases on BACH dataset.

Figures 6.16, and 6.17 visualize the performance analysis of different Ablation study cases on BreakHis and BACH datasets from the figures it can be observed that the performance of the models improved with the novelty added to the framework. Table 6.4 Shows the time taken by classifier 1, classifier 2 and hybrid classifier (Case C, D, G). It is apparent from the table that classifier 1 completed its task in 1.65 seconds, while classifier 2 took 1.52 seconds. In contrast to the default attention mechanism, the modified attention mechanism in the

proposed classifier 2 required 1.35 seconds per image, showcasing its improved computational efficiency.

Table 6.4 Presents the computational efficiency of techniques employed in the framework.

Classifiers	Time taken per image in seconds
Classifier 1 - Case C	1.65
Classifier 2 - Case D	1.52
Hybrid - Case G	1.35

6.8 Chapter Summary

This chapter introduces a deep learning framework designed for effective image classification across complex real world datasets. The proposed hybrid approach combines feature from multiple classifiers to improve adaptability and performance. Key innovations include an efficient attention mechanism that prioritizes critical features, along with selective optimization to balance computational efficiency with high accuracy. Tested on datasets such as BreakHis and BACH, this framework demonstrates strong scalability, accuracy, and adaptability, establishing itself as a valuable tool for research use. In addition to achieving high classification accuracy, this framework contributes significantly to the field by integrating resource-efficient optimization methods.

Chapter Seven: Hybrid Deep Learning Approach for Generalized Image Classification: A Focus on Robustness and Efficiency.

In this chapter a novel Hybrid Deep Learning (DL) model that addresses Generalisation problem in imbalance dataset, a common challenge affecting model performance is proposed and results were evaluated.

7.1 Introduction

Chapter 7 discusses the critical need for generalized models in automated classification, especially in scenarios where complex datasets and high-stakes decision-making are involved. Despite advances in technology, various applications still face challenges in efficient and accurate data analysis due to the diversity and intricacy of data structures and underlying patterns. Effective classification in these scenarios relies not only on precise algorithms but also on robust frameworks that can adapt to diverse datasets and minimize the risk of overfitting or bias. This chapter highlights the development of adaptable, high-performance models that can generalize effectively across varying dataset complexities. By introducing advanced deep learning techniques, the framework leverages automated feature extraction and classification, achieving enhanced accuracy and computational efficiency. Such models, which can optimize feature selection and integrate heterogeneity in data, contribute to stronger predictive performance. These advancements underscore the importance of generalized classification models that can handle broad data variations, providing a valuable foundation for further innovation across multiple application domains.

7.1.1 Key contributions

The experimental setup and detailed results analysis are presented in the following section:

(i) Development of a Novel Hybrid DL Model: In this work, a hybrid DL model is proposed, specifically designed for both binary and multi-class brain tumor classification using

MRI. The model introduces advanced techniques to address class imbalance, overfitting, and image occlusion, significantly enhancing the robustness and generalization of the model.

(ii) Advanced Noise Reduction and Occlusion Handling: The proposed model employs Gaussian filters for noise reduction and incorporates a new Patterned-GridMask technique, an enhanced variant of the standard GridMask. This innovation improves the model's ability to handle occluded images, contributing to its superior performance across a range Brain tumor dataset (Binary and Multi class datasets.).

(iii) Architectural Enhancements: In this work, a novel approach is introduced to enhance the Multi-Axis Vision Transformer (MaxViT), focusing on improving computational efficiency and generalization. These modifications are designed to optimize the model's performance, allowing for more accurate and faster classifications in both binary and multi-class tasks.

(iv) Superior Model Performance: A comprehensive comparative evaluation against existing deep learning models demonstrates the superior performance of the proposed hybrid DL model in both binary and multi-class classification tasks. These results highlight the model's robustness, adaptability, and potential for clinical applications in brain tumor diagnosis.

Methodological innovations encompassing hybrid model design, sophisticated feature selection, and addressing classifier heterogeneity, collectively propel brain tumor classification to the forefront, optimizing accuracy and generalization.

7.2 Preliminaries

In section 7.2, the Hybrid model designed to address critical data imbalance and Generalization challenges in Medical image datasets is explained. The section is divided into subsection, each focusing on a specific aspect of the proposed methodology. It begins with a

168

detailed explanation on Datasets employed, followed by outlining of the techniques employed followed by in detail explanation of the proposed model.

7.2.1 MaxViT Architecture

The MaxViT architecture represents an advanced hybrid model that integrates the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to effectively capture both local and global dependencies in image data. Its architecture features hierarchical blocks that utilize both convolutional layers and transformer-based attention mechanisms. This unique setup allows MaxViT to excel in various vision tasks, including image classification, object detection, and segmentation, achieving a notable balance between accuracy and computational efficiency.

(i) MBConv Layer: Efficient Local Feature Extraction

MaxViT's efficiency is improved by its MBConv layer, influenced by the Mobile Inverted Bottleneck Convolution. This layer employs depth wise separable convolutions to effectively extract local spatial information while minimizing computational costs. In depth wise convolution, the operation is conducted independently on each input channel employing equation (7.1):

$$Y^{(c)}(i,j) = \sum_{m,n} X^{(c)}(i+m, j+n) \cdot K^{(c)}(m,n) \tag{7.1}$$

Here, X is the input tensor and K is the convolution kernel for channel c. This approach efficiently extracts local features without mixing information across channels. Subsequently, a pointwise convolution is applied through a 1×1 convolution to project features across channels using equation (7.2):

$$Z(i,j) = \sum i,j \tag{7.2}$$

This combination enables the model to capture local features effectively while maintaining operational efficiency.

(ii) Multi-Head Self-Attention (MHSA): Capturing Global Dependencies to capture global dependencies, MaxViT incorporates Multi-Head Self-Attention (MHSA). In this mechanism, the input sequence X_{seq} is transformed into query (Q), key (K), and value (V) vectors employing equation (7.3):

$$Q = X \tag{7.3}$$

The attention mechanism computes relationships between image patches using the following equation (7.4):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d}})V \tag{7.4}$$

This allows the model to focus on the most relevant image regions, facilitating the learning of long-range dependencies.

(iii) Max Pooling Attention: Enhancing Global Attention

To enhance global attention while reducing computational costs, MaxViT introduces Max Pooling Attention (MPA), which decreases spatial dimensions before applying self-attention. This approach significantly mitigates computational requirements while preserving critical information:

The output is subsequently up sampled to match the original resolution, allowing MaxViT to efficiently capture global context.

(iv) Axial Attention: Reducing Complexity in Spatial Dimensions

MaxViT also employs axial attention, which applies attention separately along each spatial axis, reducing the complexity associated with full attention across all dimensions. For height attention, the representation in mathematical terms is given in equation (7.5):

$$X_{height} \in R^{H \times (W \cdot C)} \tag{7.5}$$

This is similarly executed for width attention, leading to lower computational complexity while maintaining essential spatial relationships.

(v) Local-Global Attention: Combining Short and Long-Range Dependencies

MaxViT's hybrid attention scheme incorporates both local and global attention mechanisms. Local attention focuses on short-range interactions within patches. By integrating both methods, MaxViT effectively captures multi-scale dependencies, making it highly effective for complex vision tasks that require both detailed analysis and global context.

(vi) Feed-Forward Network (FFN): Non-Linear Transformations

Following the attention mechanisms, a Feed-Forward Network (FFN) as formulated in equation (7.6) is employed to apply non-linear transformations, enhancing the model's expressive capability:

$$FFN(X) = GELU(XW_1 + b_1)W_2 + b_2 \qquad (7.6)$$

This non-linearity adds complexity, allowing MaxViT to perform more sophisticated transformations on the learned features.

7.2.2 Patterned GridMask

Patterned-GRIDMASK is a data augmentation technique that enhances model generalization by applying structured masking patterns to input images. Originating from the GRIDMASK concept, which involves randomly masking parts of an image in grid-like manner, Patterned-GRIDMASK employs a more organized approach. This strategic masking allows the model to concentrate on learning generalized features rather than memorizing specific details. By utilizing predefined or dynamically generated patterns for masking, this technique can be employed across various neural network architectures, including CNNs and Vision Transformers, to mitigate overfitting and improve performance on unseen data.

The core idea behind Patterned-GridMask involves applying a structured grid mask to input data. The basic grid mask is created by dividing the image into a grid and selectively removing sections according to a predefined strategy. For an input image represented as I ∈ R^(H×W×C), where H, W, and C signify the height, width, and channels of the image

respectively, the binary mask M∈R^(H×C) is generated, with regions set to 0 (masked) and others set to 1 (preserved). The application of the mask can be formulated in equation (7.7) as:

$$I' = I \circ M \qquad (7.7)$$

Here, ∘ denotes element-wise multiplication, resulting in a masked image I'∈R^(H×W×C) where specified portions of the image are removed.

The grid mask M is generated by dividing the image into blocks with a chosen block size d. The percentage of the image that is occluded is denoted as p. The mask generation formula can be described in equation (7.8) as follows:

$$M(x,y) = \{0, if\ (x\ mod\ d < d \times p), or, (y\ mod\ d < d \times p)\ 1, otherwise \qquad (7.8)$$

While traditional GRIDMASK utilizes a uniform grid, Patterned-GRIDMASK introduces additional complexity by employing various patterns to determine the masked sections. These patterns can be predefined geometric shapes or dynamically generated based on training rules, thus enriching the training experience for the model. For instance, a diagonal masking pattern can be mathematically represented by employing equation (7.9):

$$M(x,y) = \{0, if\ x - y = k, (where\ k\ is\ a\ constant)\ 1, otherwise \qquad (7.9)$$

The augmentation process comprises several layers. Initially, the input image tensor I ∈ R^(H×W×C) is loaded into the system. The mask generation layer follows, which creates the binary mask based on selected patterns. Following this, the mask is applied to the input image via element-wise multiplication employing equation (7.10):

$$I'(x,y,c) = I(x,y,c) = M(x,y) \qquad (7.10)$$

In the subsequent augmentation layer, the masked image serves as augmented input for model training, allowing the model to generalize better by concentrating on global features rather than specific details. Finally, the augmented image is processed through the main

network architecture, which may be a CNN or a Vision Transformer, effectively leading to more robust feature representations.

7.3 Proposed Methodology

Accurate and early detection of tumors is crucial for enhancing the effectiveness of treatments. Deep learning (DL) algorithms, particularly those designed for the identification and classification of malignant tumors, have shown significant potential in improving diagnostic accuracy. Researchers frequently experiment with various architectures to find the optimal model that fits their specific datasets and challenges. Customization is often required to tailor these models to specific datasets, but developing a new architecture for each tumor type does not always guarantee consistent performance across diverse datasets. To address this, we propose a novel transformer-based architecture designed for tumor classification across multiple datasets, specifically targeting brain tumors. Figure 7.1 shows the work flow of proposed model and Figure 7.2 visualizes the GRN_MLP module.
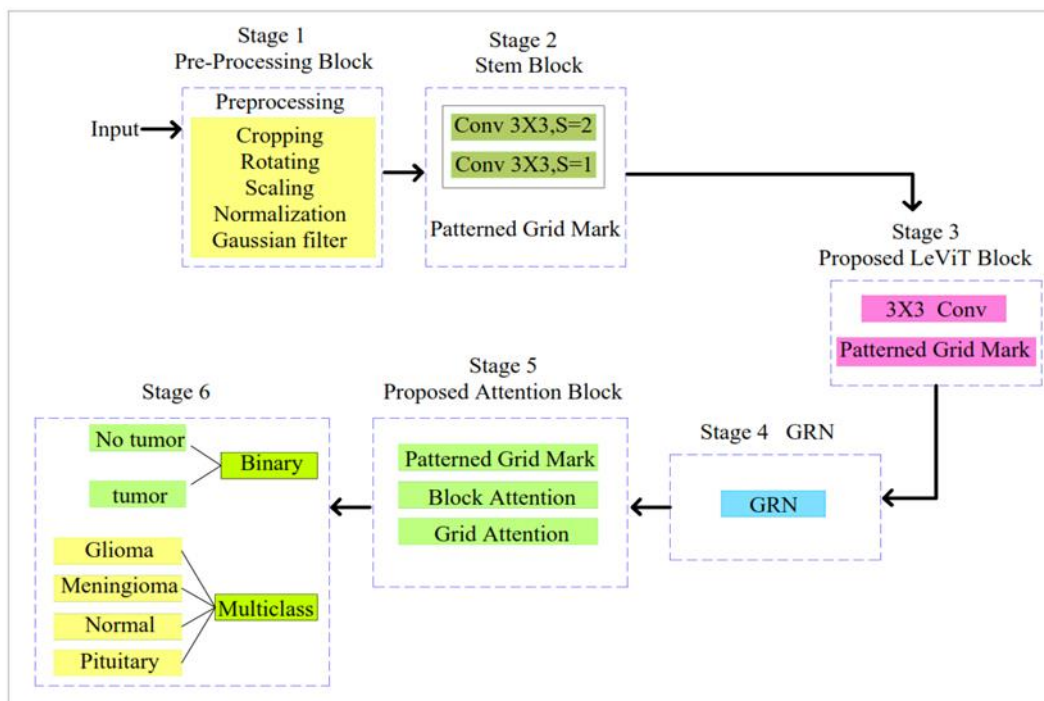
Fig 7.1 Proposed Methodology
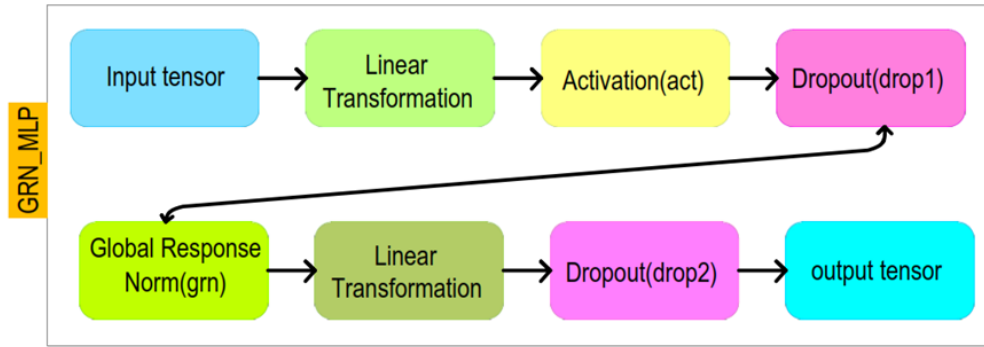
Figure 7.2 GRN_MLP

The proposed model, tailored for tumor detection, incorporates several efficient blocks to enhance feature extraction from imbalanced datasets. Key novelty of the proposed model is the integration of the Patterned Grid Mask at various blocks, aiding the network focus on important regions of the image by deleting irrelevant information. Below is a detailed explanation of the proposed model, and its components. The proposed model contains 5 blocks

Stage 1: The pre-processing block

Initially, various pre-processing techniques are applied to standardize and to enhance the input data. These pre-processing steps help in effective feature extraction. Various pre-processing techniques employed are Cropping, Rotation, Scaling, Normalization and Gaussian filters.

