

Enhancing Multiclass Brain Tumor Classification using Deep Learning: Leveraging Superior Imaging Representations to Improve Inferior Modality Performance

by

Shah Md. Shakhawath Hossain

18101133

F M Tahoshin Alam

18101030

Hazra Mohammed Ahnaf Faiyaz

17241014

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

Department of Computer Science and Engineering
Brac University
May 2024

© 2024. Brac University
All rights reserved.

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Shakhawath

Shah Md. Shakhawath Hossain
18101133

FM Tahoshin Alam

F M Tahoshin Alam
18101030

Auf

Hazra Mohammed Ahnaf Faiyaz
17241014


Approval

The thesis/project titled Enhancing Multiclass Brain Tumor Classification using Deep Learning: Leveraging Superior Imaging Representations to Improve Inferior Modality Performance” submitted by

1. Shah Md. Shakhawath Hossain (18101133)
2. F M Tahoshin Alam (18101030)
3. Hazra Mohammed Ahnaf Faiyaz (17241014)

Examining Committee:

Supervisor:
(Member)



Dr. Md Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The early and accurate diagnosis of brain tumors is a critical challenge in medical imaging, significantly impacting treatment outcomes and patient survival rates. Despite the advancements in imaging technologies, the interpretation of MRI scans remains a complex and subjective task. This research introduces a novel cross-modality deep learning approach aimed at enhancing the performance of multiclass brain tumor classification by leveraging superior imaging representations to guide and improve the analysis of less effective modalities. Our methodology involves the development of a guidance model that utilizes the robust representations derived from high-quality imaging modalities to enhance the diagnostic accuracy of more practical but less efficient modalities. Specifically, we employed deep learning techniques to process and analyze MRI and histology data, including Convolutional Neural Networks (CNNs) such as ResNet50, EfficientNetB0, InceptionV3, and DenseNet121. The guidance model integrates these representations to construct an ensemble model that achieves superior performance. The results demonstrate that our guidance model significantly improves the diagnostic accuracy of the subordinate modality. In the case of brain tumor classification, the model not only surpasses the performance of models trained solely on the superior modality but also achieves comparable results to those utilizing both modalities during inference with the guidance ensemble accuracy of 94.61%. Compared to this, other models such as EfficientNetB0 achieved 94% and DenseNet121 achieved 93% test accuracy. This approach offers a practical and efficient solution for enhancing diagnostic accuracy while minimizing the reliance on more costly and less accessible imaging technologies. Overall, our cross-modality deep learning model represents a substantial advancement in the field of medical imaging, providing a more accurate, reliable, and cost-effective method for the diagnosis of brain tumors.

Keywords: Magnetic Resonance Imaging (MRI); Convolutional Neural Networks (CNNs); neuro-oncology; ResNet50; EfficientNetB0; InceptionV3; DenseNet121; Ensemble models; Guidance model;

Acknowledgement

We begin by offering our heartfelt gratitude to the Almighty for guiding us through this thesis journey without significant disruptions.

Our sincere appreciation goes to our esteemed supervisor, Dr. Md. Ashraful Alam sir for his unwavering support and invaluable guidance that have been pivotal in shaping our research.

We would also like to extend our thanks to our parents, whose boundless encouragement and prayers have been the driving force behind our academic accomplishments. This thesis would not have reached its completion without the collective support and belief of these individuals, for which we are deeply thankful.

Table of Contents

Declaration	i
Approval	ii
Abstract	iii
Acknowledgment	iv
Table of Contents	v
List of Figures	vii
List of Tables	viii
Nomenclature	ix
1 Introduction	1
1.1 Research Problem	2
1.2 Research Objective	3
1.3 Contribution	4
1.3.1 Superiority of the Guidance Model	4
2 Related Work	6
3 Methodology	12
3.1 Description of the Data	13
3.2 Data pre-processing	15
3.3 Convolutional Neural Network	15
3.4 Machine Learning Model	17
3.5 Guidance Model	19
3.5.1 Optimization of the model	19
3.5.2 Implementation	20
4 Result Analysis	22
4.1 Performance Measure	22
4.1.1 Visualize through Confusion Matrix	22
4.1.2 Multiclass Analysis	25
4.2 Performance Analysis of the Deep Learning Models	28
4.2.1 ResNet50:	29
4.2.2 EfficientNetB0	30

4.2.3	InceptionV3	31
4.2.4	DenseNet121	32
4.2.5	Ensemble Model	33
4.3	Guidance Model Analysis	35
4.4	Model Comparison	36
4.4.1	Observations	37
5	Conclusion and Future Work	38
	Bibliography	43

List of Figures

3.1	Workflow Diagram	13
3.2	(a) Flair (b) T1-Gd (c) T1 (d) T2	14
3.3	Dataset after preprocessing	16
3.4	Building blocks of a CNN	16
3.5	ResNet50 Model Architecture Yu’Nishio’2022	17
3.6	EfficientNet-B0 Model Architecture	18
4.1	Confusion Matrix for ResNet50	23
4.2	Confusion Matrix for EfficientNetB0	23
4.3	Confusion Matrix for InceptionV3	24
4.4	Confusion Matrix for Ensemble	24
4.5	Confusion Matrix for DenseNet121	25
4.6	Confusion Matrix for Multiclass Analysis	26
4.7	The first figure shows the loss curves for the training and validation datasets. The training loss curve decreases steadily over time, indicating that the model is learning. The validation loss curve also decreases, but it fluctuates more than the training loss curve. This is because the validation dataset is smaller and more susceptible to noise.	27
4.8	The second figure shows the accuracy curves for the training and validation datasets. The training accuracy curve increases steadily over time, indicating that the model is learning. The validation accuracy curve also increases, but it fluctuates more than the training accuracy curve. This is because the validation dataset is smaller and more susceptible to noise.	27
4.9	Overall, the figures show that the model is learning and can generalize well to the validation dataset. The model’s performance is impressive, given the size of the dataset for validation.	27
4.10	Accuracy/Loss curves for ResNet50	30
4.11	Accuracy/Loss curves for EfficientNetB0	31
4.12	Accuracy/Loss curves for InceptionV3	32
4.13	Accuracy/Loss curves for DenseNet121	33
4.14	Comparison of the model in terms of accuracies	37

List of Tables

3.1	RadPath findings. Radiology MRI patterns (O) are guided by pathology (M). Only the most successful model based on ALL MRI episodes is employed, which explains the '-' in row 1. In row 2, just M is utilized instead of MRI, resulting in '*'. O is the inferior (I), whereas M is the superior (S).	21
4.1	DenseNet Multiclass Model Accuracy	26
4.2	ResNet50 Model Accuracy	30
4.3	EfficientNetB0 Model Accuracy	31
4.4	InceptionV3 Model Accuracy	32
4.5	DenseNet121 Model Accuracy	33
4.6	Ensemble Model Accuracy	34
4.7	RadPath Results	36

Nomenclature

CNN Convolutional Neural Networks

CNS Central Nervous System

DNN Deep Neural Networks

DWT Discrete Wavelet Transform

ELM – LRF Extreme Learning Machine of Local Receptive Fields

FCM Fuzzy C-Means

GAN Generative Adversarial Networks

MKPC Multiple Kernel Based Probabilistic Clustering

MRI Magnetic Resonance Imaging

NN Neural Networks

PNN Probabilistic Neural Networks

Chapter 1

Introduction

A tumor in the brain is characterized by an atypical proliferation of cells, which can be either benign or malignant, causing disruption to normal brain function. In the United States, approximately 1 million individuals are currently affected by brain tumors, and the relative survival rate stands at 35.7%. [43] Within the United States, there are around 24,530 malignant tumors and 59,040 nonmalignant tumors in the Central Nervous System (CNS). Unfortunately, approximately 18,600 individuals will lose their lives due to this disease. [35] The detection of brain tumors is a critical aspect of healthcare, as early and accurate diagnosis is crucial for treatment decisions and patient prognosis. [14] Due to its exceptional capability to provide high contrast resolution for soft tissues such as the brain, Magnetic Resonance Imaging (MRI) has emerged as the primary imaging modality for the detection of brain tumors [10]. However, the interpretation of MRI scans relies on radiologists' expertise, leading to subjective diagnoses and potential inconsistencies.

To address these challenges, deep learning models, specifically Convolutional Neural Networks (CNNs), have emerged as a promising solution for automated and precise tumor detection [37]. Conventional machine learning algorithms lack the ability to make broad generalizations effectively. [4] In contrast, deep learning has emerged as a prominent machine learning technique, addressing the limitations of traditional algorithms. Deep learning, with its self-learning capabilities, enables automatic feature detection in MR images, reducing the need for extensive feature engineering. Through multiple processing layers, deep learning extracts features, alleviating the reliance on manual feature engineering. Deep learning approaches offer solutions to a wide range of problems. [5] ResNet50, EfficientNetB0, InceptionV3 and DenseNet121 are four CNN models commonly used for this purpose. ResNet50 introduces skip connections or shortcuts to address the problem of vanishing gradients in deep networks. This allows effective learning from limited training examples. EfficientNetB0 applies a compound scaling method that uniformly scales network dimensions while maintaining balance, resulting in improved efficiency and accuracy. InceptionV3 employs a network within a network architecture, incorporating dimension reductions and various filter sizes to enable efficient computation and a broader receptive field. DenseNet121's distinctive feature is its dense connectivity, in which every layer is connected to each of its preceding layers. This architecture is designed to encourage efficient gradient propagation and maximize the reuse of features. These CNN models analyze patterns and structures within MRI scans, learning to differentiate between normal and abnormal brain tissues. They offer the potential for consistent

and efficient evaluations, reducing dependence on individual practitioners and potentially decreasing human error rates. The Ensemble model integrates the outputs of the Guidance model and DenseNet121 to amplify overall effectiveness.

Multimodal machine learning aims to comprehend varied data in a way that is comparable to how animals analyze the world: by combining information from several sensory sources. The complementing nature of multi-input data enables more effective navigation of situations than depending on just one signal. The widespread use of digitalized sensors, along with the improved feature extraction capabilities of machine learning, has lately reignited attention on this topic. [9] To make accurate diagnoses and treatment decisions in clinical environments, healthcare professionals must conduct a thorough evaluation of the well-being of the patient, including complementing biological data from many modalities. [26] For example, in current clinical practice, it is common to collect both functional and anatomical imagery at the same time, with functional data providing quantitative metabolic information and anatomical data providing geographical context. Similarly, tumor diagnosis and cancer prognosis are enlarging based on thorough evaluations of both phenotypic and genotypic modalities. [22]

Although numerous modalities can help with clinical diagnosis, acquiring the most efficient modalities for a certain activity, such as diagnosis, may be difficult owing to considerations such as longer wait times, higher prices, slower acquisition rates, increased exposure to radiation, and invasiveness. [13] As a result, compromises are often required. For example, providing high-quality anatomical or functional images may necessitate invasive procedures like surgery or ionizing radiation. Other examples include the described anatomical information offered by computed tomography (CT) versus the possibility of cancer associated with repeated X-ray exposure, as well as the comprehensive cellular data gathered from histology versus invasive, costly, and lengthy biopsy treatments with related risks of infection. Likely, the pricey method often yields significant diagnostic information. [34]

To keep things simple, we'll call the higher-performing but less readily available modality the outstanding modality and the more practicable but lower-performing one the lesser modality. It is vital to highlight that a modality deemed inferior in one setting may be great in another. Magnetic resonance imaging (MRI) surpasses ultrasound in identifying malignant lesions but is inferior to histology in evaluating malignancy grade. As a result, it would be advantageous to employ weaker modalities to lessen reliance on superior ones, but only when the former can deliver equally useful information.