- *Cropping:* This technique helps in removing irrelevant information of the images like background noise by focussing on the region of interest.

- *Rotation:* Images are rotated in random orientations to ensure that model identifies important features regardless of the orientation of the captured image.

- *Scaling:* By scaling the input images, all the images are brought to a uniform size ensuring that all images are processed at the same resolution.

- *Normalization:* This technique adjusts the pixel values by scaling them in the range 0-1 to ensure that models performance is not affected by varying brightness of the image.

174

Stage 2: Stem Block

In stem block down-sampling technique is applied on the pre-processed images to reduce the spatial resolutions of the images. Down-sampling the image resolution helps in reducing the image complexity, increasing the efficiency of the further layers of the model while learning useful features. Two convolutional layers are used where, the first layers reduce the spatial dimensions by half and the second layer helps in maintaining these dimensions while refining the feature map. The above are formulated employing equations (7.11), (7.12) and (7.13).

First convolution layer: Kernel size- 3*3, stride-2.

$$Y_1 = Conv(X_{preprocessed}, W_1) + b_1 \tag{7.11}$$

Second convolution- Kernel size-3*3, stride-1.

$$Y_2 = Conv(Y_1, W_2) + b_2 \tag{7.12}$$

After the second convolution, the patterned Grid Mask is applied.

$$Y_3 = Y_2 \odot M \tag{7.13}$$

Here the Patterned Grid Mask ensures that the stem block focuses on most important regions of the image, while ignoring less important regions. This helps in efficient feature extraction and efficient usage of computational resources. The obtained masked feature map from the stem block is given as input to the LeViT Block.

Stage 3: LeViT Block

After obtaining the masked feature map from the Stem Block, the LeViT Block applies convolution operation to capture local patterns from the input. As LeviT model is a combination of CNN and ViT both local and global features are extracted effectively.

Initially a 3*3 convolution layer is applied to capture local patterns using the following equation (7.14):

$$Y_{conv\_levit} = Conv(X_{levit}, W_{levit}) + b_{levit} \qquad (7.14)$$

After performing the convolution Patterned Grid Mask is applied to guide model's attention towards the important features for enhanced feature extraction. By doing so, the model reduces the noise in the data and ensures that the transformer focuses on important spatial relationships comparatively.

$$Y_{LeViT\_masked} = Y_{conv\_levit} \odot M \qquad (7.15)$$

Stage 4: Global Response Normalization:

After applying the patterned grid mask has been applied, the GRN is incorporated to normalize the activations of each feature map across the channels. GRN establishes an effective balance among activations ensuring that no single channel dominates the feature representation. GRN works by normalizing the activations in each channel by computing the sum of the squared activations across all channels and by adding a small constant ϵ to prevent division by zero. The normalization of the activation ensures that all channels contribute equally to the next stage of the model, improving generalization and reducing the risk of overfitting. The formula for normalization is given in equation number (7.16).

$$GRN(x_i) = \frac{x_i}{\sqrt{\Sigma_{j=1}^{N} x_j^2 + \epsilon}} \qquad (7.16)$$

Stage 5: Attention Mechanism

The feature map obtained from the LeViT block is a feature map which contains both local and global spatial features. This feature map is pass as input to the Max Self Attention block. This data transition is important because it makes sure that the Max SA module receives input that has already undergone a hybrid local, global feature extraction. The above is formulated employing equation (7.17)

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \qquad (7.17)$$

The employed attention mechanism computes the relationship between various parts of the input using the formula:

Max-SA operates by splitting into two parts a shown in equations (7.18) and (7.19):

Block Attention block: Focuses on local features.

$$Attention_B(Q_B, K_B, V_B) = softmax(\frac{Q_B K_B^T}{\sqrt{d_k}}) V_B \odot M \qquad (7.18)$$

Grid Attention block: Focuses on global features.

$$Attention_{grid}(Q_{grid}, K_{grid}, V_{grid}) = softmax(\frac{Q_{grid} K_{grid}^T}{\sqrt{d_k}}) V_{grid} \odot M \qquad (7.19)$$

Before applying block attention and grid attention, the patterned grid mask is applied to make sure that the attention mechanism focuses on relevant parts of the image.

Block Attention: By employing block attention mechanism the made is made to focus on fine-grained details of the image. By applying this features and patters from small regions of the image containing critical information can be understood by the model.

Grid Attention: By employing Grid attention mechanism the made is made to focus on capturing global dependencies of the image. By applying this features and patterns from whole image containing critical information can be understood by the model.

After applying Max-SA and GRN, model has identified both local and global features and patterns, refined through various attention mechanisms and normalized channels. The obtained features contain both low-level details and high-level structures, making them discriminative for tumor detection. So, in the final layer classification of the tumor is done from the most focused and refined set of features as shown in equation (7.20).

$$Y_{GRN\_MLP} = Linear(GRN(X_{MLP})) \qquad (7.20)$$

7.4 Experimental Setup and Result Analysis

The experimental setup and detailed results analysis are presented in the following section:

7.4.1 Experimental setup

Every experimental test is conducted concurrently on two NVIDIA GPUs, specifically A5000s with 24 GB of RAM each. 128 GB RAM system is deployed. Numerous data analysis frameworks, such as Pandas, Numpy, Seaborn, 496 Matplotlib, and Scikit-learn, were employed in the study. The entire framework is executed for ten epochs. 20% of the data were used for testing, while the remaining 80% were used for training. The proposed model was trained on the training data. The suggested model was fine-tuned utilising a variety of parameters, including hyper parameters, to improve classification accuracy and prevent overfitting. Different learning rates were applied, and it was found 506 that the default learning rate of 0.001 gave better results.

7.4.2 Results Analysis

In this study, the proposed model's ability to generalise on various datasets is analysed. The proposed model is deployed for both binary classification and multi classification. Main aim of this research is to perform experiments and prove from the results achieved that the proposed model owing to its efficient performance, and good generalisation capacity it can be commonly used as a baseline deep learning model for brain tumor classification. Additionally, In this research, Gaussian filters were applied to remove noise from brain tumor MRI scans to enhance the performance. These filters differed in strength based on the kernel size and standard deviation. We applied various combinations of parameters to the brain tumor MRI scans to evaluate their impact on performance. Later, patterned GridMask was applied to enhance the generalization performance of the deep learning model and improve its adaptability across various images.

7.4.2.1 Quantitative Analysis

Figure 7.3 provides a virtual representation of proposed model effectiveness across various binary datasets. From the Figure 7.3, it is clear that the model demonstrated exceptional results, showcasing its robustness and consistency. The proposed model achieved the following results, on the Br35h dataset, the model achieved accuracy, precision, recall, F1-Score, and MCC values of 99.63, 99.82, 99.68, 99.68, and 99.01, respectively. On the Kaggle binary dataset, the model reached accuracy, precision, recall, F1-score, and MCC values of 99.85, 99.62, 99.89, 99.58, and 98.00, respectively. For the BT-small 2c dataset, the model obtained values of 99.65, 99.25, 99.28, 99.56, and 98.93 for accuracy, precision, recall, F1-score, and MCC. Finally, on the BT-large 2c dataset, the proposed model achieved accuracy, precision, recall, F1-score, and MCC values of 99.68, 99.68, 99.51, 99.86, and 98.43, respectively. These results indicate strong and consistent performance across multiple binary datasets.
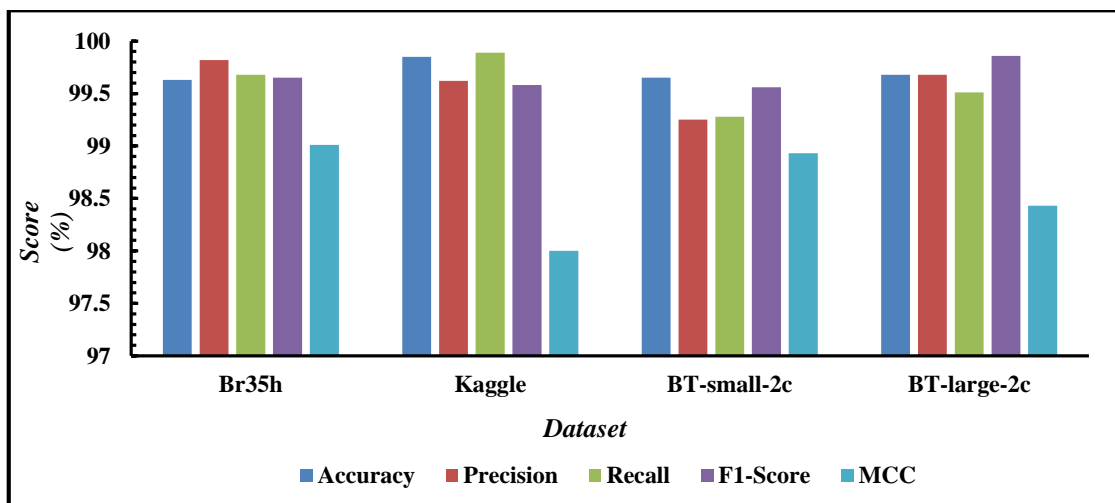
Figure 7.3 Binary classification

Notably, the model's effective handling of class imbalance played a critical role in achieving high accuracy and generalization across diverse datasets. This strong performance across all binary classification tasks highlights the model's ability to generalize well, even in

the presence of imbalanced data, further establishing its reliability for similar tasks. The visual data offers a comprehensive view of the model's capabilities, illustrating its successful handling of class imbalance and its potential for broader applications. The results not only emphasize the model's efficiency but also its adaptability and strength in producing reliable outcomes across multiple datasets proving its generalizability in binary class classification.
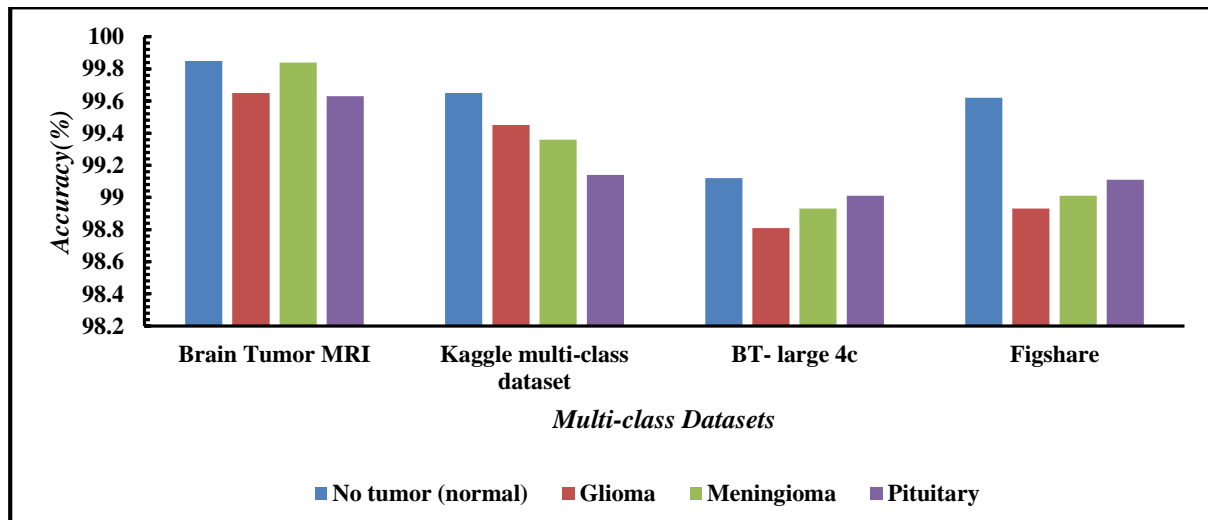
Figure 7.4 Multi class classification

Figure 7.4 visualizes the proposed model performance across different datasets, exhibiting its robustness and effectiveness in multi-class classification. On the Brain Tumor MRI dataset, the proposed model demonstrated exceptional accuracy, achieving 99.85 for the normal tumor class, 99.65 for the Glioma class, 99.84 for the Meningioma class, and 99.63 for the Pituitary class. Similarly, on the Kaggle multi-class dataset, the model achieved accuracies of 99.65, 99.45, 99.36, and 99.14 for the normal tumor, Glioma, Meningioma, and Pituitary classes, respectively.

The proposed model persistently demonstrated efficiency across multiple tumor classes, including No tumor, Glioma, Meningioma and pituitary as assessed on datasets such as the Brain Tumor MRI, Kaggle multi-class, BT-large 4c and Figshare datasets. Notably, the

proposed model achieved outstanding stability while classifying classes with significant imbalance, performing particularly well on the Brain Tumor MRI dataset, nearly followed by the Kaggle, Figshare and BT-large datasets. This stable performance, despite varying imbalance degrees, underscores the model's exceptional generalization capabilities. The visualized results indicate that the model handles complex imbalanced data reliably and achieved outstanding performance in distinguishing between the multi classes. Such consistent, high-level performance across datasets validates the proposed model's adaptability and reinforces its potential as a versatile tool in medical image classification for multi-class scenarios.

From the overall comparison of results obtained by the proposed model in both binary and multi class classification, it can be observed and understood that the proposed model not only achieved exceptional accuracy but also achieved exception results for other performance metrics employed also proving its efficacy and generalization capability on imbalanced datasets irrespective of their imbalance ratios.

The proposed model achieved exceptional performance on the Brain Tumor MRI dataset, accurately classifying normal tissue, Glioma, Meningioma, and Pituitary tumor classes. Similarly, in the Kaggle multi-class dataset, the model maintained high precision, demonstrating its effectiveness in distinguishing between these classes, despite the presence of imbalanced data. The BT-large 4c dataset, with its more complex tumor classes and larger image dimensions, further challenged the model, yet the hybrid approach ensured that it consistently delivered high performance, showcasing its scalability for large datasets. Additionally, the Figshare dataset provided further evidence of the model's reliability, as it handled various tumor types and class imbalances with ease, reinforcing its adaptability to diverse imaging conditions. The combination of visual and tabular data provides a comprehensive view of the model's capabilities, illustrating its ability to generalize effectively

across datasets. The results clearly highlight the model's precision and recall, which are significantly enhanced by the patterned Grid Mask improving tumor localization, ensuring detailed and accurate classification. Gaussian filtering contributes to the model's ability to handle noisy or low-quality images, which is particularly valuable in medical imaging scenarios where dataset quality can vary significantly.

Furthermore, the model's exceptional performance in handling class imbalance proves crucial in achieving reliable results across various tumor types, especially those that are underrepresented in training data. The successful application of this hybrid approach in multi-class classification tasks demonstrates its potential for broader applications beyond brain tumor classification, making it suitable for other complex medical imaging problems. Ultimately, the results not only emphasize the model's high accuracy but also its adaptability and scalability, proving it to be a strong candidate for real-world clinical use. The model's ability to generalize across different datasets, tumor types, and imaging conditions ensures its utility in various medical imaging tasks, where multi-class classification is essential. Its successful implementation of Patterned Grid Mask and Gaussian filtering techniques further solidifies its position as a cutting-edge solution for classifying complex tumor types, addressing class imbalance, and delivering reliable, consistent results.

Figure 7.5 below compares the performance analysis of the proposed model while classifying brain tumors in binary datasets, directly with the latest state-of-the-art techniques in the current environment. The table presents a comparison of the proposed model against recent SOTA methods across several publically available binary classification datasets, where the proposed model consistently achieves higher accuracy scores. The figure also visually highlights the model's performance over SOTA techniques, showing that the hybrid model yields improvements ranging from 1% to 8% across various binary class datasets. These

visualizations make it clear that the proposed model stands out in the current environment, constantly outperforming recent advance models also.
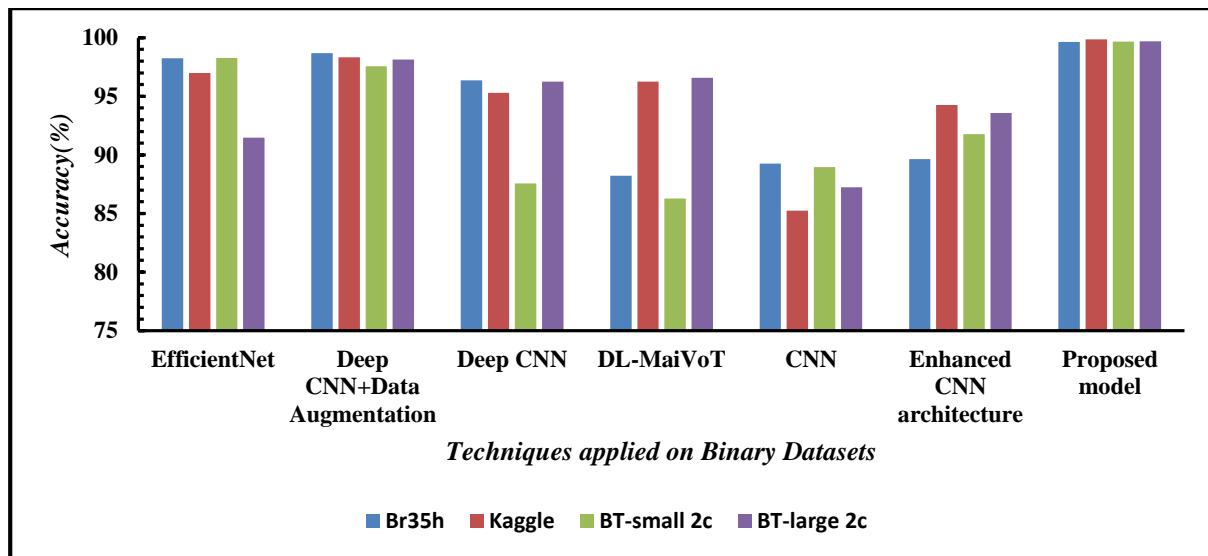
Figure 7.5 Performance of Proposed vs SOTA for Binary Class Classification

The comprehensive evaluation of the proposed model's effectiveness in multi-class classification again the latest SOTA techniques is shown in Figure 7.6 It highlights a detailed comparison across various publicly available datasets, where the proposed model constantly outperformed recent SOTA methods, achieving notable accuracy across all datasets. The visual comparison in Figure 7.6 further analyse the performance gains, explaining the clear advantage of the hybrid model in complex multi-class classification tasks. Notable, the hybrid model's ability to address data complexities infer the performance improvements of the model ranging from 1% to 5% across various datasets, visualizing its adaptability and robust classification capacity. This displayed that the proposed model underscores its potential as a leading solution for brain tumor classification, setting a new benchmark in comparison to current SOTA techniques.
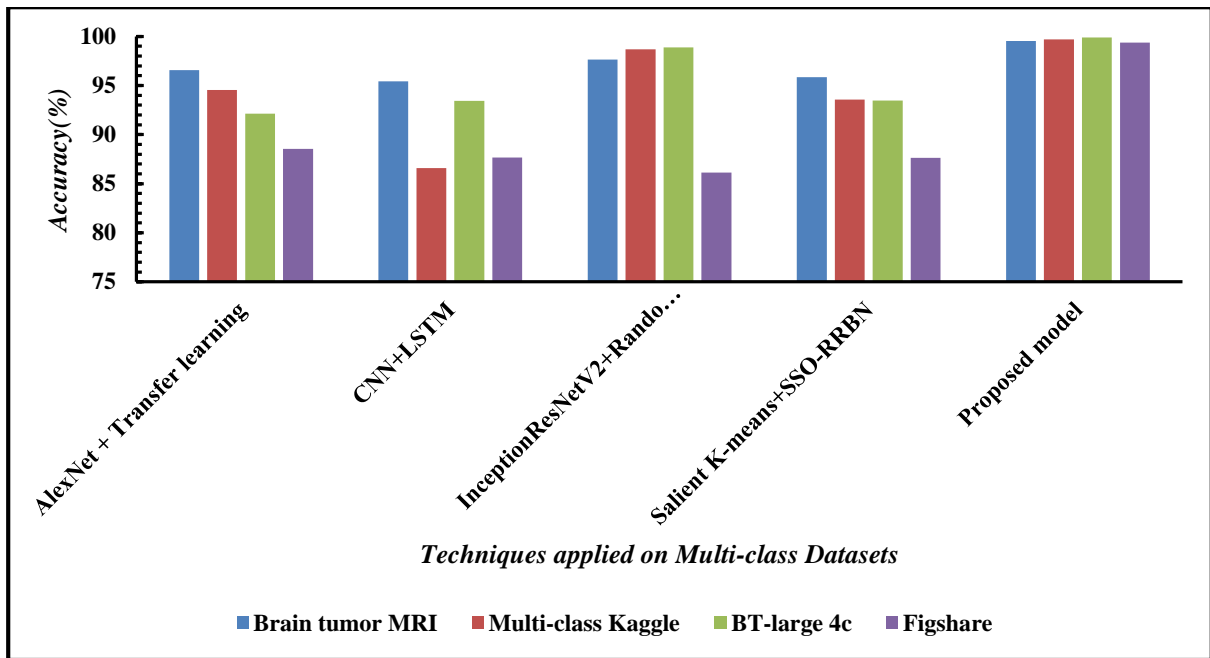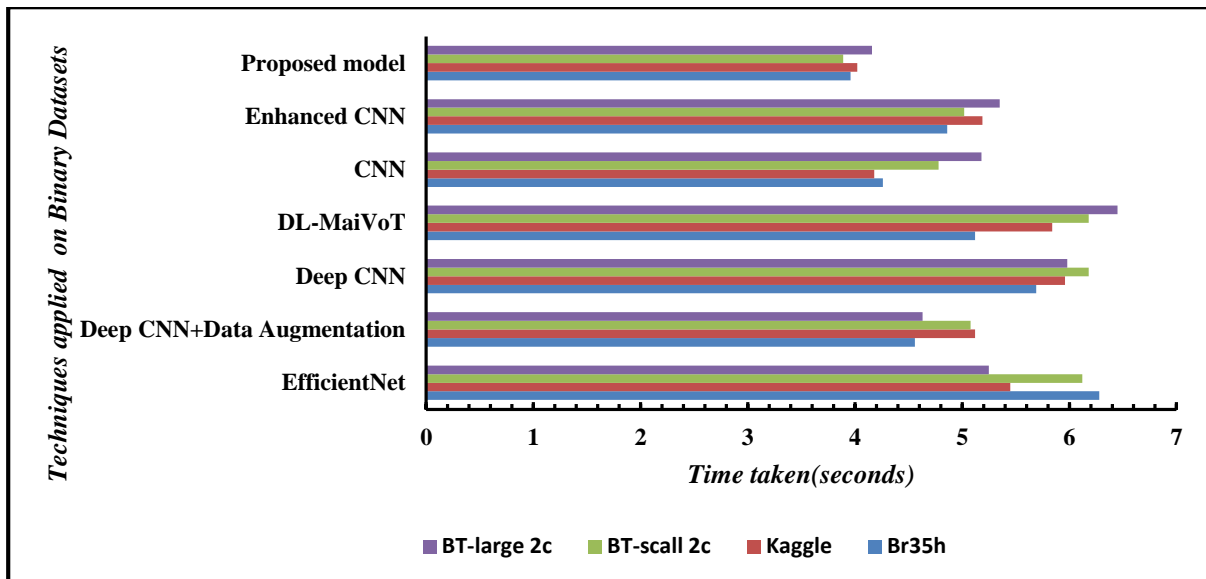
Figure 7.6 Performance of Proposed vs SOTA for Multi Class Classification

7.4.2.2 Complexity Analysis

Overall, the time and space complexity of the model can be quite high, especially for large input data and a large number of trainable parameters. However, these problems can be minimised with the aid of effective algorithms and hardware accelerators like GPUs. Furthermore, methods like early stopping and weight regularisation can assist in reducing the amount of trainable parameters as well as decreasing overfitting, both of these can lower the model's time and space complexity. For fair complexity and generalisation analysis, datasets belonging binary classification and multi- classification were considered.

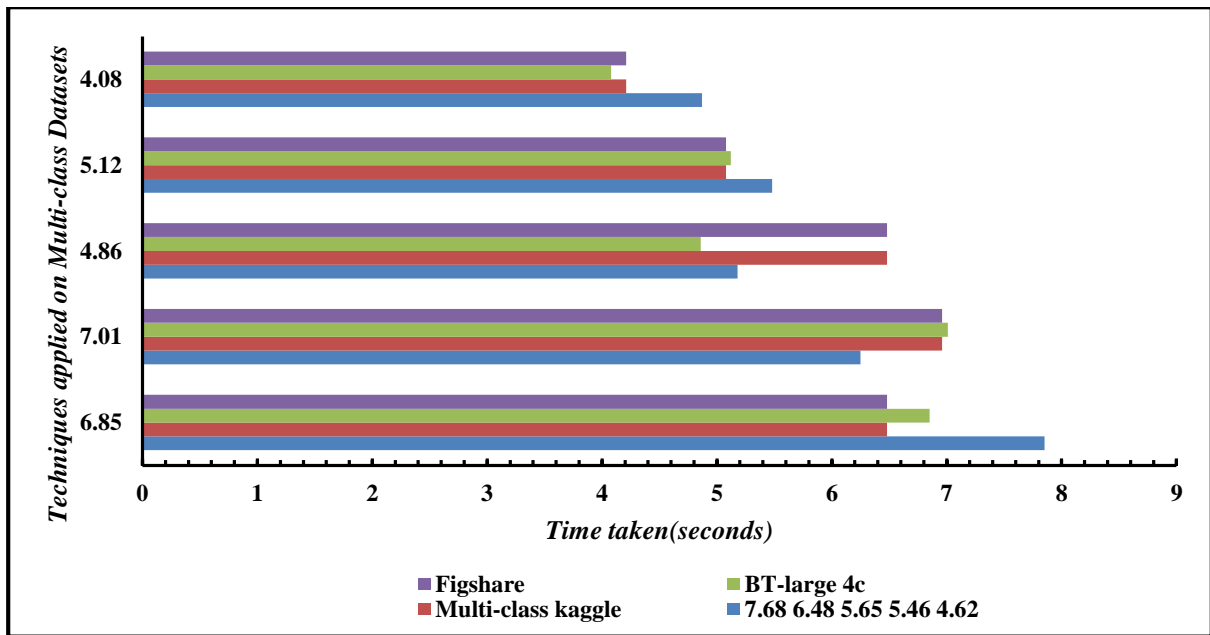*(i) Time Complexity*

Our proposed model also aims not only at improving the accuracy but also to boost the computational efficiency. Figure 7.7 provides a graphical view, with time per image on the x-axis and different classifiers on the y-axis, clearly showing that our model consistently processes images faster than other models in the current environment.

184

Figure 7.7 Performance of Proposed Model vs SOTA for Binary Class Classification

From Figure 7.8, it can be observed that the proposed model not only performed well on binary datasets but also proved computationally efficient in multi class datasets also. Results show that our proposed model processed each image in just 4.87 seconds per minute on brain tumor MRI dataset, 4.62 seconds per image on kaggle multi-class dataset, 4.08 seconds on the BT-large 4c dataset and 4.21 seconds per image on the Figshare dataset, making it the most efficient among the models tested. Figure 7.8 provides a visual comparison, with time per image on the –axis and various classifiers on the y axis, showing that our model consistently processes images faster than other SOTA techniques in the current environment.

Figure 7.8Performance of Proposed model vs SOTA vs Proposed Model for Multi Class Classification.

(ii) Space Complexity:

The space complexity of the proposed model depends on the number of trainable parameters it has. Each parameter needs memory to store its value, so the more parameters there are, the more memory is needed. Besides the parameters, memory is also required for storing input data, activations, and gradients during training. This memory use is influenced by the size of the input data, the number of units in the model, and how many training steps are needed. Since deep learning models usually have millions of parameters, the memory demand becomes quite large for big models and datasets.

The main goal of the proposed model is to improve how efficiently it uses computing resources. This is done through two main techniques:

- *Selective Attention and Feature Focus:* Techniques like **MaxViT** and **Patterned Grid Mask** work by focusing on the most important regions or features of the input data, reducing the amount of computation required. MaxViT uses a grid-based attention

186

mechanism to process only key parts of the image, while the Patterned Grid Mask selectively processes relevant areas, minimizing unnecessary calculations and improving efficiency.