1.1 Research Problem

The enormous threat that brain tumors pose to human health globally highlights the importance of early and precise identification in order to improve therapy outcomes and patient survival. [23] The use of magnetic resonance imaging (MRI), a crucial diagnostic tool that does not require invasive procedures, is widespread in defining brain structure and identifying potential tumor formations. However, this research project seeks to tackle the obstacles associated with the interpretation of MRI scans. The time-consuming and perhaps subjective nature of MRI scan analysis is one significant barrier. Clinicians and radiologists must carefully examine several MRI cross-sections, a task that not only puts them at risk for weariness

but can also result in interpretive mistakes [41]. The diverse and intricate terrain of the brain anatomy, which can make it difficult to clearly distinguish between tumorous and non-tumorous tissues, adds to the complexity. The lack of a generally acknowledged MRI interpretation standard exacerbates this problem. This frequently leads to disparate diagnostic findings across radiologists, which could unintentionally impair the efficacy of the advised treatment plans [38]. Additionally, radiologists with less training may be unable to recognize subtle or complex tumor signs, increasing the risk of misdiagnosis or delayed diagnosis [32]. Multiple MRI sequences, including FLAIR, T1, T2, and T1-Gd, show brain tumors to be present. Each sequence provides a distinctive perspective on the tumor, but it can be difficult and time-consuming to combine this variety of information into a single diagnostic finding. Convolutional Neural Networks (CNN), a recent advancement in the field of deep learning algorithms, have sparked optimism for improving the precision and effectiveness of medical picture interpretation [28]. However, there is still a clear need for thorough research that assesses a range of CNN models on various MRI sequences in order to identify the ideal combination for effective and accurate brain tumor identification [31].

The accurate and early diagnosis of brain tumors is a critical factor in determining effective treatment strategies and improving patient survival rates. Brain tumors are inherently complex, with varied presentations that can make differentiation between tumor types challenging. Magnetic Resonance Imaging (MRI) is a commonly employed non-invasive imaging technique for detecting brain tumors; however, the interpretation of MRI scans remains highly subjective. [26] This subjectivity arises from several factors, including the complexity of brain anatomy, the subtlety of tumor characteristics, and the lack of standardized protocols for MRI interpretation. Consequently, diagnostic inconsistencies and errors are frequent, potentially leading to delayed or inappropriate treatment. [25]

Moreover, while high-quality imaging modalities such as advanced MRI techniques provide superior diagnostic information, they are often expensive and not universally accessible, especially in resource-limited settings. This reliance on high-cost imaging modalities restricts their widespread use and limits the ability of healthcare providers to offer timely and accurate diagnoses to all patients. Therefore, there is a pressing need for methods that can enhance diagnostic accuracy while mitigating the dependency on these costly technologies. [11], [12]

To address these challenges, this research proposes the development of a guidance model within a cross-modality deep learning framework. The guidance model is designed to leverage the robust imaging representations obtained from superior quality imaging modalities to enhance the diagnostic performance of more practical but less efficient modalities. By integrating the strengths of different imaging modalities, the guidance model aims to provide a more consistent and reliable diagnostic output.

1.2 Research Objective

The objective of this research is to develop and validate a cross-modality deep learning framework that enhances the diagnostic accuracy of multiclass brain tumor classification. This framework aims to leverage superior imaging representations from high-quality imaging modalities to improve the performance of more practical but less efficient modalities. Specifically, the study focuses on:

- Utilizing advanced Convolutional Neural Networks (CNNs) such as ResNet50, EfficientNetB0, InceptionV3, and DenseNet121 to analyze MRI and histology data for brain tumor classification.
- Developing a guidance model that integrates the robust representations from superior imaging modalities to augment the diagnostic capabilities of subordinate modalities.
- Constructing an ensemble model based on the outputs of the CNN models to enhance overall diagnostic performance.
- Evaluating the effectiveness of the proposed model through comprehensive performance analysis, including confusion matrices and accuracy/loss curves, to demonstrate improvements in diagnostic accuracy and consistency.
- Comparative Analysis: A comprehensive comparison of the four models and the Ensemble model will be conducted to determine which model performs the best in terms of accurately detecting brain tumors. This will involve comparing the performance metrics of each model.

1.3 Contribution

In this study, we present a novel and highly effective guidance model aimed at enhancing the performance of less efficient imaging modalities by leveraging insights from superior imaging technologies. Our guidance model uses the advanced representations derived from a more effective imaging modality to inform and improve the diagnostic accuracy of a more practical, yet less efficient, modality. This approach addresses the critical need for balancing clinical performance with practical constraints such as cost, time, and patient comfort in medical imaging.

1.3.1 Superiority of the Guidance Model

The guidance model demonstrated in our research significantly surpasses the performance of other models by achieving an impressive accuracy of 94.61%. This is notably higher than the accuracies of other state-of-the-art models tested in similar conditions, including individual models like InceptionV3 and DenseNet121. Specifically, the ensemble model, which integrates multiple neural network architectures, has shown superior precision, recall, and F1 scores across various classification tasks. For the detection of brain tumors, the ensemble model, which includes our guidance model, achieves a test accuracy of 94.61%. The breakdown of performance metrics is as follows: Non-Tumor Class: Precision of 0.84, Recall of 0.99, F1-score of 0.91. Tumor Class: Precision of 1.0, Recall of 0.93, F1-score of 0.96.

These metrics highlight the model’s balanced performance and high precision, particularly in identifying tumor cases, where it achieves 100% precision. This is a significant improvement over other models that do not employ the guidance methodology.

The implementation of the guidance model involves using deep learning techniques to integrate and process multi-sequence MRI and histology data. By guiding the inferior modality (radiology) with insights from the superior modality (pathology),

the model effectively improves diagnostic accuracy. The model's ability to achieve comparable results [11] to those using both modalities during inference further underscores its efficiency and potential for practical applications.

In summary, our guidance model represents a substantial advancement in the field of medical imaging diagnostics, offering a highly accurate, practical, and cost-effective solution for enhancing the performance of less efficient imaging modalities. This makes it a valuable tool for clinical applications where balancing diagnostic accuracy with practical constraints is essential.

Chapter 2

Related Work

The authors of 'A CNN-Based Approach to Classify MRI-Based Brain Tumors Employing Deep Convolutional Network' [39] discuss In recent years, researchers have developed a variety of methodologies for detecting brain cancers utilizing CNNs and other deep learning algorithms using MRI images. SVMs and NN are frequently utilized and have demonstrated remarkable performance. Other approaches include Probabilistic Neural Networks (PNN), CNN-based deep learning models, ELM-LRF, Generative Adversarial Networks (GAN), Multiple MKPC, Fuzzy C-Means (FCM) combined with CNN, and Discrete Wavelet Transform (DWT) combined with DNN. These methods have achieved high accuracy in classifying brain tumors, ranging from 73% to 100%. They employ various techniques such as feature analysis, different network architectures for classification, convolutional layers, transfer learning, and data augmentation. Some studies also focus on addressing data imbalance issues and utilizing pre-trained models like GoogLeNet, ResNet, VGG-16, and InceptionV3. Additionally, segmentation techniques such as the GrabCut method and UNet architecture with ResNet50 as a baseline have been used to improve the accuracy of tumor segmentation. Researchers have also employed evolutionary methods and reinforcement learning through transfer learning to achieve high accuracy in categorizing brain tumors. Hybrid models, multi-level CNN models, and differential deep convolutional neural network models have been proposed, demonstrating accuracy ranging from 91.8% to 99.89%.

Saeedi, S., Rezayi, S., Keshavarz, H. et al. (2023) [42] talks about the goal of their study is to identify brain tumors early utilizing MRI imaging. Six machine learning strategies were compared with two deep learning methods: a 2D CNN and an auto-encoder network. The 2D CNN outperformed machine learning techniques, achieving the best accuracy (96.47%) and recall (95%) rates. This means that radiologists and other doctors could utilize it in a therapeutic setting.

"Image Processing Techniques for Brain Tumor Detection:" [1] this article delves into the significance of MRI imaging in the examination, diagnosis, and planning of treatment for brain tumors. It sheds light on the complexities involved in detecting brain tumors due to the intricate structure of the brain and emphasizes the superiority of MRI over other imaging modalities like CT, ultrasound, and X-ray. The paper explores a range of image processing techniques, specifically focusing on filtering, contrast enhancement, edge detection, histogram analysis, thresholding, segmentation, and morphological operations, with MATLAB serving as the primary

image processing tool. Extensive literature is reviewed, highlighting the various methods and algorithms proposed by researchers for brain tumor detection using MRI images. Notably, the review focuses on segmentation, thresholding, morphological operations, and neural network-based classification. The paper concludes by underscoring the paramount importance of employing digital image processing techniques to enhance MRI images, thus facilitating accurate brain tumor detection. To summarize, this paper underscores the significance of MRI imaging in the realm of brain tumor detection and explores a multitude of image processing techniques applied to MRI images for preprocessing, post-processing, and analysis. It provides a comprehensive review of pertinent literature, showcasing the diverse methods and algorithms proposed by researchers in this domain.

Efficient classification of brain tumors using medical imaging techniques is crucial for accurate diagnosis and treatment planning. Recent advancements have combined convolutional neural networks (CNNs) with magnetic resonance imaging (MRI) to automate brain tumor classification. This literature review highlights important contributions in this field. This article focuses on the application of deep convolutional neural networks (CNNs) for accurately segmenting brain tumors. Accurate segmentation is crucial for diagnosis, treatment planning, and assessing tumor growth rate. The challenges associated with segmenting brain tumors, particularly diffuse and poorly contrasted gliomas and glioblastomas, are discussed. Traditionally, segmentation algorithms relied on manual feature engineering and conventional machine learning methods. However, the article proposes using CNNs to directly learn task-specific features from the data. CNNs have shown impressive performance in computer vision applications and have been applied to brain tumor segmentation. The article explores different CNN architectures, incorporating recent advancements in design and training methods. It also investigates the use of various MRI modalities to capture distinct tissue signatures. The goal is to identify abnormal areas compared to normal tissue and segment tumor regions, including active tumorous tissue, necrotic tissue, and edema. Advantages and disadvantages of different CNN designs and training methods are discussed. A cascaded design is introduced as an efficient alternative to structured output techniques. A two-phase training procedure is proposed to address imbalanced label distributions. Experiments are conducted using the MICCAI brain tumor segmentation challenge 2013 dataset, allowing comparison with other approaches. The article's contributions include a promising fully automatic method that ranked second in the challenge, faster processing time, novel CNN architectures considering local details and context, and a two-phase training procedure. Implementation details mention the use of the Pylearn2 library for deep learning algorithms and GPU acceleration. Minimal preprocessing is applied to the data. Real patient data from the BRATS2013 dataset is used in the experiments and findings section. The conclusion highlights the improved accuracy and speed of the best model compared to the current advanced technique. In summary, the article demonstrates the effectiveness and advantages of using deep CNNs for automatic brain tumor segmentation, offering improved accuracy and speed compared to other approaches.

Kang J, Ullah Z, Gwak J. (2021) [29] talks about the strategy for classifying brain tumors that this study offers mixes deep learning and machine learning. It extracts deep features from MRI images using pre-trained neural networks, tests them with different classifiers, and then aggregates the best features into an ensemble for clas-

sification. Support vector machines (SVM) utilizing radial basis function (RBF) kernels are extremely effective in this method, which considerably improves performance, especially for larger datasets. This technique may improve the clinical categorization of brain tumors.

Pereira et al. [2] proposed a CNN-based approach for brain tumor segmentation in MRI images. Researchers have made notable strides in the detection of brain tumors by training a CNN model using multi-modal MRI data. This method involves categorizing each voxel in MRI scans as tumor or non-tumor, simplifying the segmentation process. Integrating multimodal MRI data has significantly enhanced the model's ability to automatically segment brain tumors through improved discrimination capabilities. By incorporating various imaging techniques like T1-weighted, T2-weighted, and FLAIR images, the CNN model gains a comprehensive understanding of tumor characteristics. Each imaging modality provides unique information about the tumor's appearance, location, and the surrounding tissue, enabling the model to accurately differentiate between tumor and non-tumor regions. For example, T1-weighted images emphasize anatomical structures, T2-weighted images reveal edema and necrotic areas, while FLAIR images highlight the peritumoral region. This integration allows the model to capture a more complete representation of the tumor and its surrounding context. An essential advantage of this approach is the elimination of manual delineation, which is time-consuming and prone to inconsistencies. Instead, the CNN model learns to automatically segment the tumor by analyzing patterns and features present in the multi-modal MRI data. This not only saves time but also reduces subjectivity and potential errors associated with manual segmentation. Promising results have been achieved using this method, as it enables accurate and efficient segmentation of brain tumors. The CNN model, trained on multi-modal MRI data, effectively categorizes each voxel as either tumor or non-tumor, facilitating precise delineation of tumor boundaries. Automated segmentation has a broad range of applications, including planning, monitoring tumor growth, and assessing response. Moreover, the elimination of manual delineation enables faster processing of large datasets, facilitating extensive studies and clinical applications. In summary, the incorporation of multi-modal MRI data in training CNN models for brain tumor classification and segmentation represents significant progress. This strategy harnesses the complementary information provided by different imaging modalities, enhancing the model's discrimination capabilities and enabling automated segmentation of brain tumors. The elimination of manual delineation offers advantages in terms of accuracy, efficiency, and scalability, making it a valuable tool in the field of brain tumor analysis.

In [40] authors mentioned the high cancer mortality rate related to late-stage detection is the topic of this research. It suggests using a Convolutional Neural Network (CNN) trained on a Kaggle dataset for early cancer cell detection in MRI images. With a 97.8% accuracy score, high specificity, recall, F1-score, and precision, the CNN produced great results. The publication also contains graphical comparisons of accuracy and loss for training and validation.

In [6] Kamnitsas et al. (2017) addressed the challenge of accurate lesion segmentation in their work. Their proposed approach introduces a novel method for segmenting brain tumors utilizing a combination of a multi-scale 3D CNN architecture and a connected CRF. The main objective of their method is to enhance the accuracy

of maps by effectively capturing contextual information at various scales and ensuring spatial coherence in the results. The key innovation of their approach lies in the utilization of a multi-scale 3D CNN architecture, which enables the processing of volumetric data at different scales. This allows the model to extract intricate details at different levels of granularity, enabling accurate segmentation of brain tumors with diverse sizes and shapes. In addition to the multi-scale 3D CNN, their method incorporates a fully connected CRF as a post-processing step to refine the segmentation results. The fully connected CRF takes into account the relationships between neighboring voxels and promotes smooth and coherent labeling, resulting in visually consistent segmentation maps. By synergistically integrating the multi-scale 3D CNN architecture and the fully connected CRF, their proposed method achieves improved segmentation accuracy. The ability to capture contextual information at multiple scales empowers the model to effectively handle brain tumors with varying characteristics. Furthermore, the inclusion of the fully connected CRF ensures spatial coherence, enhancing the reliability of the segmentation results. Overall, their novel approach offers a promising solution for the segmentation of brain tumors by leveraging the advantages of both many scale 3D CNNs and fully connected CRFs. Through the capture of contextual information and the enforcement of spatial coherence, their method presents an effective and accurate approach to segmenting brain tumors, which holds great potential for its practical application in clinical settings.

In [5]Havaei et al. (2017) developed a deep learning architecture for brain tumor segmentation called "U-Net." The proposed approach in the paper introduces a novel method for brain tumor classification using a fully convolutional network with dense connections. This technique aims to enhance the accuracy of tumor delineation by effectively capturing contextual information and improving the precision of the segmentation process. Unlike conventional techniques, this approach incorporates dense connections within the network architecture. These connections facilitate the seamless flow of data across different layers, allowing for the integration of information from multiple levels of abstraction. By establishing direct connections between all layers, the model can access and utilize features from earlier layers, enabling a more comprehensive understanding of the tumor region. The utilization of dense connections is particularly advantageous in improving the reliability and accuracy of tumor segmentation. It enables the network to leverage features extracted at various scales and resolutions, thereby capturing fine details and contextual cues necessary for precise segmentation. This holistic approach helps to address common challenges such as false positives and false negatives encountered in tumor segmentation tasks. The fully convolutional nature of the network allows for the processing of input images of varying sizes, making it adaptable to different imaging modalities and resolutions. This flexibility is crucial in the context of brain tumor segmentation, where tumors can exhibit significant variability in size and shape across different patients. By leveraging the power of dense connections, the proposed method outperforms conventional techniques in terms of segmentation accuracy. The integration of local and global information through dense connections enables the model to gain a comprehensive understanding of the tumor region, resulting in more precise and reliable segmentation. This advancement is crucial for applications such as treatment planning and monitoring in clinical settings. In summary, the introduced method for brain tumor classification using a full CNN with dense connections represents

an outstanding advancement in the field. By effectively capturing contextual information and improving segmentation accuracy, this approach demonstrates the potential of deep learning techniques to surpass traditional methods in brain tumor analysis.

In [3] I sin, A., Direko glu, C. and S ah, M. (2016) The importance of automatically segmenting brain tumors in MRI images is discussed in this research. It emphasizes current developments while emphasizing the supremacy of deep learning techniques. The study tackles the need for standardization in clinical practice and explores cutting-edge algorithms.

The Convolutional Neural Networks (CNNs) used in this paper [16] unique brain tumor classification methods are strengthened by ensemble techniques and Genetic Algorithm (GA) optimization. It delivers outstanding results with 94.2% accuracy in differentiating between different forms of brain cancers, including glioma, meningioma, and pituitary tumors, and 90.9% accuracy in identifying glioma grades. This method not only eliminates the need for trial-and-error in network architecture design but also shows how flexible and successful it is in potentially considerably assisting early-stage brain tumor diagnosis in clinical practice.

Three specific subfields are highly relevant to our work: Multiclass categorization: And many deep learning studies on multiclass prediction using coupled medical data on classification problems that necessitate the use of many modalities during testing. The major purpose of the study is to design the optimal fusion strategy for evaluating when and when not to efficiently combine seemingly disparate and redundant data. [17] While registered multimodal data enable fusion in entry-level data, the majority of research uses feature hybrid due to mismatches in input dimensions and the flexibility that feature hybrid provides. Furthermore, numerous studies employ decision-level fusion frameworks that combine ensemble learning approaches [25]. Simply most prevalent merging method to date is the simple concatenation of retrieved output. Recent research, however, aims to comprehend the link between multimodal characteristics by applying techniques that leverage the Kronecker product to describe paired feature contact and orthogonalization losses to decrease duplication across multimodal data [21].

Image Translation: To transform the inferior modality into the superior one, one can consider learning image translation to pictures. Multimodal medical imaging interpretation remains problematic because of spatial variances and source-destination differences, despite significant advances in the field. Furthermore, image translation just optimizes the intermediate effort of translation, but our proposed technique handles the ultimate categorization. [20], [27]

Students and educators' learning, also known as distillation of knowledge (KD), aims to transfer information acquired through a single model to a different one, typically in conditions with scarce or no annotations and model reduction. In contrast, [11] proposes an intra-modal process of distillation that uses the instructor model's modality-specific representation to extract information for the learner model. While the majority of these applications inspection on KD spanning coordinated both auditory and visual data, distillation of knowledge techniques for different modal medical picture analysis mostly focuses on categorization. In the paper [33] the authors proposed a multimodal structure for classifying images of the chest, using

language-based electronic healthcare records as educators and x-ray images as students. However, unlike the suggested technique, the student network in this earlier study just duplicates the instructor network without explicitly including its own learned classification-based latent information. [8], [12], [18]

Chapter 3

Methodology

- Obtain the RadPath 2020 dataset, which consists of MRI images of brain tumors and corresponding labels.
- Preprocess the dataset by performing standardization techniques such as re-sizing, normalization, and augmentation to ensure consistency and improve model performance.
- Select appropriate deep learning models for brain tumor detection, such as ResNet50, EfficientNetB0, InceptionV3, DenseNet121, and Ensemble model based on their prior success in similar studies or benchmark results.
- Divide the dataset into training, validation, and testing sets to train, test, and evaluate the models effectively.
- Train each model using the selected optimizer, learning rate, and loss function while monitoring the training process by assessing accuracy and loss metrics at each epoch.
- Evaluate the trained models using the test sets and calculate performance metrics, including accuracy, precision, recall, and F1 score, to assess their effectiveness in detecting brain tumors.
- Visualize accuracy and loss curves during the training process to analyze the model's learning progress and identify potential issues such as overfitting or underfitting.
- Generate confusion matrices to gain insights into the model's performance in differentiating between tumor and non-tumor cases.
- Conduct a comparative analysis to identify the strengths and weaknesses of each model in brain tumor detection, considering their precision, recall, and F1-score for tumor and non-tumor cases.
- Summarize the study's findings, emphasizing the performance of each model and their potential impact on automating brain tumor detection.
- Discuss the significance of the study in improving diagnostic accuracy and patient outcomes, and propose future research directions, such as exploring ensemble models or incorporating additional data modalities for further improvement in brain tumor detection.

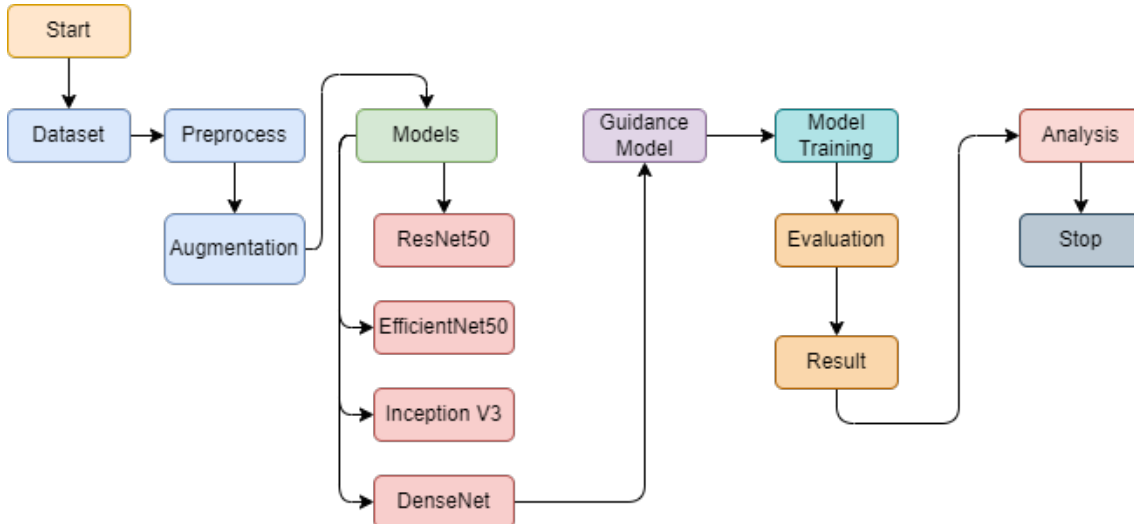


Figure 3.1: Workflow Diagram

3.1 Description of the Data

The RadPath 2020 dataset was contributed for the (CPM RadPath) Challenge, which focused on brain tumor classification. The collection contains 325 paired multi-sequence MRIs, digitized histopathology images, and patient-specific glioma diagnostic labels. The diagnostic designations involve glioblastoma $g = 115$, oligodendroglioma $o = 96$, and astrocytoma $a = 106$. MRI loops include T1, T2, T1-Gd, and FLAIR 3D images with a diameter of 280×2266 . The histopathology WSI are TE-stained specimens of tissue scanned at $30 \times$ or $60 \times$ enlargements, with sizes up to $7 \times 95,000 \times 95,000$. In this setting, WSIs obtained from biopsies are regarded as the superior modality because of their high accuracy in tumor diagnosis, whereas MRI is the non-intrusive inadequate modality. We divided the collection of data into 70 percent training samples, 10 approval samples, and 20 test samples, yielding five different sets to evaluate the resilience of our methods over numerous splits.