- *Simplified and Lightweight Operations:* **Gaussian filters** and **GRN_MLP** contribute by simplifying operations. Gaussian filters provide efficient noise reduction with minimal computational cost, improving input quality for faster processing. GRN_MLP reduces parameter complexity and computational waste by learning to ignore irrelevant features, leading to quicker convergence and lower resource consumption.

In conclusion, our results show that the proposed model achieves both high accuracy and computational efficiency, making it an excellent choice for real-world applications. Whether used for binary or multi-class classification, the model consistently outperforms state-of-the-art methods by processing each image faster and with greater accuracy. This combination of strong performance and low computational demand makes it an ideal solution for a range of practical uses, delivering meaningful improvements over current hybrid techniques. Its versatility and efficiency emphasize its potential for broad applications across diverse classification tasks.

The following figures visually represent the performance metrics achieved by the proposed hybrid model, which utilizes patterned Grid Masking and Gaussian filtering techniques on datasets like the Brain Tumor MRI, Kaggle multi-class, BT-large 4c, and Figshare datasets for multi-class classification, as well as various binary datasets. These figures underscore the model's strong performance, demonstrating its robustness, consistency, and adaptability across multiple datasets.

The proposed model effectively addresses class imbalance and the lack of generalized models in medical imaging using the Patterned Grid Masking technique, which improves feature extraction and tumor segmentation. Gaussian filters further enhance noise reduction,

187

ensuring accurate classification of even rare tumor types. This approach significantly boosts the model's accuracy and generalization across all tumor types.

7.4.2.3 Ablation study

This section provides deeper analysis of proposed model on various study cases.

- **Case A:** MaXViT base model is applied for both binary and multi-class classification.

- **Case B**: Introducing LeViT to the architecture.

- **Case C:** Integrating GRN based MLP module to the architecture.

- **Case D:** Integrating Gaussian Filters to the architecture.

- **Case E:** Integrating Patterned Grid Mask to the architecture.

- **Case F:** Proposed Model

The case study reveals a clear improvement in accuracy as advanced techniques are integrated into to model. The Bar graph from Figure 7.9 illustrates the impact of various model modifications on binary classification datasets. In Case A which applies MAXViT alone, the accuracy is modest, just below 80%. By introducing LeViT in Case B yields a slight improvement, but it remains insufficient. The integration of GRN-based model in case C further improved the accuracy, indicating that this module is beneficial for enhancing the model's ability to classify accurately. Likewise, the integration of Case D, suggest that these filters added a meaningful contribution to the model accuracy. In case E, the accuracy reached a notably higher level, with values recorded at 95%, explaining the effect of this integration. Finally, the proposed model in Case F achieves near-perfect accuracy, approaching 100% across all datasets, suggesting it is well generalised technique. Overall, this study highlights that integration of these techniques incrementally improves model performance, with the proposed model demonstrating the highest effectiveness and generalisation for binary classification.
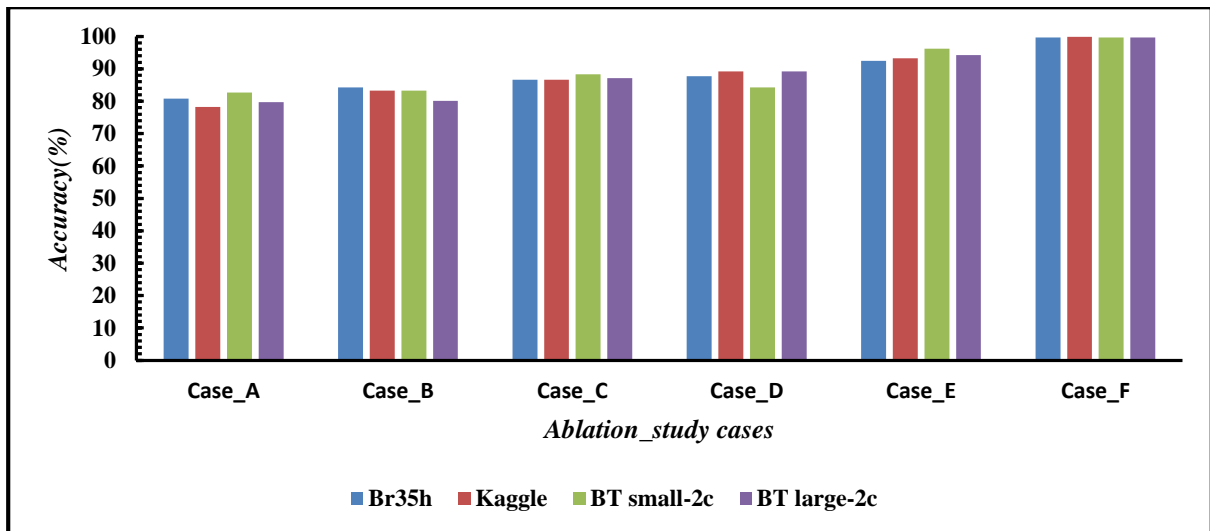
Figure 7.9 Ablation Study on Binary Classification

Figure 7.10 shows how different versions of a model perform on a multi-class classification task. Starting with Case A, where only MAXViT is used, the accuracy is fairly low. Each following case adds a new feature to the model, which generally improves accuracy. For instance, adding a GRN-based MLP module in Case C and Gaussian Filters in Case D make the model better, but it's only when the Patterned Grid Mask is added in Case E that we see a big boost in accuracy, getting close to 95%. The final model in Case F, which combines all these techniques, achieves the highest accuracy, almost reaching perfect performance. This shows that each addition helps, but the combined model is the most effective.
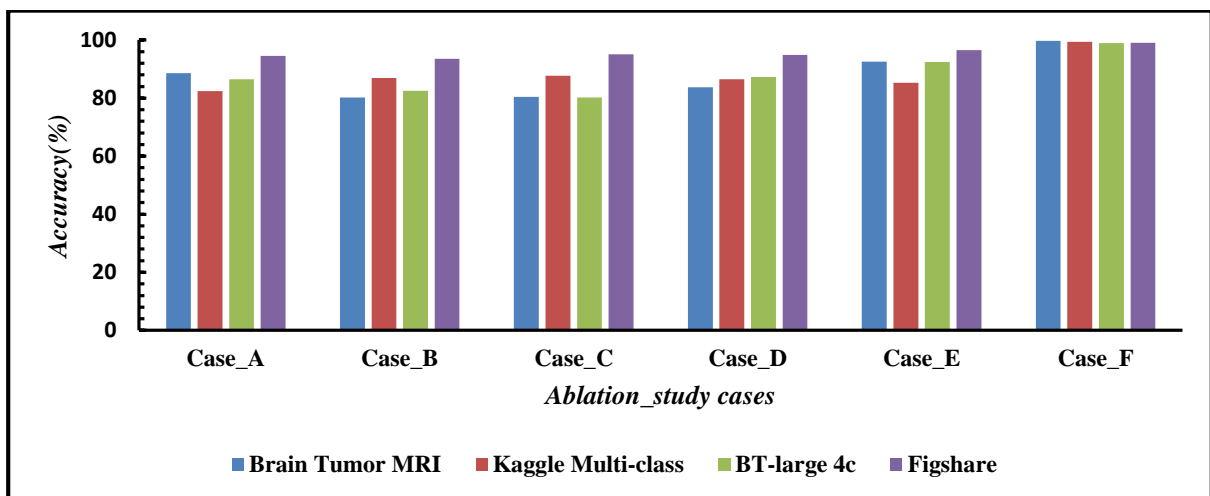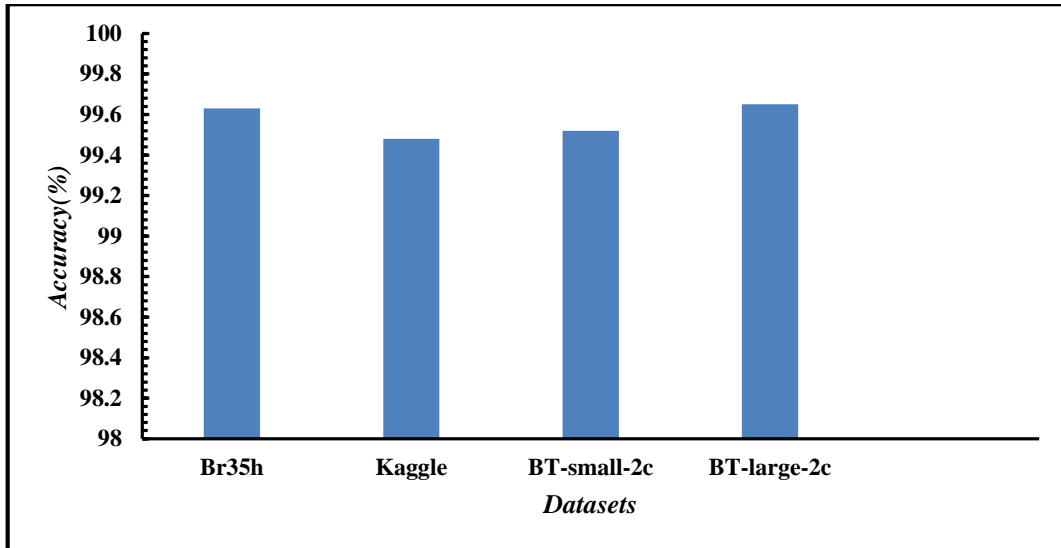

Figure 7.10 Ablation Study on Multi-Class Classification

7.4.2.4 Generalisation Analysis

The following Figure 7.11 provide an analysis of the model's generalization ability across multiple datasets. In binary classification, the model achieved exceptional performance on the Br35h dataset, showing superior results compared to other datasets. This performance highlights its effectiveness in identifying the specific features required for accurate classification within the Br35h dataset. To further evaluate generalization, we conducted an analysis where the model was trained exclusively on the Br35h dataset and then tested on images from three other datasets that were not part of the training phase. This approach enabled us to assess how well the model could adapt to unseen data from entirely different sources. The results, illustrated in the figure, reveal that the proposed model maintained strong accuracy when applied to these previously unseen datasets. This suggests that the model can effectively generalize, even when trained on a single dataset and exposed to data from different sources during testing.

Such robust generalization is essential for binary classification in medical imaging, as it indicates that the model can perform reliably across diverse datasets. The ability to generalize in this way supports the model's practical utility, especially in real-world applications where unseen data often differ from the original training dataset. This cross-dataset performance highlights the model's adaptability and its potential to be a reliable tool in various clinical settings, as it is not strictly bound to the specific characteristics of one dataset but can extend its learned representations across other binary classification tasks.
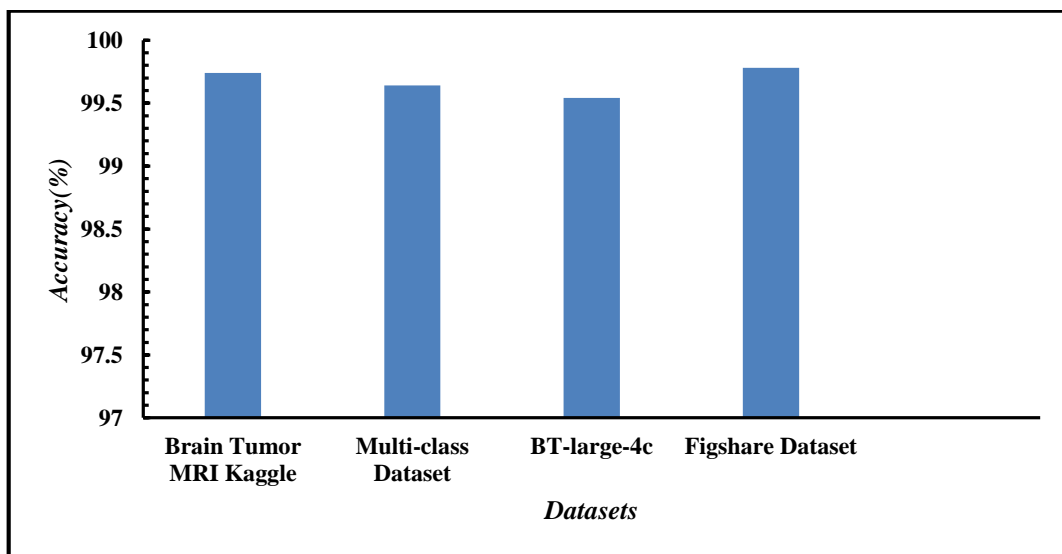
Figure 7.11 Generalisation Analysis on Binary Classification

The following Figure 7.12 provides an analysis of the model's generalization ability across multiple datasets. In multi-class classification, the model achieved exceptional performance on the Brain Tumor MRI dataset, showing superior results compared to other datasets. This performance demonstrates its effectiveness in identifying specific features required for accurate classification across multiple classes within the Br35h dataset.

To further evaluate generalization, we conducted an analysis where the model was trained exclusively on the Brain Tumor MRI dataset and then tested on images from three other datasets that were not included in the training phase. This approach allowed us to assess how well the model could adapt to unseen data across multiple classes from entirely different sources. The results, illustrated in the figure, reveal that the proposed model maintained strong accuracy when applied to these previously unseen datasets, even in a multi-class classification. This suggests that the model can effectively generalize to a range of classes, even when trained on a single dataset and exposed to data from other sources during testing.

Such robust generalization is essential for multi-class classification in medical imaging, as it indicates that the model can reliably distinguish between multiple classes across diverse

datasets. This ability to generalize supports the model's practical utility, especially in real-world applications where unseen data often vary significantly from the original training dataset. The cross-dataset performance highlights the model's adaptability and its potential to serve as a reliable tool in various clinical settings, as it is not strictly limited to the specific characteristics of one dataset but can extend its learned representations to multi-class classification tasks across other datasets.

Figure 7.12 Generalisation Analysis on Multi class Classification

7.5 Chapter Summary

This chapter presents a comprehensive exploration of a novel hybrid deep learning (DL) model designed to improve classification accuracy and robustness in binary and multi-class scenarios. A primary focus is placed on addressing class imbalance, reducing overfitting, and handling image occlusions. To achieve these goals, the model incorporates several innovative techniques, including the use of Gaussian filters for noise reduction and a Patterned-GridMask, which enhances the standard GridMask for more effective occlusion handling. These improvements allow the model to generalize effectively across diverse datasets, even under challenging conditions. Further architectural advancements are introduced, specifically enhancing the Multi-Axis Vision Transformer (MaxViT) for optimized computational

192

efficiency and scalability. These modifications improve the model's performance, enabling faster and more accurate classifications. A rigorous comparative evaluation demonstrates the model's superior performance relative to existing DL models, confirming its robustness and adaptability across multiple datasets. The chapter concludes with insights into the model's potential applications, underlining its capacity to meet complex classification requirements with enhanced accuracy, resilience, and efficiency.