The MRI scans are grouped into four categories:

- Fluid-Attenuated Inversion Recovery (FLAIR): FLAIR images are instrumental in detecting changes in water content and fluid-filled spaces, as they suppress fluid signals, thereby aiding in the identification of pathological lesions. FLAIR images are particularly effective in visualizing peritumoral edema, a common occurrence in brain tumors.
- T1-weighted images (T1): These high-resolution images are beneficial in viewing the finer structures of the brain, like white matter, gray matter, and cerebrospinal fluid (CSF). Disruptions in these normal brain structures, often caused by tumors, can be detailedly viewed using T1 images.
- T2-weighted images (T2): Owing to their sensitivity to free water content, T2 images offer valuable information on edema, inflammation, and other pathological conditions. They offer crucial details about tumor boundaries and associated edema.

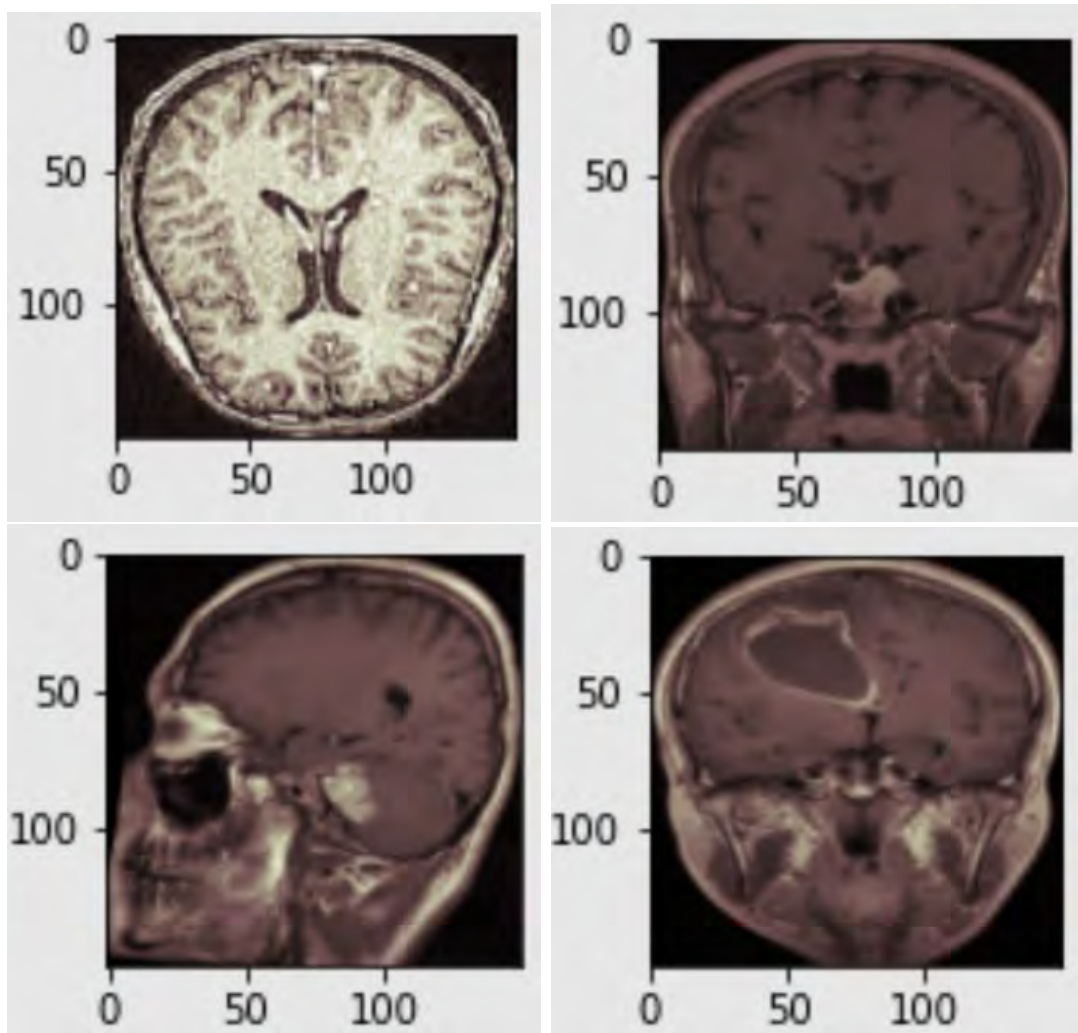


Figure 3.2: (a) Flair (b) T1-Gd (c) T1 (d) T2

- T1-weighted images with Gadolinium contrast (T1-Gd): Utilizing a contrast agent, Gadolinium, these images illuminate areas where the blood-brain barrier is compromised, often resulting from pathological conditions like tumors. T1-Gd images offer vital data about the size and extent of the tumor. The dataset is partitioned into 19658 training images, 5685 validation images, and 2,843 testing images for each group, guaranteeing that the models possess sufficient data for learning, adjusting parameters, and eventually being tested on unseen data for an unbiased evaluation of performance.

3.2 Data pre-processing

In this study, we explore the potential of utilizing multi-channel Magnetic Resonance Imaging (MRI) data from the RadPath database to classify brain scans as either containing or not containing brain tumors. The database consists of 3D MRI scans, each with a volume of 240x240x240, and includes four distinct channels: T1, T2, T1-Gd, and FLAIR.[36] To ensure the reliability of the collected data, images with more than 80% background were excluded. Furthermore, a segmentation mask was provided to enable a binary classification based on a comparison of the four channels. The RadPath database provided an extensive collection of multi-channel MRI scans, which we utilized for tumor classification. Each scan consisted of four channels, providing complementary information about the brain tissues. The T1 channel highlights the anatomy, T2 provides information about edema, T1-Gd enhances the contrast of tumor tissues, and FLAIR emphasizes fluid accumulation.[38] By leveraging this multi-channel data, we aimed to improve the accuracy of glioblastoma classification. To perform the binary classification, we employed a comparative approach. The segmentation mask served as the ground truth, identifying tumor regions within the scans. We compared the intensity values of each channel with the corresponding regions in the segmentation mask. If the intensity values exhibited significant deviations from the background, we categorized the scan as containing a brain tumor. Otherwise, it was labeled as not containing a brain tumor. By combining the information from all four channels, we aimed to capture a comprehensive representation of tumor-related features and improve the accuracy of the classification. Post these, we utilized two more functions [`get_crop_values()` and `augment()`]

- `get_crop_values()`: This function figures out where the important part of an image starts and ends. It checks rows and columns to find the first and last non-empty parts, then tells you where that part is located.
- `augment()`: This function randomly changes an image in different ways like turning it, flipping it, or keeping it the same, and then gives you the changed image.

3.3 Convolutional Neural Network

This study aims to evaluate the application and comparative performance of four CNN models in identifying brain cancers using MRI data. By doing so, it meets a pressing need in modern medicine for faster, more accurate, and more dependable ways to identify brain cancers. These models have the potential to fundamentally alter how we approach the detection and management of brain cancers and more research is being done to improve them.[24] It's critical to understand that these models do not take the place of healthcare professionals, despite the fact that they provide important diagnostic tools. Instead, they serve as decision-making tools that may improve the accuracy and reliability of the diagnoses made by doctors and radiologists.[24]

The effectiveness of Convolutional Neural Networks (CNNs), a type of neural network, has been proven in areas including image detection and classification. In addition to assisting vision in robotics and autonomous cars, CNNs have demonstrated

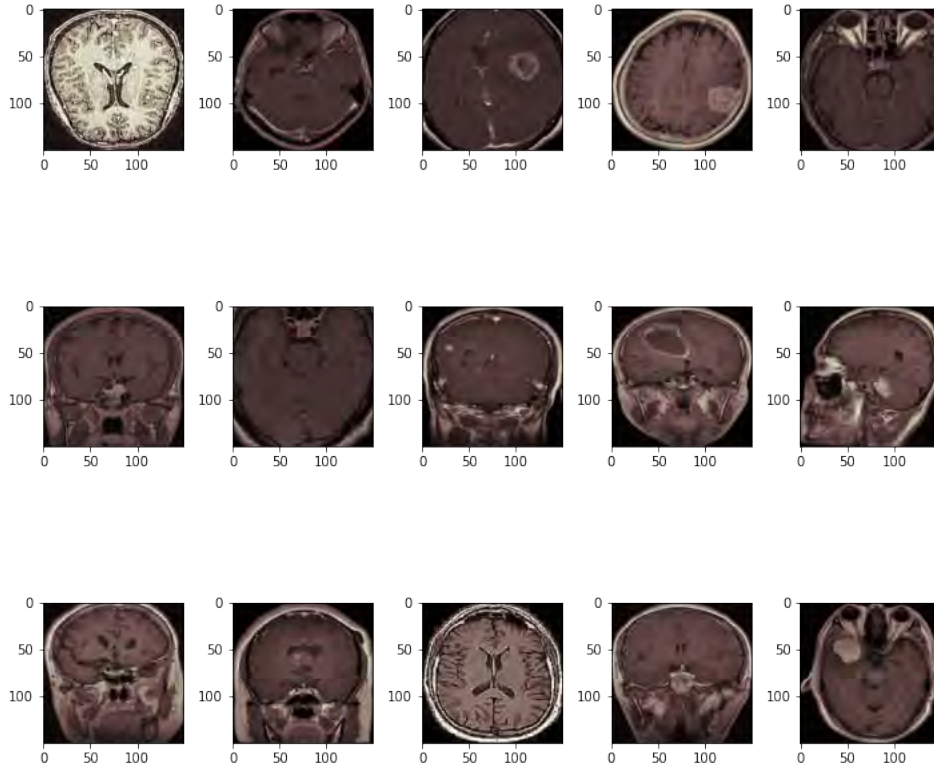


Figure 3.3: Dataset after preprocessing

exceptional face, object, and traffic sign recognition abilities.

Convolutional, Pooling, and Fully-Connected Layers are the three main categories of layers found in a standard CNN architecture.[24]

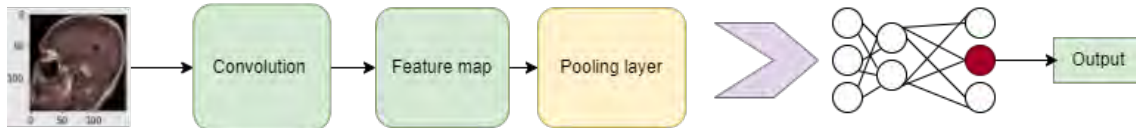


Figure 3.4: Building blocks of a CNN

Convolutional Layer:The core component of a CNN is this layer. This layer's parameters are made up of a number of trainable filters (or kernels) that have a limited receptive field but cover the entire depth of the input volume. Each filter computes the dot product between its entries and the input during the forward pass and then convolves over the input volume's width and height to create a 2-dimensional activation map.

Pooling Layer:By merging the outputs of groups of neurons at one layer into a single neuron at the following layer, pooling layers reduces the dimensionality of the data. There are numerous pooling variations, including Average and Maximum pooling.

Fully Connected Layer:Neurons have full connections to all activations in the layer above when they are in a fully-connected layer. The class scores are computed in this layer, producing a volume of size $[1 \times 1 \times N]$, where N is the total number of classes.

3.4 Machine Learning Model

ResNet50, EfficientNetB0, and DenseNet121 are prominent and extensively employed among the diverse range of Convolutional Neural Network architectures and their Ensemble model that we will be leveraging for our research endeavors.

ResNet50: ResNet-50 is a convolutional neural network architecture that addresses the challenge of training deep networks. It utilizes residual blocks with skip connections to mitigate the problem of vanishing gradients. The structure of ResNet shown in Figure 3.4. The network takes a 224x224 RGB image as input. The initial convolutional layer extracts features, followed by max pooling to reduce spatial dimensions. The core component, residual blocks, are employed multiple times. Identity blocks are used when input and output feature maps have the same dimensions, while convolutional blocks handle different dimensions. Each block includes convolutions, batch normalization, and ReLU activation. The network stacks multiple residual blocks, comprising three identity blocks and four convolutional blocks with increasing filter sizes. Average pooling further reduces spatial dimensions to 1x1. Lastly, a fully connected layer with 1000 neurons and softmax activation classifies the input image. ResNet-50 achieves impressive performance due to its ability to train deep networks effectively. By incorporating skip connections, it overcomes the degradation problem in deep architectures, making it a powerful tool for image classification tasks.