Chapter Eight: CONCLUSION AND FUTURE SCOPE

This chapter presents a comprehensive summary of the research work done. It includes the research summary of the work done in Section 8.1. The chapter also presents the future aspects of the research work performed in Section 8.2 and how the study can help the future researchers in the said domain.

8.1 Research Summary

This research significantly advances the field of image processing and classification by addressing critical challenges such as dataset imbalance, generalization, and computational efficiency in machine learning (ML) and deep learning (DL) models. Through the development and application of innovative hybrid approaches, it demonstrates the effectiveness of combining data-level and algorithmic-level modifications to enhance model performance.

The study begins with an in-depth exploration of class imbalance, its characteristics, and metrics for measurement. A thorough review of existing approaches for addressing class imbalance in deep learning revealed that standalone data-level or algorithmic-level solutions often fall short in delivering optimal outcomes. From extensive experiments, it was established that hybrid strategies combining both levels are more effective in mitigating imbalance and improving model performance. This led to the development of ConvNexT-CRL, a hybrid model incorporating advanced data augmentation techniques alongside the CRL loss function. Applied to the BreakHis dataset, this model achieved a remarkable accuracy of 99.2%, outperforming existing methods by 18%, thus validating the superiority of hybrid approaches.

Further research extended these principles to tumor detection and classification tasks. At the data level, generative adversarial networks (GANs) were employed to generate diverse synthetic images, addressing data scarcity and increasing dataset variety. Traditional

augmentation techniques, including rotation, flipping, scaling, and contrast adjustments, were also applied to enhance robustness and generalization. On the algorithmic front, Vision Transformers (ViT) were integrated with auxiliary artificial neural networks (ANNs), enabling the capture of complex spatial and geometric features. This multi-output architecture effectively improved the model's decision-making capabilities. The combined approach delivered enhanced performance, achieving testing accuracies of 99.8% for binary classification and 99.3% for multi-class tasks, showcasing its potential in real-world applications.

In this research, to optimize resource utilization and reduce computational overhead, a novel SL(t)-ViT model was introduced, leveraging a multi-scale attention mechanism as a more efficient alternative to traditional self-attention. This mechanism effectively captured multi-scale features while significantly minimizing computational costs. By integrating this approach with GANs for diverse and high-quality data augmentation, the model addressed challenges such as image blurriness and model collapse, leading to substantial gains in accuracy.

For binary classification, it achieved testing accuracies of 99.8, 99.7, 99.6, 99.54, and 99.5, while for multi-class classification, it delivered 99.3, 98.6, 98.5, and 98.2. These results represent significant improvements, with gains of 1% to 2% for binary classification and 9% to 10% for multi-class classification compared to state-of-the-art techniques. The model's efficiency and accuracy, driven by its innovative multi-scale attention mechanism and diverse data augmentation strategies, position it as a leading solution for brain tumor identification, excelling in both performance and resource optimization.

Further to address the optimization problem by improving computational efficiency, a computationally efficient DL framework was proposed, addressing dataset imbalance through a hybrid model design, ensuring both accuracy and speed in breast cancer image classification.

Proposed novel framework excels in accuracy and generalization across medical imaging datasets, providing a robust tool for precise diagnostics.

The proposed model integrated features from two classifiers, Inception ResNet V2 and Vision Transformers (ViT), to enhance the classification of breast cancer. This synergistic blend enhances adaptability, ensuring consistent performance across diverse dataset scales. A key contribution is the introduction of an Efficient Attention Mechanism within one of the classifiers, optimizing focus on critical features for improved accuracy and computational efficiency. Further, a Resource-Efficient Optimization model through feature selection is proposed, streamlining computational usage without compromising accuracy. Addressing the inherent heterogeneity within classifiers, our framework integrates high dimensional features comprehensively, leading to more accurate tumor class predictions. This consideration of heterogeneity marks a significant leap forward in precision for breast cancer diagnosis.

An extensive analysis on datasets, BreakHis and BACH, that are imbalanced in nature is conducted by evaluating complexity, performance, and resource usage. Comprehensive evaluation using the datasets and standard performance metrics accuracy, precision, Recall, F1-score, MCC reveals the model's high efficacy, achieving a testing accuracy of 99.36 and 99.4, with precision, recall, F1-score and MCC scores of 99.19, 98.7, 98.98, 98.52 and 98.9, 99.86, 99.3, 98.8 on the BreakHis and BACH datasets, respectively. Our proposed model outperforms state-of-the-art techniques, demonstrating superior accuracy across different datasets, with improvements ranging from 0.25% to 15% on the BACH dataset and from 0.36% to 15.02% on the BreakHis dataset. The collective contributions, from framework and hybrid model design to feature selection and classifier heterogeneity consideration, establish a holistic and state-of-the-art approach, significantly improving accuracy and establishing optimization in breast cancer classification from MRI images.

Additionally, this research introduced innovative techniques to tackle generalization challenges in brain tumor classification. Gaussian noise reduction was employed to enhance image clarity, while a Patterned-Grid Mask selectively emphasized critical regions, preserving essential tumor details. Modifications to the Multi-Axis Vision Transformer (MaxViT) architecture, such as replacing MBConv blocks with LeViT blocks and incorporating GRN-based MLPs, significantly improved computational efficiency. These enhancements resulted in superior classification accuracy, robustness, and speed, outperforming state-of-the-art models across multiple datasets.

Throughout the research, the importance of multi-faceted approaches became evident. Hybrid methods consistently demonstrated superior performance, achieving up to 10% better results than existing techniques. Models incorporating Grid Attention, Block Attention, and Patterned-Grid Masks generalized well across underrepresented classes and untrained datasets. Furthermore, the integration of efficient attention mechanisms, such as the Nystrom Attention Mechanism, substantially reduced computational overhead while maintaining high accuracy, making these methods highly scalable and resource-efficient. The proposed model demonstrated outstanding performance in both binary and multi-class classification, achieving high accuracy, speed, and robustness across various tasks. Extensive evaluations across multiple datasets yielded testing accuracies of 99.63, 99.85, 99.65, and 99.65 for binary classification, and 99.47, 99.4, 98.96, and 99.11 for multi-class classification. These results reveal the model's superior performance, with a significant improvement of 1% to 2% in binary classification and 9% to 10% in multi-class classification over state-of-the-art methods.

In conclusion, this thesis highlights the efficacy of hybrid frameworks in overcoming critical challenges of class imbalance, generalization, and computational efficiency. By combining advanced data augmentation techniques, innovative attention mechanisms, and

hybrid architectures, this research not only achieves exceptional results but also sets a foundation for future advancements in AI-driven imaging solutions. These contributions pave the way for more reliable, efficient, and accessible tools, with broad implications for both research and real-world applications.

8.2 Future Aspects

Following are the future aspects of the research work performed:

- Results for AutoML implementation can be improved by using more advanced models for the ensemble with hyper-parameter optimization.

- More complex evolutionary strategies can be considered for finding the best models. Apart from working on model generation, research on AutoML can also be done in data preparation, feature engineering, and model evaluation. Thus, creating a complete platform to make addressing class imbalance very trivial to implement for the user by encapsulating all the complexities within an AutoML framework.

- Also, for addressing class imbalance, future work in this domain includes exploring non-hierarchical structures for the ensembles and different ways to optimize them. It is also essential to see how this framework performs across domains.

- The proposed model for improving generalization, can be extended to enhances image generalization, making it adaptable across various applications. With advanced techniques, researchers can develop a comprehensive generalized system that processes data directly, avoiding pre-conversion steps and preserving effectiveness. This adaptable framework allows for improved accuracy across diverse use cases, supporting generalization in new contexts.

8.3 Potential Industrial Applications

(i) Addressing Class Imbalance in Healthcare The methodologies developed in this research for addressing class imbalance hold transformative potential for healthcare applications, ensuring more reliable and accurate diagnostic and decision-making systems. Key applications include:

(ii) Improved Disease Detection and Diagnosis By mitigating the impact of class imbalance in medical datasets, the proposed frameworks enable the development of diagnostic tools that accurately identify rare conditions such as specific cancer subtypes or uncommon genetic disorders. This ensures that underrepresented cases are detected with higher precision, reducing the risk of misdiagnosis.

(iii) Enhanced Performance in Medical Imaging The hybrid models effectively balance datasets in medical imaging, ensuring consistent classification of tumors, lesions, and other abnormalities. This is particularly impactful in imaging modalities like MRI and histopathology, where class imbalance often skews the performance of traditional machine learning models.

(iv) Equitable Diagnostic Systems Addressing class imbalance helps create diagnostic systems that perform equitably across diverse patient populations, reducing healthcare disparities. This is critical for ensuring that minority or underrepresented groups receive accurate and reliable diagnoses.

(v) Robust AI for Rare Diseases By generating synthetic data and applying advanced augmentation techniques, the proposed methods improve the robustness of AI models in detecting and diagnosing rare diseases, where data scarcity often leads to poor model performance.

(vi) Precision in Multi-Class Healthcare Tasks The ability to handle imbalance in multi-class settings ensures that all disease categories, including those with fewer samples, are classified accurately. This has direct applications in areas such as multi-class tumor grading or identifying different stages of a disease.

(vii) Optimized Clinical Trial Analysis Balancing class representation in clinical trial datasets enhances the reliability of AI-driven analysis, ensuring that treatment effects are accurately measured across all patient subgroups.

These advancements in addressing class imbalance pave the way for fairer, more accurate, and inclusive healthcare solutions, ultimately improving patient care and outcomes.

References

[1] Y. Bengio, *Learning Deep Architectures for AI*. 2009. doi: 10.1561/9781601982957.

[2] E. J *et al.*, "Medical deep learning-A systematic meta-review," *Comput. Methods Programs Biomed.*, vol. 221, Jun. 2022, doi: 10.1016/j.cmpb.2022.106874.

[3] P. Ruiz-Ponce, D. Ortiz-Perez, J. Garcia-Rodriguez, and B. Kiefer, "POSEIDON: A Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments," *Sensors*, vol. 23, no. 7, p. 3691, Apr. 2023, doi: 10.3390/s23073691.

[4] A. Aslam and E. Curry, "A Survey on Object Detection for the Internet of Multimedia Things (IoMT) using Deep Learning and Event-based Middleware: Approaches, Challenges, and Future Directions," *Image Vis. Comput.*, vol. 106, p. 104095, Feb. 2021, doi: 10.1016/j.imavis.2020.104095.

[5] J. Rüter, U. Durak, and J. C. Dauer, "Investigating the Sim-to-Real Generalizability of Deep Learning Object Detection Models," *J. Imaging*, vol. 10, no. 10, Art. no. 10, Oct. 2024, doi: 10.3390/jimaging10100259.

[6] B. Benchamardimath and R. Hegadi, "A STUDY ON THE IMPORTANCE OF IMAGE PROCESSING AND ITS APLLICATIONS," pp. 2321–7308, Apr. 2020.

[7] R. Dastres and M. Soori, "Advanced Image Processing Systems," *Int. J. Imaging Robot.*, vol. 21, Feb. 2021.

[8] S. Dewangan, "Importance & Applications of Digital Image Processing," 2016. Accessed: Nov. 14, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Importance-%26-Applications-of-Digital-Image-Dewangan/f7ed4d79193713035de783949ff065837eb781dd.

[9] B. Crockett, L. Romero Cortés, and J. Azana, "Passive Amplification and Noise Mitigation of Optical Signals Through Talbot Processing," *J. Light. Technol.*, vol. PP, pp. 1–15, Jan. 2022, doi: 10.1109/JLT.2022.3212708.

[10] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3D random noise filtering for absorption optical projection tomography," *Phys. Med. Biol.*, vol. 55, no. 18, p. 5401, Aug. 2010, doi: 10.1088/0031-9155/55/18/009.

[11] W. Jasim and R. Mohammed, "A Survey on Segmentation Techniques for Image Processing," *Iraqi J. Electr. Electron. Eng.*, vol. 17, pp. 73–93, Dec. 2021, doi: 10.37917/ijeee.17.2.10.

[12] N. Sharma, M. Mishra, and M. Shrivastava, "COLOUR IMAGE SEGMENTATION TECHNIQUES AND ISSUES: AN APPROACH," vol. 1, no. 4, 2012.

[13] A. P. Sheppard, R. M. Sok, and H. Averdunk, "Techniques for image enhancement and segmentation of tomographic images of porous materials," *Phys. Stat. Mech. Its Appl.*, vol. 339, no. 1, pp. 145–151, Aug. 2004, doi: 10.1016/j.physa.2004.03.057.

[14] R. C. Gonzalez, R. E. Woods, and B. R. Masters, "Digital Image Processing, Third Edition," *J. Biomed. Opt.*, vol. 14, no. 2, p. 029901, 2009, doi: 10.1117/1.3115362.

[15] A. K. Jain, *Fundamentals of digital image processing*. USA: Prentice-Hall, Inc., 1989.

[16] Y. Pourasad and F. Cavallaro, "A Novel Image Processing Approach to Enhancement and Compression of X-ray Images," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 13, Art. no. 13, Jan. 2021, doi: 10.3390/ijerph18136724.

[17] N. Kumari, P. Sharma, and I. Kansal, "Systematic Review of Image Computing Under Some Basic and Improvised Techniques," 2023, pp. 274–301. doi: 10.1201/9781003367161-27.

[18] A. Beghdadi, C. Larabi, A. Bouzerdoum, and K. Iftekharuddin, "A Survey of Perceptual Image Processing Methods," *Signal Process. Image Commun.*, vol. 28, pp. 811–831, Sep. 2013, doi: 10.1016/j.image.2013.06.003.

[19] Y. Ma, K. Zhan, and Z. Wang, "Image Segmentation," in *Applications of Pulse-Coupled Neural Networks*, Y. Ma, K. Zhan, and Z. Wang, Eds., Berlin, Heidelberg: Springer, 2010, pp. 27–42. doi: 10.1007/978-3-642-13745-7_3.