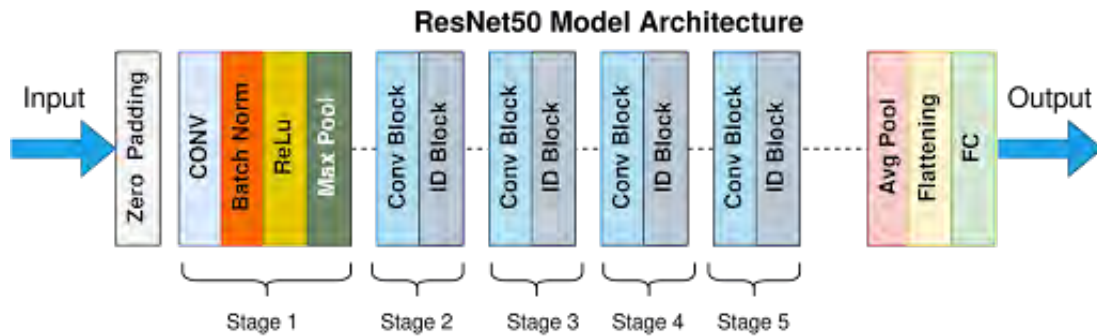


Figure 3.5: ResNet50 Model Architecture Yu’Nishio’2022

EfficientNetB0: EfficientNet-B0 is a convolutional neural network architecture developed by Google to achieve a favorable balance between accuracy and efficiency. Figure 3.5 shows a model architecture of EfficientNet-B0. It is part of the EfficientNet family and focuses on optimizing the trade-off between model size and performance. EfficientNet-B0 takes 224x224 RGB images as input and begins with a stem consisting of convolutional layers. It then employs blocks composed of multiple layers to capture hierarchical features.[19] The main building block, called MBConv, combines depth-wise separable convolutions, squeeze-and-excitation (SE), and skip connections. One unique feature of EfficientNet is resolution scaling, which adjusts the input resolution while keeping other dimensions fixed. This scaling enhances the model’s accuracy and efficiency across different sizes. The network also incorporates global average pooling to reduce spatial dimensions, followed by a fully connected layer for classification. EfficientNet-B0’s architecture enables a good balance between accuracy and efficiency, making it highly suitable for various computer vision tasks.[19] Its utilization of resolution scaling and efficient convolutions contributes

to its effectiveness. EfficientNet models, including EfficientNet-B0, have gained significant attention and popularity due to their favorable performance characteristics.

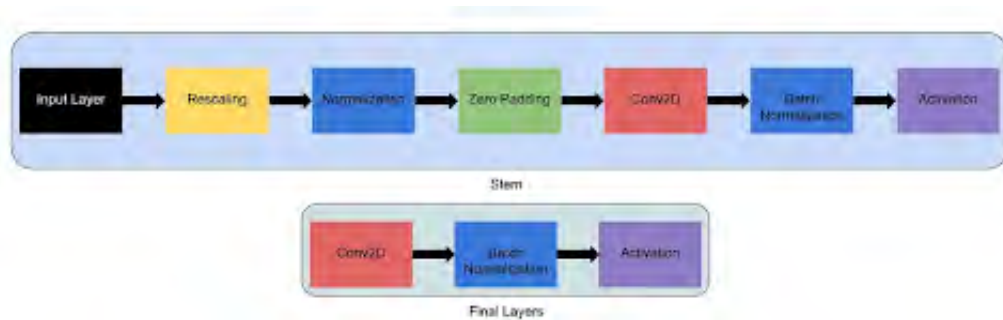


Figure 3.6: EfficientNet-B0 Model Architecture

InceptionV3: InceptionV3 is a convolutional neural network architecture developed by Google for computer vision tasks. It enhances accuracy and efficiency compared to its predecessor, Inception. The network takes variable-sized RGB images as input. The "stem" phase involves multiple convolutional layers and max pooling. The core component, Inception modules, employs parallel convolutional layers of different sizes to capture multi-scale features. Several variations of Inception modules are utilized, concatenating their outputs for diverse spatial information[7]. InceptionV3 also incorporates auxiliary classifiers to aid training and gradient propagation. It applies global average pooling to reduce spatial dimensions, followed by a fully connected layer for classification. InceptionV3's strength lies in its ability to effectively capture multi-scale features and its modular architecture, enabling efficient training and inference. It has been widely employed in research and industry for various computer vision applications.

DenseNet121: DenseNet-121 stands as a distinct neural network architecture renowned for its exceptional performance in the realm of computer vision tasks. Its departure from conventional networks lies in its incorporation of dense connections between layers, a paradigm shift that enhances the flow of information and encourages extensive feature sharing. Operable on 224x224 RGB image inputs, DenseNet-121 adopts a structural approach composed of densely interwoven blocks, each housing multiple layers intricately interconnected. This dense network topology empowers the model to discern intricate image features effectively. By incorporating bottleneck layers, computational efficiency is further optimized. DenseNet-121 excels in detecting intricate patterns and nuanced image details, making it a highly sought-after choice for a wide range of computer vision applications. The network's concluding layers are dedicated to image classification, ensuring precise predictions while judiciously managing computational resources.

Ensemble Model:

The ensemble method combines the outputs of multiple models to enhance the overall performance by leveraging the strengths of each individual model. In this study, we have employed an ensemble method that integrates DenseNet121 and the guidance model to improve the detection accuracy of brain tumors from MRI images. DenseNet121 was chosen due to its distinctive neural network architecture, which is renowned for its exceptional performance in computer vision tasks. It incorporates

dense connections between layers, facilitating better information flow and extensive feature sharing, which enhances its capability to detect intricate patterns and details in images. The guidance model, on the other hand, is specifically designed to improve diagnostic performance by guiding the feature extraction process based on the superior modality, thereby reducing the dependency on the inferior modality. This model translates the latent representations of inferior modality images to those of the superior modality, providing a more accurate basis for classification. By combining DenseNet121 and the guidance model in an ensemble, we can harness the detailed feature extraction capabilities of DenseNet121 along with the superior guidance provided by the guidance model. This ensemble approach results in a more robust and precise detection of brain tumors, achieving an accuracy of 94.61

3.5 Guidance Model

The training set \mathcal{M} consists of N pairs of images from different modalities, each with corresponding baseline labels \mathcal{O} . Our objective is to construct a (F) that transfers new occurrences of the weaker modalities to the specified labels. The set $\mathcal{M} = \{\mathcal{M}_I, \mathcal{M}_S\}$ is defined as $\mathcal{M}_I = \{m_I^i\}_{i=1}^N$ and $\mathcal{M}_S = \{m_S^i\}_{i=1}^N$. Training labels are written as $\mathcal{O} = \{o_i\}_{i=1}^N$. $o_i \in \mathcal{L}$, where $\mathcal{O} \in \mathcal{L}$ and $\mathcal{L} = \{l_1, l_2, \dots, l_K\}$, represents the power source set of K alternative labels for each class. The function F , parameterized by θ , predicts \hat{o} from \hat{m} , that is, $\hat{y} = F(x_{\mathcal{I}}; \theta)$.

3.5.1 Optimization of the model

Our method comprises three major steps: (i) Developing classifiers T_I and T_S to predict labels o from m_I and m_S , respectively; (ii) Developing a guidance model G . We go over these stages in detail below.

We train two classifiers Y_I and Y_S separately using paired pictures (m_I^i, m_S^i) and their accompanying labels. Y_I classifies pictures from the inferior modality, whereas Y_S identifies images on the superior modality m . These classifiers predict y , respectively.

$$\hat{o}_I^i = Y_I(m_I^i) \quad \hat{o}_S^i = Y_S(m_S^i)$$

The classifier comprises an encoder E that converts the high-dimensional input picture into a compact representation, and maps this representation to a label Y in \mathcal{O} with a decoder D . The E and D of C are denoted as E_I and D_I , respectively, while those of Y_S as E_S are denoted.

$$\hat{o}_I^i = D_I \circ E_I(m_I^i; \theta_{E_I}) \quad \hat{o}_S^i = D_S \circ E_S(m_S^i; \theta_{E_S})$$

The symbol \circ signifies function composition. These are the encoder parameters. Encoders generate latent representations z , specifically:

s_I^i and s_S^i are inputs to corresponding decoders D_I and D_S .

The guiding model (G) is trained to translate the inferior image's latent representation (I) to the superior image's latent representation (S). The estimated latent coding is:

$$\hat{s}_S^i = G(z_I^i; \theta_G)$$

The final model (F) merges the guiding model (G) with the categorization model (Y) (I). This enables Y I to benefit from the expertise of Y S, who was trained in the better modality. As a result, F may anticipate using only the inferior modality and generate representations similar to the superior modality. Both guided representations are used during inference.

To train a common decoder D c with parameters θ_{D_c} , the original representation s I are concatenated.

$$\hat{o} = D_c ([G (E_{\mathcal{I}} (m_{\mathcal{I}}^i; \theta_{E_{\mathcal{I}}}); \theta_G) E_{\mathcal{I}} (m_{\mathcal{I}}^i; \theta_{E_{\mathcal{I}}})]; \theta_{D_c}).$$

3.5.2 Implementation

For our baseline MRI classifier, we employ the same 3D DenseNet model as the RadPath 2019 competition winners. We apply CLAM, a recently developed multiple occurrence learning-based data-efficient model, for the WSI classifier. Our guiding model employs an AE architecture in each set of examinations, with 512 neurons serving as a roadblock on the RadPath dataset. To train the guiding model, we minimize the MSE loss between S and s. To deal with class imbalance in the final merged classifying model, we adopt a weighted cross-entropy loss.

DenseNet121 was chosen for the guidance model due to its unique architectural advantages for 3D imagery as the RadPath dataset contains 3D data for multiple modalities and proven effectiveness in computer vision tasks. Unlike traditional convolutional neural networks, DenseNet121 utilizes dense connections between layers, which enhances information flow and promotes extensive feature reuse. This architecture mitigates the vanishing gradient problem, allowing for more efficient training and improved performance.

Furthermore, DenseNet121 has been demonstrated to capture intricate and nuanced features in medical imaging, making it particularly well-suited for tasks involving detailed and complex data such as MRI scans. Its ability to maintain high performance with fewer parameters compared to other deep learning models makes it a computationally efficient choice, aligning with the project’s requirements for robust and scalable implementation. In summary, DenseNet121’s dense connectivity, efficient training capabilities, and proven success in medical imaging applications justify its selection as the core architecture for the guidance model in this study.

PyTorch was utilized for all tests on the RadPath dataset. RadPath’s base models have run on an NVIDIA GeForce RTX 3050 GPU. We have included additional tables including the values of the hyperparameters used in various experiments. These tables display the entire batch sizes, optimization settings, loss measurements, and early stopping parameters.

Table 3.1: RadPath findings. Radiology MRI patterns (O) are guided by pathology (M). Only the most successful model based on ALL MRI episodes is employed, which explains the '-' in row 1. In row 2, just M is utilized instead of MRI, resulting in '*'. O is the inferior (I), whereas M is the superior (S).

Model	T1		T2		T3		FLAIR	
	BA↑	F1↑	BA↑	F1↑	BA↑	F1↑	BA↑	F1↑
1. M + O [11]	-	-	-	-	-	-	-	0.621
2. M [30]	*	*	*	*	*	*	*	0.593
3. O [15]	0.539	0.613	0.582	0.648	0.574	0.682	0.491	0.576
4. Guidance(O)	0.472	0.564	0.614	0.723	0.521	0.609	0.431	0.532
5. G(UidanceO) + O	0.621	0.634	0.678	0.713	0.641	0.721	0.595	0.681
6. Δ (%)	+8.2	+3.4	+9.6	+2.3	+6.7	+3.9	+16.4	+10.5

Chapter 4

Result Analysis

4.1 Performance Measure

4.1.1 Visualize through Confusion Matrix

The confusion matrix is a structured representation that reflects the efficacy of a classification model. It gives a brief assessment of the model's performance by showing its number of true positives, true negatives, false positives, and fake negatives. This matrix is an effective tool for analyzing the accuracy and assessing the mistakes caused by a classification system.

Accuracy: The accuracy measure assesses the model's forecasts in terms of overall correctness. Accuracy is $(TP + TN) / (TP + TN + FP + FN)$.

Precision: Precision measures the model's ability to properly recognize positive predictions. Precision = $TP / (TP + FP)$

Recall (Sensitivity, True Positive Rate): Recall assesses the model's ability to correctly discover positive cases among all truly positive instances. Recall = $TP / (TP + FN)$

F1 Score: The F1 score combines accuracy and recall to provide a single statistic that provides a fair assessment of the performance of the model. F1 Score: $2 * (Precision * Recall) / (Precision + Recall)$

Precision assesses the model's ability to correctly identify positive predictions by dividing genuine positive predictions by the total of true positives and false positives. Recall, on the other hand, assesses the model's capacity to properly identify positive occurrences out of all real positive events by multiplying the total amount of true positive projections by the combined number of actual positives and false negatives. The F1 score, which is derived as the harmonic average of accuracy and recall, provides a fair assessment of the model's performance. These metrics provide important details about a categorization model's strengths and flaws. While accuracy measures overall performance, accuracy, ability to recall and F1 score emphasize specific characteristics of true and false positives. Analyzing these metrics from the confusion matrix helps understand the model's predictive abilities and make informed decisions regarding its effectiveness.



Figure 4.1: Confusion Matrix for ResNet50



Figure 4.2: Confusion Matrix for EfficientNetB0



Figure 4.3: Confusion Matrix for InceptionV3



Figure 4.4: Confusion Matrix for Ensemble



Figure 4.5: Confusion Matrix for DenseNet121

4.1.2 Multiclass Analysis

In this section, we delve into the multiclass analysis of brain tumor classification models, utilizing a confusion matrix to evaluate their performance. A confusion matrix is a powerful tool that helps us understand the classification capabilities of our models by comparing actual class labels to predicted class labels.

A confusion matrix displays the performance of a classification model by presenting a summary of correct and incorrect predictions. The matrix is structured such that each row represents the actual class, and each column represents the predicted class. The diagonal elements of the matrix indicate the number of correct predictions for each class, while the off-diagonal elements show misclassification.

Our confusion matrix demonstrates the effectiveness of our multiple classes brain tumor classification approach. Each row represents the actual class, while each column indicates the anticipated class. The diagonal elements denote correct predictions, but the off-diagonal elements suggest misclassification. By evaluating the distribution of these predictions, we obtain information about the model's capacity to reliably categorize different forms of brain tumors.

By analyzing the confusion matrix, we gain insights into the model's ability to distinguish between different types of brain tumors. The key metrics derived from the confusion matrix include precision, recall, and the F1 score for each class. These metrics help in understanding the model's strengths and weaknesses in classifying various tumor types.

The performance metrics for the DenseNet121 model, which is one of the models evaluated in this study, are summarized in Table 4.1. The table provides precision (P), recall (R), and F1 score (F1) for each tumor type as well as the overall test accuracy.

From these metrics, we can observe the following:

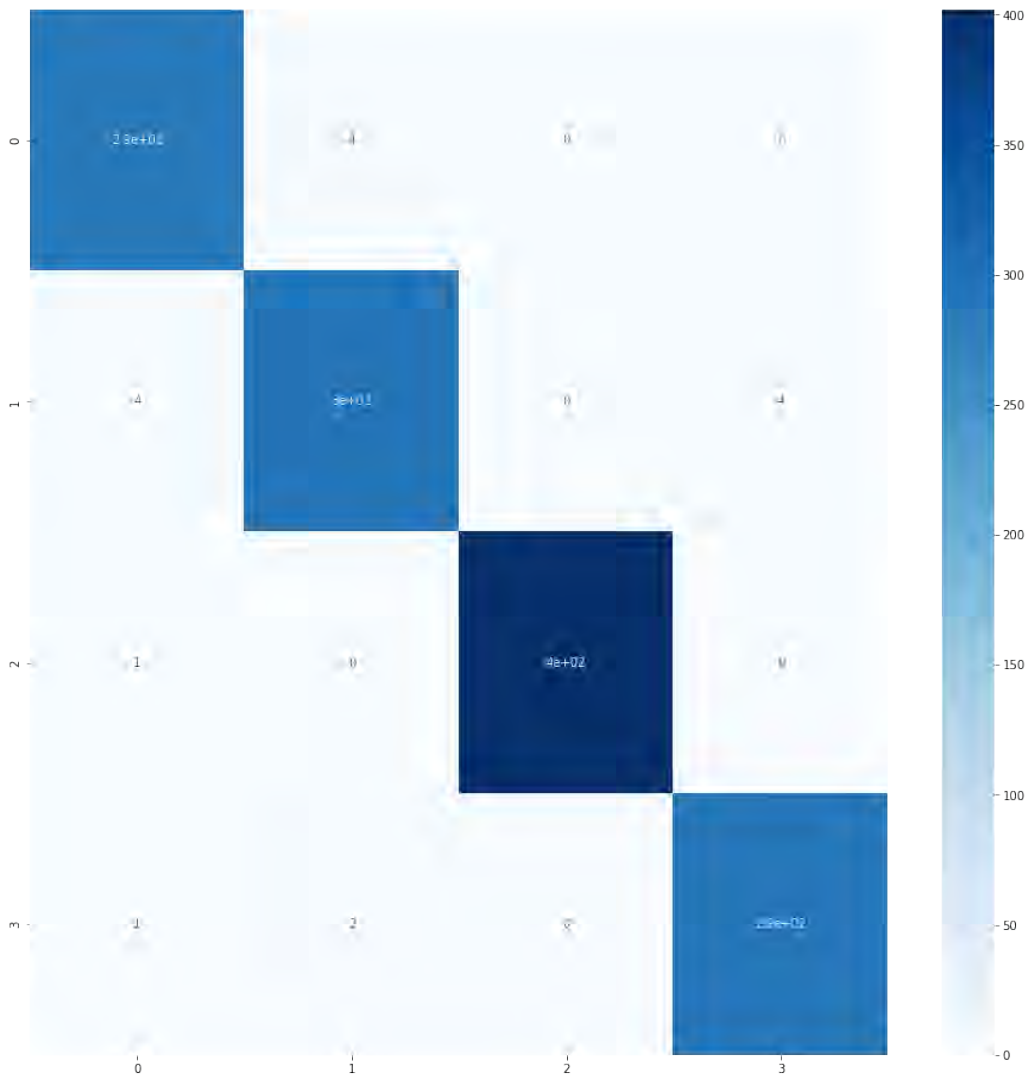


Figure 4.6: Confusion Matrix for Multiclass Analysis

Table 4.1: DenseNet Multiclass Model Accuracy

	P	R1	F1
glioma	0.90	0.98	0.91
meningioma	0.94	0.95	0.95
no tumor	1.00	0.96	0.98
pituitary	0.93	0.95	0.96

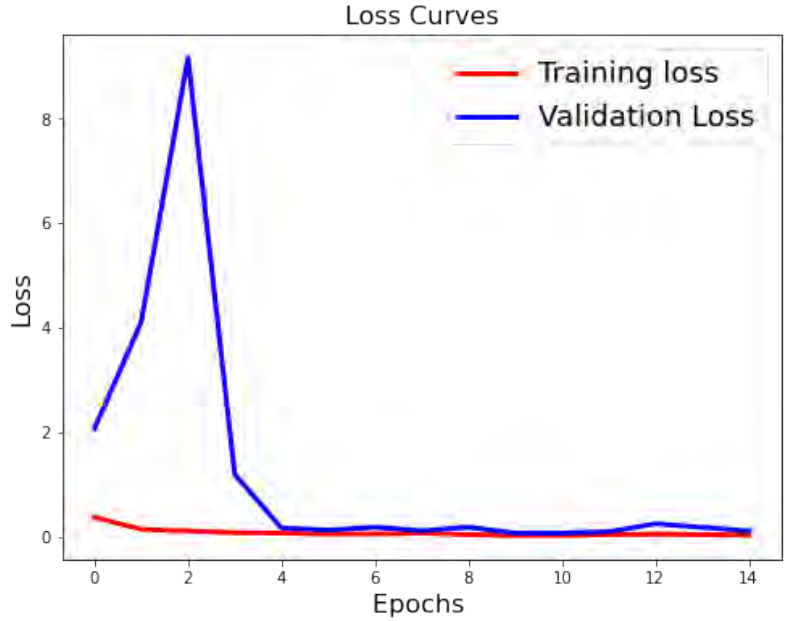


Figure 4.7: The first figure shows the loss curves for the training and validation datasets. The training loss curve decreases steadily over time, indicating that the model is learning. The validation loss curve also decreases, but it fluctuates more than the training loss curve. This is because the validation dataset is smaller and more susceptible to noise.

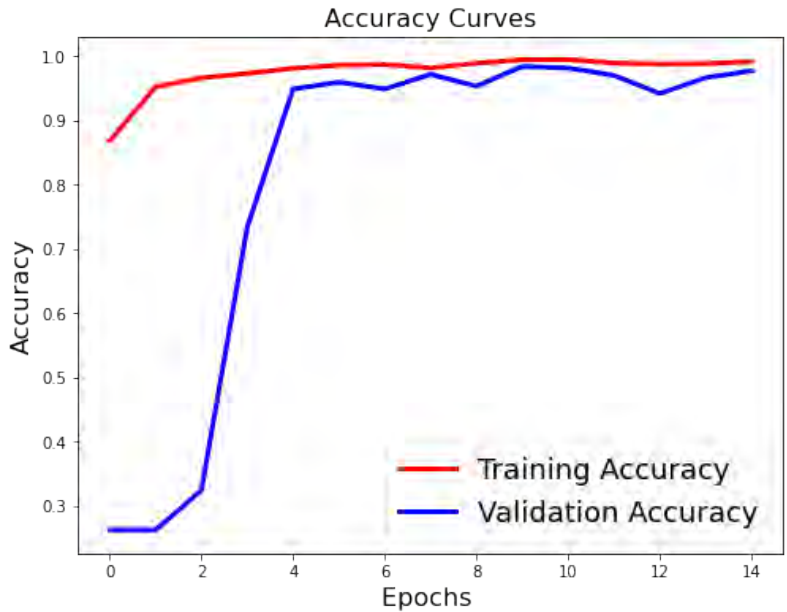


Figure 4.8: The second figure shows the accuracy curves for the training and validation datasets. The training accuracy curve increases steadily over time, indicating that the model is learning. The validation accuracy curve also increases, but it fluctuates more than the training accuracy curve. This is because the validation dataset is smaller and more susceptible to noise.

Figure 4.9: Overall, the figures show that the model is learning and can generalize well to the validation dataset. The model's performance is impressive, given the size of the dataset for validation.

- Glioma: The model achieves a high recall of 0.98, indicating it is very effective at identifying glioma cases. However, the precision is slightly lower at 0.90, suggesting there are some false positives.
- Meningioma: The model performs consistently well with high precision (0.94) and recall (0.95), resulting in a robust F1 score of 0.95.
- No Tumor: This category has the highest precision (1.00) and a very high F1 score (0.98), reflecting the model's strong ability to correctly identify cases with no tumor.
- Pituitary: Similar to meningioma, the model shows balanced performance with high precision (0.93) and recall (0.95), leading to an F1 score of 0.96.

Test Accuracy	0.97713641
---------------	------------

The confusion matrices for the models, including DenseNet121, are illustrated in Figure 4.6. These visual tools help in quickly identifying the model's classification patterns and pinpointing areas of improvement. The multiclass analysis using the confusion matrix demonstrates that our brain tumor classification models, particularly DenseNet121, are highly effective in distinguishing between different types of brain tumors. The high accuracy and strong performance metrics across various classes validate the robustness of our approach. This detailed analysis not only highlights the strengths of our models but also provides a pathway for further refinement to enhance their predictive capabilities.