[20] U. Qidwai and C. H. Chen, *Digital Image Processing: An Algorithmic Approach with MATLAB*. New York: Chapman and Hall/CRC, 2009. doi: 10.1201/9781420079517.

[21] M. Soori, B. Arezoo, and R. Dastres, "Artificial Intelligence, Machine Learning and Deep Learning in Advanced Robotics, A Review," Apr. 2023, doi: 10.1016/j.cogr.2023.04.001.

[22] N. Rane, M. Paramesha, S. Choudhary, and J. Rane, "Artificial Intelligence, Machine Learning, and Deep Learning for Advanced Business Strategies: A Review," vol. 2, pp. 147–171, Jul. 2024, doi: 10.5281/zenodo.12208298.

[23] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 1, pp. 131–156, Jan. 1997, doi: 10.1016/S1088-467X(97)00008-5.

[24] J. Rane, S. Mallick, Ö. Kaya, and N. Rane, "Scalable and adaptive deep learning algorithms for large-scale machine learning systems," 2024, pp. 39–92. doi: 10.70593/978-81-981271-0-5_2.

[25] V. Nasteski, "An overview of the supervised machine learning methods," *HORIZONS.B*, vol. 4, pp. 51–62, Dec. 2017, doi: 10.20544/HORIZONS.B.04.1.17.P05.

[26] E. Morales and H. J. Escalante, "A brief introduction to supervised, unsupervised, and reinforcement learning," 2022, pp. 111–129. doi: 10.1016/B978-0-12-820125-1.00017-8.

[27] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[28] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 281, Dec. 2019, doi: 10.1186/s12911-019-1004-8.

[29] R. C. Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, no. 20, p. 1920, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.

[30] "Deep Learning," MIT Press. Accessed: Nov. 15, 2024. [Online]. Available: https://mitpress.mit.edu/9780262035613/deep-learning/

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–44, May 2015, doi: 10.1038/nature14539.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[33] "RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks | IEEE Conference Publication | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8769811

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[35] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv 14091556*, Sep. 2014.

[36] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.

[37] "Representation Learning: A Review and New Perspectives." Accessed: Nov. 16, 2024. [Online]. Available:

https://www.researchgate.net/publication/240308775_Representation_Learning_A_Review_and_New_Perspectives

[38] "Long Short-Term Memory | Neural Computation." Accessed: Nov. 15, 2024. [Online]. Available: https://dl.acm.org/doi/10.1162/neco.1997.9.8.1735

[39] S. Kim, "Bishop Pattern Recognition and Machine Learning", Accessed: Nov. 15, 2024. [Online]. Available: https://www.academia.edu/17851990/Bishop_Pattern_Recognition_and_Machine_Learning

[40] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. 2009.

[41] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010950718922.

[42] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," vol. 14, Mar. 2001.

[43] J. Bergstra, J. Bergstra, Y. Bengio, and Y. Bengio, "Random Search for Hyper-Parameter Optimization".

[44] "Multiple Electric Energy Consumption Forecasting Using a Cluster-Based Strategy for Transfer Learning in Smart Building." Accessed: Nov. 15, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/20/9/2668

[45] "Managing minority opinions in large-scale group decision making based on community detection and group polarization | Computers and Industrial Engineering." Accessed: Nov. 15, 2024. [Online]. Available: https://dl.acm.org/doi/10.1016/j.cie.2022.108337

[46] H. V. Long *et al.*, "A New Approach for Construction of Geodemographic Segmentation Model and Prediction Analysis," *Comput. Intell. Neurosci.*, vol. 2019, p. 9252837, 2019, doi: 10.1155/2019/9252837.

[47] "Optimized Ensemble Machine Learning Framework for High Dimensional Imbalanced Bio Assays | IIETA." Accessed: Nov. 15, 2024. [Online]. Available: https://www.iieta.org/journals/ria/paper/10.18280/ria.330509

[48] A. Pregowska, M. Osial, and W. Urbańska, "The Application of Artificial Intelligence in the Effective Battery Life Cycle in the Closed Circular Economy Model—A Perspective," *Recycling*, vol. 7, no. 6, Art. no. 6, Dec. 2022, doi: 10.3390/recycling7060081.

[49] J. Pan, J. Liu, J. Zhang, X. Li, D. Quan, and Y. Li, "Depression Detection Using an Automatic Sleep Staging Method With an Interpretable Channel-Temporal Attention Mechanism," *IEEE Trans. Cogn. Dev. Syst.*, vol. 16, no. 4, pp. 1418–1432, Aug. 2024, doi: 10.1109/TCDS.2024.3358022.

[50] "Dealing with the Class Imbalance Problem in the Detection of Fake Job Descriptions," *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 521–535, Feb. 2021, doi: 10.32604/cmc.2021.015645.

[51] Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework," *Sensors*, vol. 20, no. 5, Art. no. 5, Jan. 2020, doi: 10.3390/s20051399.

[52] D. Sembroiz, D. Careglio, S. Ricciardi, and U. Fiore, "Planning and Operational energy optimization solutions for Smart Buildings," *Inf. Sci.*, vol. 476, Jun. 2018, doi: 10.1016/j.ins.2018.06.003.

[53] B. An and Y. Suh, "Identifying financial statement fraud with decision rules obtained from Modified Random Forest," *Data Technol. Appl.*, vol. ahead-of-print, May 2020, doi: 10.1108/DTA-11-2019-0208.

[54] K. W. De Bock, K. Coussement, and S. Lessmann, "Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach," *Eur. J. Oper. Res.*, vol. 285, no. 2, pp. 612–630, 2020.

[55] "55. Mazurowski, M.A.; Habas, P.A.; Zurada, J.M.; Lo, J.Y.; Baker, J.A.; Tourassi, G.D. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Netw. 2008, 21, 427–436. - Google Search." Accessed: Nov. 15, 2024. [Online].

[56] D. Goyal, A. Choudhary, B. S. Pabla, and S. S. Dhami, "Support vector machines based non-contact fault diagnosis system for bearings," *J. Intell. Manuf.*, vol. 31, no. 5, pp. 1275–1289, 2020.

[57] T. Fawcett and F. Provost, "Adaptive Fraud Detection," *Data Min. Knowl. Discov.*, vol. 1, pp. 291–316, Sep. 1997, doi: 10.1023/A:1009700419189.

[58] X. Gao, Z. Chen, S. Tang, Y. Zhang, and J. Li, "Adaptive weighted imbalance learning with application to abnormal activity recognition," *Neurocomputing*, vol. 173, pp. 1927–1935, Jan. 2016, doi: 10.1016/j.neucom.2015.09.064.

[59] M. Koziarski, B. Kwolek, and B. Cyganek, "Convolutional Neural Network-Based Classification of Histopathological Images Affected by Data Imbalance: 17th International Conference, ICAISC 2018, Zakopane, Poland, June 3-7, 2018, Proceedings, Part II," 2019, pp. 1–11. doi: 10.1007/978-3-030-12177-8_1.

[60] D. Olszewski, "A probabilistic approach to fraud detection in telecommunications," *Knowl.-Based Syst.*, vol. 26, pp. 246–258, Feb. 2012, doi: 10.1016/j.knosys.2011.08.018.

[61] I. Lin, O. Loyola-González, R. Monroy, and M. A. Medina-Pérez, "A Review of Fuzzy and Pattern-Based Approaches for Class Imbalance Problems," *Appl. Sci.*, vol. 11, no. 14, Art. no. 14, Jan. 2021, doi: 10.3390/app11146310.

[62] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, *Safe-Level-SMOTE: Safe-Level-Synthetic Minority Over-Sampling TEchnique for Handling the Class Imbalanced Problem*, vol. 5476. 2009, p. 482. doi: 10.1007/978-3-642-01307-2_43.

[63] C. Zhang and T. Chen, "An active learning framework for content-based information retrieval," *IEEE Trans. Multimed.*, vol. 4, no. 2, pp. 260–268, Jun. 2002, doi: 10.1109/TMM.2002.1017738.

[64] M. Kubat, R. C. Holte, and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Radar Images," *Mach. Learn.*, vol. 30, no. 2, pp. 195–215, Feb. 1998, doi: 10.1023/A:1007452223027.

[65] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," *J King Saud Univ Comput Inf Sci*, vol. 35, no. 1, pp. 145–174, Jan. 2023, doi: 10.1016/j.jksuci.2022.11.008.

[66] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[67] Z.-B. Zhu and Z.-H. Song, "Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis," *Chem. Eng. Res. Des. - CHEM ENG RES DES*, vol. 88, pp. 936–951, Aug. 2010, doi: 10.1016/j.cherd.2010.01.005.

[68] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007, doi: 10.1016/j.patcog.2006.12.019.

[69] G. Yu, C. Domeniconi, H. Rangwala, G. Zhang, and Z. Yu, "Transductive multi-label ensemble classification for protein function prediction," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, in KDD '12.

New York, NY, USA: Association for Computing Machinery, Aug. 2012, pp. 1077–1085. doi: 10.1145/2339530.2339700.

[70] Y. Zhang, Y. Wang, X.-Y. Liu, S. Mi, and M.-L. Zhang, "Large-Scale Multi-Label Classification using Unknown Streaming Images," *Pattern Recognit.*, vol. 99, p. 107100, Nov. 2019, doi: 10.1016/j.patcog.2019.107100.

[71] "Multi-label classification with label clusters," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/385526036_Multi-label_classification_with_label_clusters

[72] "Learning multi-label scene classification," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/222430151_Learning_multi-label_scene_classification

[73] "Classifier Chains for Multi-label Classification." Accessed: Nov. 16, 2024. [Online]. Available:https://www.researchgate.net/publication/227319967_Classifier_Chains_for_Multi-label_Classification

[74] H. Han, W.-Y. Wang, and B.-H. Mao, *Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning*, vol. 3644. 2005, p. 887. doi: 10.1007/11538059_91.

[75] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Jun. 09, 2011, *arXiv*: arXiv:1106.1813. doi: 10.48550/arXiv.1106.1813.

[76] "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning." Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/225129029_Borderline-SMOTE_A_New_Over-Sampling_Method_in_Imbalanced_Data_Sets_Learning

[77] M. Z. Abedin, C. Guotai, and F. E. Moula, "Weighted SMOTE-Ensemble Algorithms: Evidence from Chinese Imbalance Credit Approval Instances," in *2019 2nd International Conference on Data Intelligence and Security (ICDIS)*, Jun. 2019, pp. 208–211. doi: 10.1109/ICDIS.2019.00038.

[78] M.-L. Zhang and Z.-H. Zhou, *A k-nearest neighbor based algorithm for multi-label classification*, vol. 2. 2005, p. 721 Vol. 2. doi: 10.1109/GRC.2005.1547385.

[79] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2*, in IJCAI'01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 2001, pp. 973–978.

[80] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

[81] "Breast cancer." Accessed: Nov. 15, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[82] "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries - Sung - 2021 - CA: A Cancer Journal for Clinicians - Wiley Online Library." Accessed: Nov. 15, 2024. [Online]. Available: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660

[83] H. Bahramiabarghouei, E. Porter, A. Santorelli, B. Gosselin, M. Popović, and L. A. Rusch, "Flexible 16 Antenna Array for Microwave Breast Cancer Detection," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2516–2525, Oct. 2015, doi: 10.1109/TBME.2015.2434956.

[84] S. Lei *et al.*, "Breast cancer incidence and mortality in women in China: temporal trends and projections to 2030," *Cancer Biol. Med.*, vol. 18, no. 3, pp. 900–909, Aug. 2021, doi: 10.20892/j.issn.2095-3941.2020.0523.

[85] Y. Liang, H. Zhang, X. Song, and Q. Yang, "Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets," *Semin. Cancer Biol.*, vol. 60, pp. 14–27, Feb. 2020, doi: 10.1016/j.semcancer.2019.08.012.

[86] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," Feb. 06, 2015, *arXiv*: arXiv:1502.01852. doi: 10.48550/arXiv.1502.01852.

[87] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, Jan. 2016, doi: 10.1038/nature16961.

[88] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.

[89] "Deep Learning - an overview | ScienceDirect Topics." Accessed: Nov. 16, 2024. [Online]. Available: https://www.sciencedirect.com/topics/engineering/deep-learning

[90] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[91] M. Ghafoorian *et al.*, "Location Sensitive Deep Convolutional Neural Networks for Segmentation of White Matter Hyperintensities," *Sci. Rep.*, vol. 7, no. 1, p. 5110, Jul. 2017, doi: 10.1038/s41598-017-05300-5.

[92] P. Rajpurkar *et al.*, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," Dec. 25, 2017, *arXiv*: arXiv:1711.05225. doi: 10.48550/arXiv.1711.05225.

[93] M. Sinclair *et al.*, "Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2018, pp. 714–717, Jul. 2018, doi: 10.1109/EMBC.2018.8512278.

[94] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[95] S. li, S. Wenfeng, H. Qin, and A. Hao, "Deep Variance Network: An Iterative, Improved CNN Framework for Unbalanced Training Datasets," *Pattern Recognit.*, vol. 81, Apr. 2018, doi: 10.1016/j.patcog.2018.03.035.

[96] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Oct. 13, 2018, *arXiv*: arXiv:1710.05381. doi: 10.48550/arXiv.1710.05381.

[97] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sanchez, "Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1273–1284, May 2016, doi: 10.1109/TMI.2016.2526689.

[98] J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, "Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks," *Med. Image Anal.*, vol. 34, pp. 123–136, Dec. 2016, doi: 10.1016/j.media.2016.04.004.

[99] B. Dashtbozorg, J. Zhang, F. Huang, and B. M. Ter Haar Romeny, "Retinal Microaneurysms Detection Using Local Convergence Index Features," *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.*, vol. 27, no. 7, pp. 3300–3315, Jul. 2018, doi: 10.1109/TIP.2018.2815345.

[100] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognit.*, vol. 36, pp. 2967–2991, Dec. 2003, doi: 10.1016/S0031-3203(03)00192-4.

[101] A. Bria, N. Karssemeijer, and F. Tortorella, "Learning from unbalanced data: A cascade-based approach for detecting clustered microcalcifications," *Med. Image Anal.*, vol. 18, no. 2, pp. 241–252, Feb. 2014, doi: 10.1016/j.media.2013.10.014.

[102] A. Carass *et al.*, "Longitudinal multiple sclerosis lesion segmentation: Resource and challenge," *NeuroImage*, vol. 148, pp. 77–102, Mar. 2017, doi: 10.1016/j.neuroimage.2016.12.064.

[103] A. Bria, C. Marrocco, and F. Tortorella, "Addressing class imbalance in deep learning for small lesion detection on medical images," *Comput. Biol. Med.*, vol. 120, p. 103735, May 2020, doi: 10.1016/j.compbiomed.2020.103735.

[104] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017, doi: 10.1016/j.eswa.2016.12.035.

[105] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding Data Augmentation for Classification: When to Warp?," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2016, pp. 1–6. doi: 10.1109/DICTA.2016.7797091.

[106] "Deep Over-sampling Framework for Classifying Imbalanced Data | SpringerLink." Accessed: Nov. 16, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-71249-9_46

[107] A. Shrivastava, A. Gupta, and R. Girshick, "Training Region-Based Object Detectors with Online Hard Example Mining," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 761–769. doi: 10.1109/CVPR.2016.89.

[108] "MCS-based balancing techniques for skewed classes: An empirical comparison | IEEE Conference Publication | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/4761359

[109] Y. Zhang, B. Zhang, F. Coenen, J. Xiao, and W. Lu, "One-class kernel subspace ensemble for medical image classification," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 17, Feb. 2014, doi: 10.1186/1687-6180-2014-17.

[110] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical Image Classification Using a Light-Weighted Hybrid Neural Network Based on PCANet and DenseNet," *IEEE Access*, vol. 8, pp. 24697–24712, 2020, doi: 10.1109/ACCESS.2020.2971225.

[111] P. Kumar and B. Vijayakumar, "Brain Tumour Mr Image Segmentation and Classification Using by PCA and RBF Kernel Based Support Vector Machine," 2015. Accessed: Nov. 16, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Brain-Tumour-Mr-Image-Segmentation-and-Using-by-PCA-Kumar-Vijayakumar/e2754fa03bf147352b1d42875f7347a309bc8206

[112] "Principal Component Analysis in Medical Image Processing: A Study." Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/272576742_Principal_Component_Analysis_in_Medical_Image_Processing_A_Study

[113] S. Kumar, C. Dabas, and S. Godara, "Classification of Brain MRI Tumor Images: A Hybrid Approach," *Procedia Comput. Sci.*, vol. 122, pp. 510–517, Jan. 2017, doi: 10.1016/j.procs.2017.11.400.

[114] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomed. Signal Process. Control*, vol. 39, pp. 139–161, Jan. 2018, doi: 10.1016/j.bspc.2017.07.007.

[115] K. Jiang, J. Lu, and K. Xia, "A Novel Algorithm for Imbalance Data Classification Based on Genetic Algorithm Improved SMOTE," *Arab. J. Sci. Eng.*, vol. 41, no. 8, pp. 3255–3266, Aug. 2016, doi: 10.1007/s13369-016-2179-2.

[116] K. Jiang, J. Lu, and K. Xia, "A Novel Algorithm for Imbalance Data Classification Based on Genetic Algorithm Improved SMOTE," *Arab. J. Sci. Eng.*, vol. 41, no. 8, pp. 3255–3266, Aug. 2016, doi: 10.1007/s13369-016-2179-2.

[117] "Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus | Multimedia Systems." Accessed: Nov. 16, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00530-021-00817-2

[118] D. Devarriya, C. Gulati, V. Mansharamani, A. Sakalle, and A. Bhardwaj, "Unbalanced breast cancer data classification using novel fitness functions in genetic programming," *Expert Syst. Appl.*, vol. 140, p. 112866, Feb. 2020, doi: 10.1016/j.eswa.2019.112866.

[119] "A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data - ScienceDirect." Accessed: Nov. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046420300940

[120] "A Random Forest Based Approach for One Class Classification in Medical Imaging," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/278814349_A_Random_Forest_Based_Approach_for_One_Class_Classification_in_Medical_Imaging

[121] M. Bader-El-Den, E. Teitei, and T. Perry, "Biased Random Forest For Dealing With the Class Imbalance Problem," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2163–2172, Jul. 2019, doi: 10.1109/TNNLS.2018.2878400.

[122] D. Müller, I. Soto-Rey, and F. Kramer, "An Analysis on Ensemble Learning optimized Medical Image Classification with Deep Convolutional Neural Networks," Apr. 13, 2022, *arXiv*: arXiv:2201.11440. doi: 10.48550/arXiv.2201.11440.

[123] N. V. Shamna and B. A. Musthafa, "Feature Extraction Method using HoG with LTP for Content-Based Medical Image Retrieval," *Int. J. Electr. Comput. Eng. Syst.*, vol. 14, no. 3, Art. no. 3, Mar. 2023, doi: 10.32985/ijeces.14.3.4.

[124] J. Liu, F. Guo, H. Gao, Z. Huang, Y. Zhang, and H. Zhou, "Image classification method on class imbalance datasets using multi-scale CNN and two-stage transfer learning," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14179–14197, Nov. 2021, doi: 10.1007/s00521-021-06066-8.

[125] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, and H. Chen, "Medical image classification using spatial adjacent histogram based on adaptive local binary patterns," *Comput. Biol. Med.*, vol. 72, pp. 185–200, May 2016, doi: 10.1016/j.compbiomed.2016.03.010.

[126] "Research on Medical Image Classification Based on Machine Learning | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9091175

[127] "Deep convolutional neural network based medical image classification for disease diagnosis | Journal of Big Data | Full Text." Accessed: Nov. 16, 2024. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0276-2

[128] "Texture based medical image classification by using multi-scale gabor rotation-invariant local binary pattern (MGRLBP)." Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/321017410_Texture_based_medical_image_classification_by_using_multi-scale_gabor_rotation-invariant_local_binary_pattern_MGRLBP

[129] K. M. Poloni, I. A. Duarte de Oliveira, R. Tam, and R. J. Ferrari, "Brain MR image classification for Alzheimer's disease diagnosis using structural hippocampal

asymmetrical attributes from directional 3-D log-Gabor filter responses," *Neurocomputing*, vol. 419, pp. 126–135, Jan. 2021, doi: 10.1016/j.neucom.2020.07.102.

[130] A. H. Barshooi and A. Amirkhani, "A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-Ray images," *Biomed. Signal Process. Control*, vol. 72, p. 103326, Feb. 2022, doi: 10.1016/j.bspc.2021.103326.

[131] "Computer-aided diagnostic network for brain tumor classification employing modulated Gabor filter banks | The Visual Computer." Accessed: Nov. 16, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00371-020-01977-4

[132] "Gabor Filter Algorithm for medical image processing: evolution in Big Data context | IEEE Conference Publication | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9151681

[133] "Medical Image Coloring Based on Gabor Filtering for Internet of Medical Things | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9106393

[134] R. M. Azawi, D. A. Abdulah, J. M. Abbas, and I. T. Ibrahim, "Brain Tumors Classification by Using Gray Level Co-occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network," *Diyala J. Med.*, vol. 14, no. 2, Art. no. 2, Aug. 2018.

[135] "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix | PLOS ONE." Accessed: Nov. 16, 2024. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267955

[136] R. Biswas, A. Nath, and S. Roy, "Mammogram Classification Using Gray-Level Co-occurrence Matrix for Diagnosis of Breast Cancer," in *2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*, Sep. 2016, pp. 161–166. doi: 10.1109/ICMETE.2016.85.

[137] "Detection and Classification of Brain Tumor from MRI Medical Image using Wavelet Transform and PSO based LLRBFNN Algorithm." Accessed: Nov. 16, 2024. [Online]. Available:
https://www.researchgate.net/publication/325772339_Detection_and_Classification_of_Brain_Tumor_from_MRI_Medical_Image_using_Wavelet_Transform_and_PSO_based_LLRBFNN_Algorithm

[138] "Brain Tumor Classification in Magnetic Resonance Images Using Deep Learning and Wavelet Transform." Accessed: Nov. 16, 2024. [Online]. Available: https://www.scirp.org/journal/paperinformation?paperid=100953

[139] "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9505667

[140] "An overview of classification algorithms for imbalanced datasets," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/292018027_An_overview_of_classification_algorithms_for_imbalanced_datasets

[141] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," *Inf. Sci.*, vol. 465, pp. 1–20, Oct. 2018, doi: 10.1016/j.ins.2018.06.056.

[142] P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 4, pp. 829–837, Apr. 2019, doi: 10.1007/s00432-018-02834-7.

[143] "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results | IEEE Conference Publication | IEEE Xplore."

Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9078901

[144] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, "Clustering-based undersampling in class-imbalanced data," *Inf. Sci.*, vol. 409–410, pp. 17–26, Oct. 2017, doi: 10.1016/j.ins.2017.05.008.

[145] "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks." Accessed: Nov. 16, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/13/6/4006

[146] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Mar. 2013, doi: 10.1186/1471-2105-14-106.

[147] L. Gao, L. Zhang, C. Liu, and S. Wu, "Handling imbalanced medical image data: A deep-learning-based one-class classification approach," *Artif. Intell. Med.*, vol. 108, p. 101935, Aug. 2020, doi: 10.1016/j.artmed.2020.101935.

[148] "Research on expansion and classification of imbalanced data based on SMOTE algorithm | Scientific Reports." Accessed: Nov. 16, 2024. [Online]. Available: https://www.nature.com/articles/s41598-021-03430-5

[149] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf. Sci.*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.

[150] H. Guan and M. Liu, "Domain Adaptation for Medical Image Analysis: A Survey," Feb. 18, 2021, *arXiv*: arXiv:2102.09508. doi: 10.48550/arXiv.2102.09508.

[151] C. Gan, L. Wang, and Z. Zhang, "Multi-entity sentiment analysis using self-attention based hierarchical dilated convolutional neural network," *Future Gener. Comput. Syst.*, vol. 112, May 2020, doi: 10.1016/j.future.2020.05.022.

[152] Y. Luo, H. Feng, X. Weng, K. Huang, and H. Zheng, "A novel oversampling method based on SeqGAN for imbalanced text classification," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 2891–2894. doi: 10.1109/BigData47090.2019.9006138.

[153] "Deep Learning." Accessed: Nov. 16, 2024. [Online]. Available: https://mitpress.mit.edu/9780262035613/deep-learning/

[154] S. Niu, B. Li, X. Wang, and H. Lin, "Defect Image Sample Generation With GAN for Improving Defect Recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1611–1622, Jul. 2020, doi: 10.1109/TASE.2020.2967415.

[155] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.

[156] F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training," Nov. 29, 2017, *arXiv*: arXiv:1711.06606. doi: 10.48550/arXiv.1711.06606.

[157] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Feb. 07, 2018, *arXiv*: arXiv:1708.02002. doi: 10.48550/arXiv.1708.02002.

[158] S. Ali and N. Bouguila, "Variational Learning of Beta-Liouville Hidden Markov Models for Infrared Action Recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019, pp. 898–906. doi: 10.1109/CVPRW.2019.00119.

[159] D. Lamrani, M. A. Bouqentar, M. A. Mahjoubi, S. Hamida, B. Cherradi, and L. Bahatti, "Automatic Localization and Classification of Brain Tumors using Deep Residual Networks and MRI Images," in *2024 4th International Conference on Innovative*

Research in Applied Science, Engineering and Technology (IRASET), May 2024, pp. 1–8. doi: 10.1109/IRASET60544.2024.10548396.

[160] Z. Liu, J. Peng, X. Guo, S. Chen, and L. Liu, "Breast cancer classification method based on improved VGG16 using mammography images," *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 2, p. 100885, Jun. 2024, doi: 10.1016/j.jrras.2024.100885.

[161] "SovaSeg-Net: Scale Invariant Ovarian Tumors Segmentation from Ultrasound Images | IEEE Conference Publication | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10647995

[162] "CACBL-Net: a lightweight skin cancer detection system for portable diagnostic devices using deep learning based channel attention and adaptive class balanced focal loss function | Multimedia Tools and Applications." Accessed: Nov. 16, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11042-024-19485-1

[163] "U-Net Segmentation of Lung Cancer CT Scans for 3D Rendering," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/384098742_U-Net_Segmentation_of_Lung_Cancer_CT_Scans_for_3D_Rendering

[164] H. M. Rai, J. Yoo, and S. Dashkevych, "Two-headed UNetEfficientNets for parallel execution of segmentation and classification of brain tumors: incorporating postprocessing techniques with connected component labelling," *J. Cancer Res. Clin. Oncol.*, vol. 150, no. 4, p. 220, Apr. 2024, doi: 10.1007/s00432-024-05718-1.

[165] "Enhancing Brain Tumor Classification Using Data-Efficient Image Transformer," ResearchGate. Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/380450640_Enhancing_Brain_Tumor_Classification_Using_Data-Efficient_Image_Transformer

[166] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 9260–9269. doi: 10.1109/CVPR.2019.00949.

[167] R. D. Prayogo, N. Hamid, and H. Nambo, "An Improved Transfer Learning-Based Model with Data Augmentation for Brain Tumor Detection," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Feb. 2024, pp. 602–607. doi: 10.1109/ICAIIC60209.2024.10463471.

[168] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, Dec. 2007, doi: 10.1016/j.patcog.2007.04.009.

[169] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," in *Knowledge Discovery in Databases: PKDD 2003*, vol. 2838, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., in Lecture Notes in Computer Science, vol. 2838. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 107–119. doi: 10.1007/978-3-540-39804-2_12.

[170] W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf, and Y. A. Bangash, "An In-Depth Analysis of IoT Security Requirements, Challenges, and Their Countermeasures via Software-Defined Security," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10250–10276, Oct. 2020, doi: 10.1109/JIOT.2020.2997651.

[171] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *Eur. J. Oper. Res.*, vol. 247, no. 1, pp. 124–136, Nov. 2015, doi: 10.1016/j.ejor.2015.05.030.

[172] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based

Approaches," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[173] "A texture-Based Video Processing Methodology Using Bayesian Data Fusion for Autonomous Crack Detection on Metallic Surfaces - Chen - 2017 - Computer-Aided Civil and Infrastructure Engineering - Wiley Online Library." Accessed: Nov. 16, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/mice.12256

[174] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, no. 1, pp. 239–263, May 2002, doi: 10.1016/S0004-3702(02)00190-X.

[175] "Transformers in medical image analysis | Intelligent Medicine." Accessed: Nov. 16, 2024. [Online]. Available: https://mednexus.org/doi/full/10.1016/j.imed.2022.07.002

[176] "LETCP: A Label-Efficient Transformer-Based Contrastive Pre-Training Method for Brain Tumor Segmentation." Accessed: Nov. 16, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/12/21/11016

[177] Y. Lee and P. Kang, "AnoViT: Unsupervised Anomaly Detection and Localization with Vision Transformer-based Encoder-Decoder," Mar. 21, 2022, *arXiv*: arXiv:2203.10808. doi: 10.48550/arXiv.2203.10808.