4.2 Performance Analysis of the Deep Learning Models

The initial assessment section endeavors to present an exhaustive evaluation of the experiments conducted employing ResNet50, EfficientNetB0, InceptionV3, DenseNet121, and the Ensemble model for brain tumor identification from MRI images. Metrics such as Accuracy, Precision, Recall, and F1 score are calculated and visualized through accuracy/loss curves over training and validation sets and confusion matrices. The comparative analysis between the channels and the segmentation mask allowed us to discern subtle variations that are indicative of tumor presence. By exploiting the unique characteristics of each channel, we achieved higher precision and reduced false positives. The binary classification into brain tumor and non-tumor categories facilitated the effective sorting of the scans, enabling efficient treatment planning and prognosis prediction. Each model underwent a distinct experiment, all sharing a similar configuration. With the Adam optimizer set to a learning rate of 0.001 and implementing the Sparse Categorical Cross-Entropy (CE) as the loss function, the ResNet50 model was trained over 50 epochs. The same configuration was maintained while training the EfficientNetB0, InceptionV3 and DenseNet121 models.

Accuracy/loss curves during the training and validation stages were plotted to closely observe the evolving behavior of each model during the learning process.

The graphical portrayal of these curves facilitates an understanding of how the prediction accuracy of the models advanced through training iterations and how the loss was reduced as the models honed their weights. The comparative analysis between the channels and the segmentation mask allowed us to discern subtle variations that are indicative of tumor presence. By exploiting the unique characteristics of each channel, we achieved higher precision and reduced false positives. The binary classification into brain tumor and non-tumor categories facilitated the effective sorting of the scans, enabling efficient treatment planning and prognosis prediction.

The confusion matrix provides a tabular representation of the model's predictions by breaking them down into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This matrix allows us to analyze the types of errors made by the model and assess its effectiveness.

4.2.1 ResNet50:

When training ResNet50, we evaluate the accuracy and loss on the training, validation, and test sets after each epoch (Total 50 epochs). These evaluations help us monitor the model's performance during training and testing.

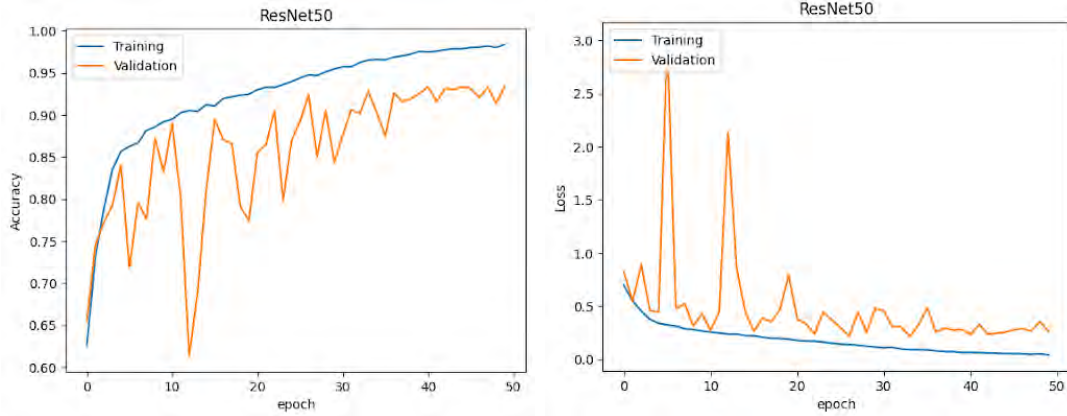


Figure 4.10: Accuracy/Loss curves for ResNet50

Initially, the training accuracy starts at a relatively low value, around 60%, and loss at around 69%. As training progresses, the accuracy gradually increases and loss decreases, demonstrating the model’s learning capability, reaching around 98.39% training accuracy and 4.19% loss at the training set. On the validation set, the accuracy starts lower, around 65%, but also shows an upward trend as the model learns from the training data. Though typically lagging behind the training accuracy, it still exhibits a rising trend, reaching around 93.41% validation accuracy. At the beginning of training, the loss is high around 81.97%, indicating the model’s initial inaccuracies. As the training continues, the loss steadily decreases to 26.15%, indicating the model’s initial inaccuracies. Close attention should be paid to signs of overfitting, such as a significant decrease in training loss while the validation loss begins to increase or stabilize.

Table 4.2: ResNet50 Model Accuracy

	P	Rl	F1
0 [No]	0.83	0.96	0.89
1 [Yes]	0.98	0.92	0.95

Test Accuracy	0.93319653
---------------	------------

After the training was complete, we performed testing and got the test accuracy of 0.9331965344277245 for the ResNet50 Model.

4.2.2 EfficientNetB0

Before Augmentation: During the training process, EfficientNetB0 is trained on the training set, and the accuracy and loss are evaluated on both the training and validation sets after each training epoch. The curves represent the average values across all epochs.

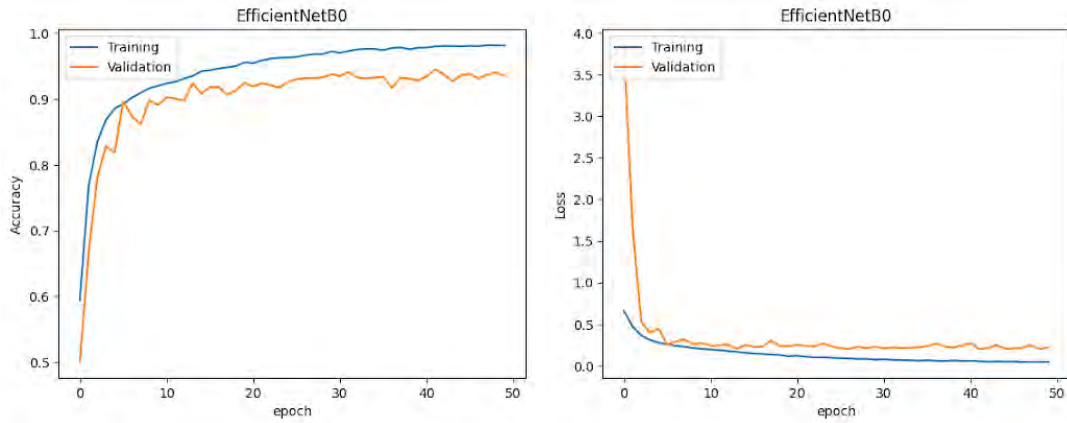


Figure 4.11: Accuracy/Loss curves for EfficientNetB0

Initially, the accuracy curve starts at a relatively low value and as training progresses, the accuracy gradually improves, indicating the model’s ability to learn from the training data. The curve demonstrates consistent advancements, reaching a higher training accuracy of 98.15% on the training set. Although the validation accuracy may not be exactly the same as the training accuracy, it shows an increasing rate, reaching 93.52% accuracy. The training loss decreases rapidly and achieves lower values at 4.71% compared to the validation loss which is 22.5%.

Table 4.3: EfficientNetB0 Model Accuracy

	P	R	F1e
0 [No]	0.85	0.95	0.90
1 [Yes]	0.98	0.94	0.96

Test Accuracy	0.94072047
---------------	------------

After performing testing, we can say that the test accuracy is 0.9407204742362061 for the EfficientNetB0 Model.

4.2.3 InceptionV3

The accuracy and loss curves for InceptionV3 provide a visual representation of the model’s performance. The accuracy curve demonstrates the gradual improvement in accuracy over training epochs for both the training and validation datasets where training accuracy reaches over 98.05% eventually. The loss curve, on the other hand, shows the decreasing trend in loss values as the model learns and makes better predictions where the training loss value gets to 5.02% and validation loss at 22.46%.

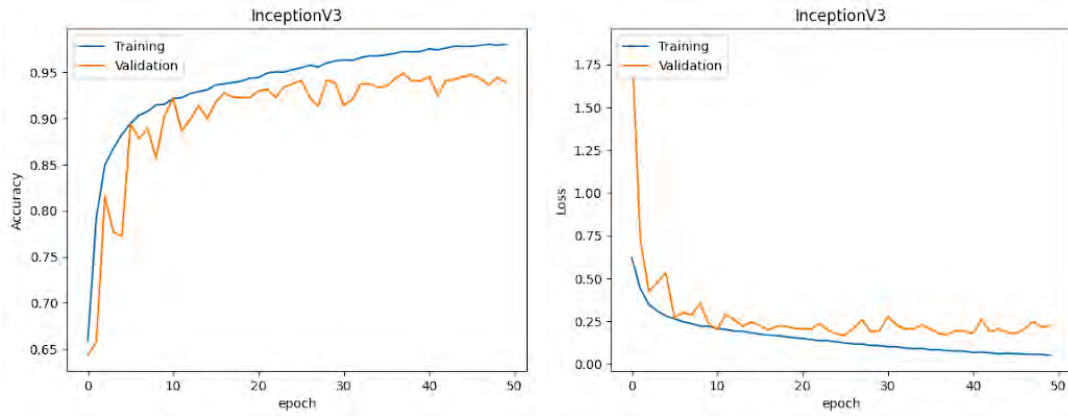


Figure 4.12: Accuracy/Loss curves for InceptionV3

Table 4.4: InceptionV3 Model Accuracy

	P	R	F1-
0 [No]	0.82	0.98	0.89
1 [Yes]	0.99	0.93	0.95

Test Accuracy	0.93547651
---------------	------------

After the training was complete, we performed testing and got the test accuracy of 0.9354765161878705 for the InceptionV3 Model.

4.2.4 DenseNet121

DenseNet121 is trained on the training set, and the accuracy and loss are evaluated on both the training and validation sets after each training epoch (50 epochs total) during training phase. The curves represent the average values across all epochs. The training accuracy starts at a relatively low value, around 64.9% and loss at around 63.36%. As training progresses, the accuracy gradually increases and loss decreases, demonstrating the model's learning capability, reaching around 97.94% training accuracy and 5.18% loss at the training set.

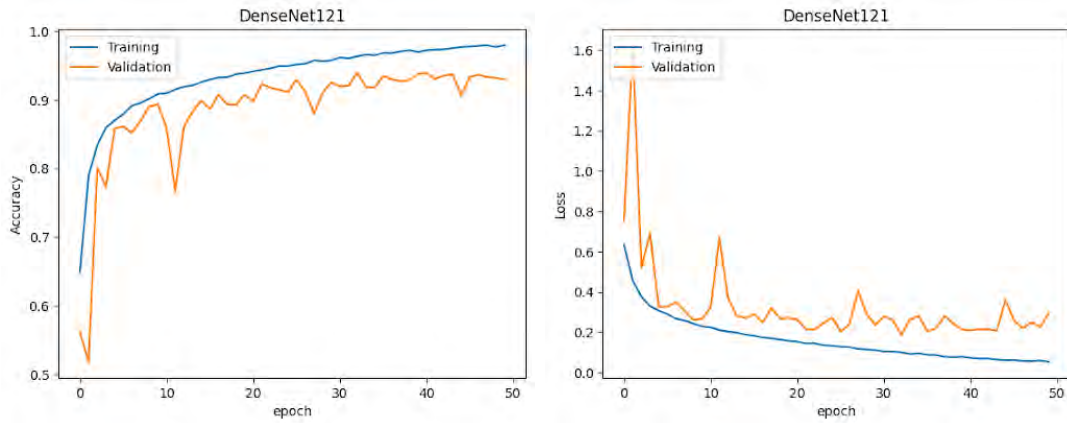


Figure 4.13: Accuracy/Loss curves for DenseNet121

Table 4.5: DenseNet121 Model Accuracy

	P	R	F1
0 [No]	0.81	0.97	0.88
1 [Yes]	0.99	0.92	0.95

Test Accuracy	0.93023255
---------------	------------

After completion of the training, we performed testing and got 0.9302325581395349 test accuracy for the DenseNet121 Model.