[178] L. Luque, J. A. Ottesen, A. Bjørnerud, K. E. Emblem, and B. J. MacIntosh, "Reducing Annotator's Burden: Cross-Pseudo Supervision for Brain Tumor Segmentation," *Proc. North. Lights Deep Learn. Workshop*, vol. 4, Jan. 2023, doi: 10.7557/18.6815.

[179] S. Mehta *et al.*, "End-to-End diagnosis of breast biopsy images with transformers," *Med. Image Anal.*, vol. 79, p. 102466, Jul. 2022, doi: 10.1016/j.media.2022.102466.

[180] J. Zahálka, S. Rudinac, and M. Worring, "Interactive Multimodal Learning for Venue Recommendation," *IEEE Trans. Multimed.*, vol. 17, no. 12, pp. 2235–2244, Dec. 2015, doi: 10.1109/TMM.2015.2480007.

[181] "Irrelevant attributes and imbalanced classes in multi-label text-categorization domains - IOS Press." Accessed: Nov. 16, 2024. [Online]. Available: https://content.iospress.com/articles/intelligent-data-analysis/ida00499

[182] L. Cao and Y. Zhai, "Imbalanced Data Classification Based on a Hybrid Resampling SVM Method," in *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 1533–1536. doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.275.

[183] B. Tang and H. He, "GIR-based ensemble sampling approaches for imbalanced learning," *Pattern Recognit.*, vol. 71, pp. 306–319, Nov. 2017, doi: 10.1016/j.patcog.2017.06.019.

[184] Y. Song *et al.*, "Large Margin Local Estimate With Applications to Medical Image Classification," *IEEE Trans. Med. Imaging*, vol. 34, no. 6, pp. 1362–1377, Jun. 2015, doi: 10.1109/TMI.2015.2393954.

[185] Q. Dong, S. Gong, and X. Zhu, "Imbalanced Deep Learning by Minority Class Incremental Rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1367–1381, Jun. 2019, doi: 10.1109/TPAMI.2018.2832629.

[186] S. Wang and X. Yao, "Using Class Imbalance Learning for Software Defect Prediction," *IEEE Trans. Reliab.*, vol. 62, no. 2, pp. 434–443, Jun. 2013, doi: 10.1109/TR.2013.2259203.

[187] G. Kim, B. K. Chae, and D. L. Olson, "A support vector machine (SVM) approach to imbalanced datasets of customer responses: comparison with other customer response

models," *Serv. Bus.*, vol. 7, no. 1, pp. 167–182, Mar. 2013, doi: 10.1007/s11628-012-0147-9.

[188] S. Barua, Md. M. Islam, X. Yao, and K. Murase, "MWMOTE–Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 405–425, Feb. 2014, doi: 10.1109/TKDE.2012.232.

[189] L. Song, D. Li, X. Zeng, Y. Wu, L. Guo, and Q. Zou, "nDNA-prot: identification of DNA-binding proteins based on unbalanced classification," *BMC Bioinformatics*, vol. 15, no. 1, p. 298, Sep. 2014, doi: 10.1186/1471-2105-15-298.

[190] D. Zhang, J. Ma, J. Yi, X. Niu, and X. Xu, "An ensemble method for unbalanced sentiment classification," in *2015 11th International Conference on Natural Computation (ICNC)*, Aug. 2015, pp. 440–445. doi: 10.1109/ICNC.2015.7378029.

[191] L. Cao and Y. Zhai, "Imbalanced Data Classification Based on a Hybrid Resampling SVM Method," in *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 1533–1536. doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.275.

[192] Ö. F. Arar and K. Ayan, "Software defect prediction using cost-sensitive neural network," *Appl. Soft Comput.*, vol. 33, pp. 263–277, Aug. 2015, doi: 10.1016/j.asoc.2015.04.045.

[193] C. Y. Wang, L. L. Hu, M. Z. Guo, X. Y. Liu, and Q. Zou, "imDC: an ensemble learning method for imbalanced classification with miRNA data," *Genet. Mol. Res. GMR*, vol. 14, no. 1, pp. 123–133, Jan. 2015, doi: 10.4238/2015.January.15.15.

[194] R. Sowah, M. Agebure, G. Mills, K. M. Koumadi, and S. Fiawoo, "New Cluster Undersampling Technique for Class Imbalance Learning," *Int. J. Mach. Learn. Comput.*, vol. 6, pp. 205–214, Jun. 2016, doi: 10.18178/ijmlc.2016.6.3.599.

[195] T. Lee, K. Lee, and C. O. Kim, "Performance of Machine Learning Algorithms for Class-Imbalanced Process Fault Detection Problems," *IEEE Trans. Semicond. Manuf.*, vol. 29, pp. 1–1, Nov. 2016, doi: 10.1109/TSM.2016.2602226.

[196] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," *2016 IEEE Int. Conf. Image Process. ICIP*, pp. 3713–3717, Sep. 2016, doi: 10.1109/ICIP.2016.7533053.

[197] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning Deep Representation for Imbalanced Classification," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 5375–5384. doi: 10.1109/CVPR.2016.580.

[198] J. Du, C.-M. Vong, C.-M. Pun, P.-K. Wong, and W.-F. Ip, "Post-boosting of classification boundary for imbalanced data using geometric mean," *Neural Netw. Off. J. Int. Neural Netw. Soc.*, vol. 96, pp. 101–114, Dec. 2017, doi: 10.1016/j.neunet.2017.09.004.

[199] S. Wan, Y. Duan, and Q. Zou, "HPSLPred: An Ensemble Multi-Label Classifier for Human Protein Subcellular Location Prediction with Imbalanced Source," *Proteomics*, vol. 17, no. 17–18, Sep. 2017, doi: 10.1002/pmic.201700262.

[200] W. Ding, D.-Y. Huang, Z. Chen, X. Yu, and W. Lin, *Facial action recognition using very deep networks for highly imbalanced class distribution*. 2017, p. 1372. doi: 10.1109/APSIPA.2017.8282246.

[201] S. Pouyanfar *et al.*, *Dynamic Sampling in Convolutional Neural Networks for Imbalanced Data Classification*. 2018. doi: 10.1109/MIPR.2018.00027.

[202] "Predicting Hospital Readmission via Cost-Sensitive Deep Learning | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8338085

[203] P. Jonnalagedda, D. Schmolze, and B. Bhanu, "[Regular Paper] MVPNets: Multi-viewing Path Deep Learning Neural Networks for Magnification Invariant Diagnosis in Breast Cancer," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, Oct. 2018, pp. 189–194. doi: 10.1109/BIBE.2018.00044.

[204] Z. Wang, L. Zhang, X. Shu, Q. Lv, and Z. Yi, "An End-to-End Mammogram Diagnosis: A New Multi-Instance and Multiscale Method Based on Single-Image Feature," *IEEE Trans. Cogn. Dev. Syst.*, vol. 13, no. 3, pp. 535–545, Sep. 2021, doi: 10.1109/TCDS.2019.2963682.

[205] M. Mahrooghy *et al.*, "Pharmacokinetic Tumor Heterogeneity as a Prognostic Biomarker for Classifying Breast Cancer Recurrence Risk," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1585–1594, Jun. 2015, doi: 10.1109/TBME.2015.2395812.

[206] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Sci. Rep.*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.

[207] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with Deep Learning," *Sci. Rep.*, vol. 8, no. 1, p. 4165, Mar. 2018, doi: 10.1038/s41598-018-22437-z.

[208] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health Inf. Sci. Syst.*, vol. 6, no. 1, p. 18, Sep. 2018, doi: 10.1007/s13755-018-0057-x.

[209] F. Riaz, F. Abid, I. U. Din, B.-S. Kim, A. Almogren, and S. U. Durar, "Identification of Secondary Breast Cancer in Vital Organs through the Integration of Machine Learning and Microarrays," *Electronics*, vol. 11, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/electronics11121879.

[210] M. Obayya *et al.*, "Hyperparameter Optimizer with Deep Learning-Based Decision-Support Systems for Histopathological Breast Cancer Diagnosis," *Cancers*, vol. 15, no. 3, p. 885, Jan. 2023, doi: 10.3390/cancers15030885.

[211] M. M. Srikantamurthy, V. P. S. Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park, "Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning," *BMC Med. Imaging*, vol. 23, no. 1, p. 19, Jan. 2023, doi: 10.1186/s12880-023-00964-0.

[212] "GARL-Net: Graph Based Adaptive Regularized Learning Deep Network for Breast Cancer Classification | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 16, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10025729

[213] C. Lyle, M. Rowland, W. Dabney, M. Kwiatkowska, and Y. Gal, "Learning Dynamics and Generalization in Reinforcement Learning," arXiv.org. Accessed: Nov. 16, 2024. [Online]. Available: https://arxiv.org/abs/2206.02126v1

[214] K. Kawaguchi, L. P. Kaelbling, and Y. Bengio, "Generalization in Deep Learning," Aug. 22, 2023, *arXiv*: arXiv:1710.05468. doi: 10.48550/arXiv.1710.05468.

[215] A. Vakanski and M. Xian, "Evaluation of Complexity Measures for Deep Learning Generalization in Medical Image Analysis," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, Gold Coast, Australia: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/MLSP52302.2021.9596501.

[216] E. Yagis, A. G. S. De Herrera, and L. Citi, "Generalization Performance of Deep Learning Models in Neurodegenerative Disease Classification," in *2019 IEEE*

International Conference on Bioinformatics and Biomedicine (BIBM), Nov. 2019, pp. 1692–1698. doi: 10.1109/BIBM47256.2019.8983088.

[217]   J. Zhang, T. Liu, and D. Tao, "An Optimal Transport Analysis on Generalization in Deep Learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, pp. 1–12, Sep. 2021, doi: 10.1109/TNNLS.2021.3109942.

[218]   H.-T. Cheng *et al.*, "Wide & Deep Learning for Recommender Systems," arXiv.org. Accessed: Nov. 16, 2024. [Online]. Available: https://arxiv.org/abs/1606.07792v1

[219]   A. G. Wilson and P. Izmailov, "Bayesian Deep Learning and a Probabilistic Perspective of Generalization," Mar. 30, 2022, *arXiv*: arXiv:2002.08791. doi: 10.48550/arXiv.2002.08791.

[220]   A. Boyd, K. Bowyer, and A. Czajka, "Human-Aided Saliency Maps Improve Generalization of Deep Learning," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA: IEEE, Jan. 2022, pp. 1255–1264. doi: 10.1109/WACV51458.2022.00132.

[221]   D. Coccomini, N. Messina, C. Gennaro, and F. Falchi, "Combining EfficientNet and Vision Transformers for Video Deepfake Detection," Jan. 20, 2022, *arXiv*: arXiv:2107.02612. doi: 10.48550/arXiv.2107.02612.

[222]   N. Mungoli, "Adaptive Feature Fusion: Enhancing Generalization in Deep Learning Models," Apr. 04, 2023, *arXiv*: arXiv:2304.03290. doi: 10.48550/arXiv.2304.03290.

[223]   "Deep Learning Approach for Software Maintainability Metrics Prediction." Accessed: Nov. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/332665366_Deep_Learning_Approach_for_Software_Maintainability_Metrics_Prediction

[224]   Q. H. Doan, T. Le, and D.-K. Thai, "Optimization strategies of neural networks for impact damage classification of RC panels in a small dataset," *Appl. Soft Comput.*, vol. 102, p. 107100, Apr. 2021, doi: 10.1016/j.asoc.2021.107100.

[225]   M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, p. 102026, Jan. 2022, doi: 10.1016/j.compmedimag.2021.102026.

[226]   S. Rasheed, K. Rehman, and M. S. H. Akash, "An insight into the risk factors of brain tumors and their therapeutic interventions," *Biomed. Pharmacother. Biomedecine Pharmacother.*, vol. 143, p. 112119, Nov. 2021, doi: 10.1016/j.biopha.2021.112119.

[227]   J. E. Villanueva-Meyer, M. C. Mabray, and S. Cha, "Current Clinical Brain Tumor Imaging," *Neurosurgery*, vol. 81, no. 3, pp. 397–415, Sep. 2017, doi: 10.1093/neuros/nyx103.

[228]   "Noninvasive Classification of Glioma Subtypes Using Multiparametric MRI to Improve Deep Learning." Accessed: Nov. 16, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/12/12/3063

[229]   A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, "Vision Transformers in medical computer vision—A contemplative retrospection," *Eng. Appl. Artif. Intell.*, vol. 122, p. 106126, Jun. 2023, doi: 10.1016/j.engappai.2023.106126.

[230]   J. Cheng *et al.*, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation," *PLOS ONE*, vol. 11, no. 6, p. e0157112, Jun. 2016, doi: 10.1371/journal.pone.0157112.

[231]   "datasets." figshare, Oct. 05, 2017. doi: 10.6084/m9.figshare.5472970.v1.

[232]   "Brain Tumor Classification (MRI)." Accessed: Nov. 16, 2024. [Online]. Available: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

[233] "Brain MRI Images for Brain Tumor Detection." Accessed: Nov. 16, 2024. [Online]. Available: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

[234] "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers." Accessed: Nov. 16, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/21/6/2222

PLAGIARISM VERIFICATION

Title of the Thesis: "Design and Evaluation of Image Processing Model Using Machine Learning Approaches"

Total Pages: 236

Name of the Scholar: Allam Venkata Siva Swetha

Supervisor: Prof. Kapil Sharma

Department: Information Technology

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software used: Turnitin Similarity Index: 9% Word Count: 59,121 Words

Date: 16/12/2024

(Prof. Kapil Sharma)

Candidate's Signature   Signature of Supervisor

Thesis print.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:74200402

Submission Date

Dec 16, 2024, 2:13 PM GMT+5:30

Download Date

Dec 16, 2024, 2:24 PM GMT+5:30

File Name

Thesis print.pdf

File Size

3.9 MB

236 Pages

59,121 Words

354,588 Characters

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Cited Text
- Small Matches (less than 10 words)
- Crossref database
- Crossref posted content database

Match Groups

319 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

2 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% 🌐 Internet sources

2% 📖 Publications

5% 👤 Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

> Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.
>
> A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.