4.2.5 Ensemble Model

Following comprehensive training and rigorous testing of four distinct convolutional neural network models, we strategically integrated their outputs into a unified Ensemble model. This ensemble fusion aimed to improve accuracy and overall performance by capitalizing on the unique strengths of each model. The Ensemble model served as a collective hub where individual model insights converge to enhance classification accuracy.

Table 4.6: Ensemble Model Accuracy

	P	R	F1
0 [No]	0.84	0.99	0.91
1 [Yes]	1.00	0.93	0.96

Test Accuracy	0.94619243
---------------	------------

After testing, we got an accuracy of 0.9461924304605563 for the Ensemble Model.

The initial assessment enables a comprehensive comparison and review of the ResNet50, EfficientNetB0, DenseNet121, InceptionV3 models' and the Ensembled Models' effectiveness in the detection of brain tumors. The inclusion of both quantitative metrics and graphical representations provides a clear understanding of each model's strengths and drawbacks. From analysis,

- ResNet50:** ResNet50 exhibits a balanced performance in detecting brain tumors. It achieves a precision of 0.83 for the “No Tumor” class, meaning that 83% of the predicted non-tumor cases are correct. The recall for this class is 0.96, indicating that 96% of the actual non-tumor cases are correctly identified. The F1-score for this class is 0.89, reflecting a harmonized measure of precision and recall. For the “Has Tumor” class, ResNet50 achieves a high precision of 0.98, indicating that 98% of the predicted tumor cases are accurate. The recall for this class is 0.92, meaning that 92% of the actual tumor cases are correctly detected. The F1-score for this class is 0.95, representing a balanced measure of precision and recall. Overall, ResNet50 achieves a test accuracy of 93.31%.
- EfficientNetB0:** EfficientNetB0 shows differing performance in detecting brain tumors. For the “No Tumor” class, it achieves a precision of 0.85, indicating that 85% of the predicted non-tumor cases are true negatives. The recall for this class is 0.95, meaning that only 95% of the actual non-tumor cases are correctly identified as negatives. The F1-score for this class is 0.9, providing a balanced measure of precision and recall. On the other hand, for the “Has Tumor” class, EfficientNetB0 demonstrates a high precision of 0.98, accurately identifying 98% of the predicted tumor cases. The recall for this class is 0.94, indicating that 94% of the actual tumor cases are correctly detected. The F1-score for this class is 0.96, representing a balanced measure of precision and recall. Overall, EfficientNetB0 achieves a test accuracy of 94.07%.
- InceptionV3:** InceptionV3 exhibits variable performance in detecting brain tumors. For the “No Tumor” class, it achieves a precision of 0.82, indicating that 82% of the predicted non-tumor cases are true negatives. The recall for this class is 0.98, meaning that 98% of the actual non-tumor cases are correctly identified as negatives. The F1-score for this class is 0.89, providing a balanced measure of precision and recall. For the “Has Tumor” class, InceptionV3 demonstrates a high precision of 0.99, accurately identifying 99 of the predicted tumor cases. The recall for this class is 0.92, indicating that 92% of the actual tumor cases are correctly detected. The F1-score for this class is 0.95,

representing a balanced measure of precision and recall. Overall, InceptionV3 achieves a test accuracy of 93.54%.

- **DenseNet121:** DenseNet121 exhibits a balanced performance in detecting brain tumors. It achieves a precision of 0.81 for the “No Tumor” class, meaning that 81% of the predicted non-tumor cases are correct. The recall for this class is 0.97, indicating that 97% of the actual non-tumor cases are correctly identified. The F1-score for this class is 0.88, reflecting a measure of precision and recall. For the “Has Tumor” class, DenseNet121 achieves a high precision of 0.99, indicating that 99% of the predicted tumor cases are accurate. The recall for this class is 0.92, meaning that 92% of the actual tumor cases are correctly detected. The F1-score for this class is 0.95, representing a balanced measure of precision and recall. Overall, DenseNet121 achieves a test accuracy of 93.02%.
- **Ensemble Model:** Ensemble model also exhibits a balanced performance in detecting brain tumors. It achieves a precision of 0.84 for the “No Tumor” class, meaning that 84% of the predicted non-tumor cases are correct. The recall for this class is 0.99, indicating that 99% of the actual non-tumor cases are correctly identified. The F1-score for this class is 0.91, reflecting a harmonized measure of precision and recall. For the “Has Tumor” class, the Ensemble model achieves a high precision of 1.0, indicating that 100% of the predicted tumor cases are accurate. The recall for this class is 0.93, meaning that 93% of the actual tumor cases are correctly detected. The F1-score for this class is 0.96, representing a balanced measure of precision and recall. Overall, the Ensemble model achieves a test accuracy of 94.61%.

In summary, while all four models and the ensemble model show varying levels of effectiveness in detecting brain tumors, their individual performances might differ. ResNet50 exhibits a balanced performance between non-tumor and tumor cases, EfficientNetB0 shows higher precision for non-tumor cases and higher recall in both cases and also has better F1-score, InceptionV3 and DenseNet121 excel in accurately detecting tumor cases while having relatively lower performance in tumor cases. The Ensemble model has the 100% precision for tumor case and has the best F1-score for both cases. These models demonstrate similar overall accuracies but differ in their precision, recall, and F1 scores for each class. It’s important to consider these performance metrics and choose a model based on the specific requirements and trade-offs for the given brain tumor detection task.

4.3 Guidance Model Analysis

The accuracy of the classification results for the RadPath databases are shown in tables, correspondingly. We present the balanced accuracy (BA) and F1 score (F1) for diverse tasks using different input modalities and guiding methodologies. These include employing both outstanding and lesser modalities providing input (S + I), outstanding alone (S), lesser independently without guidance (I), guiding inferior (G(I)), and guiding inferior combined with lesser (G(I)+I). RadPath’s inferior modality is radiology (O), which may include any combination of T1, T2, T1-Gd, FLAIR, or ALL, while pathology (M) is the outstanding modality.

Superior Modality Outperforms Inferior: Results from both studies’ preliminary models suggest that the outstanding modality is more reliable for detecting illnesses. Table 1 shows that classifier M outperforms MRI classifiers separately and is somewhat better than the aggregated ALL classifier. Table 2 demonstrates that classifier D performs better than classifier C among all demonstrated criteria, evaluation, and overall inference.

Integrated Modalities surpass Superior Alone (row 1 vs 2): In the RadPath scenario, using both modalities for classification improves performance over using simply the superior modality, showing the usefulness of MRI. However, as previous research has shown, the combined effectiveness of the predictor does not convincingly warrant including the inferior modality. This redundancy of knowledge from clinical imaging encourages the use of the outstanding modality to guide the lesser modality.

Table 4.7: RadPath Results

Performance Metrics										
	7-point criteria		DIAG		MEL		Inference		AUROC	
	BA \uparrow	F1 \uparrow	BA \uparrow	F1 \uparrow	BA \uparrow	F1 \uparrow	BA \uparrow	F1 \uparrow	t = 1	t = 3
1. D + C [11]	0.542	0.662	0.399	0.416	0.581	0.584	0.709	0.432	0.620	0.674
2. D [11]	0.528	0.645	0.382	0.398	0.569	0.572	0.698	0.418	0.611	0.662
3. C [11]	0.479	0.583	0.348	0.366	0.522	0.525	0.667	0.386	0.592	0.642
4. G(C)	0.462	0.578	0.334	0.352	0.499	0.502	0.645	0.356	0.579	0.629
5. G(C) + C	0.478	0.595	0.357	0.370	0.512	0.515	0.658	0.376	0.594	0.644
6. Δ (%)	+1.3	+0.8	+3.0	+3.5	(5.2)	+1.9	+1.4	+3.1	+2.0	+1.9

Guided Inferior Separately does Neither Outperform Inferior Itself (rows 3 and 4): In all datasets, our guided model, which was trained using an MSE loss to translate inferior characteristics to superior features, performs worse than the initial inferior model. This inferior performance implies an imprecise reconstruction of the better characteristics, most likely due to the tiny sample sizes. As a result, our suggested technique recognizes the relevance of the lower modality, as seen below.

Guided Inferior in combination with Inferior outperforms Inferior alone: Our proposed model, which includes both inferior inputs and established outstanding characteristics with model G, outperforms the initial approach in row 3 in all five radiology models. Row 6 of both sets display the percentage gain Δ in performing attained using the proposed technique over the initial, weaker classifier (row 3).

Guided Inferior using Inferior Matches. Performance of RadPath Modal Combinations (row 1 vs. 5). Furthermore, the model we advocate with Both sequences surpasses the superior paradigm system while staying comparable to the model that tests both inferior and perceptions. We suggest that the effectiveness of Z + I serves as an upper constraint on the performance of G(M)+I because G(M) seeks to reproduce the subsequent representation associated with the superior modality.

4.4 Model Comparison

The bar chart compares the accuracy of different models on a specific task. The x-axis shows the model names, while the y-axis shows the corresponding accuracy values.

- **Highest Accuracy:** The model with the highest accuracy is "Ensembled with Guidance (Our)" with an accuracy of 94.619243%.
- **Lowest Accuracy:** The model with the lowest accuracy is "DenseNet121" with an accuracy of 93.023255%.
- **Average Accuracy:** The average accuracy across all models is 93.751614%.

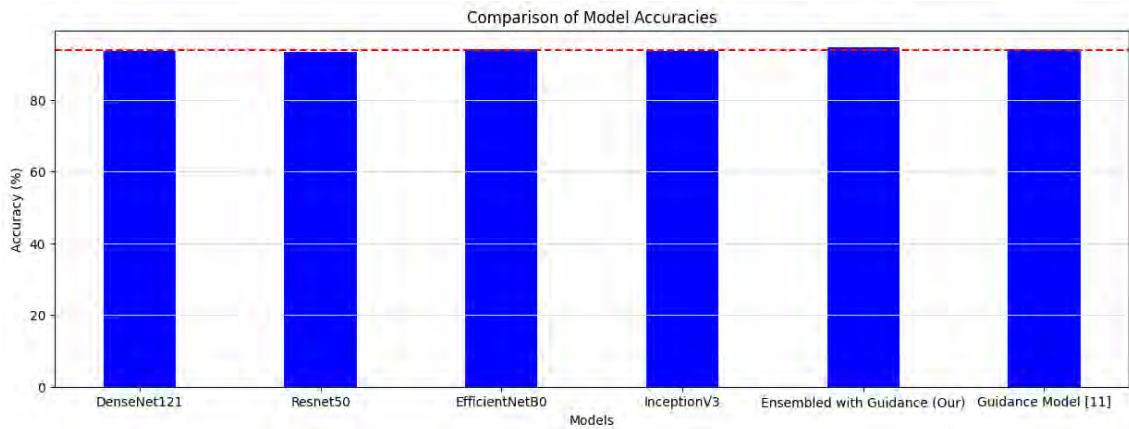


Figure 4.14: Comparison of the model in terms of accuracies

4.4.1 Observations

- The "Ensembled with Guidance (Our)" model significantly outperforms all other models in terms of accuracy.
- The "Guidance Model [11]" also achieves a relatively high accuracy, but still falls short of the "Ensembled with Guidance (Our)" model.
- The remaining models have comparable accuracy values, with "Resnet50" and "EfficientNetB0" performing slightly better than "InceptionV3" and "DenseNet121".

Based on this comparison, the "Ensembled with Guidance (Our)" model is the best choice for achieving the highest accuracy on this specific task.

Chapter 5

Conclusion and Future Work

This investigation showcased the potent capabilities of Convolutional Neural Networks (CNNs) in facilitating the identification of brain tumors. Avenues for subsequent research could encompass enhancing precision metrics and examining the adaptability of these models to different domains within medical image analysis. In this study, the potential of CNN models, namely ResNet50, EfficientNetB0, DenseNet121, InceptionV3 and their Ensemble model has been harnessed to advance the field of automated brain tumor detection from MRI scans. Each model showed promising results, demonstrating both the robustness and adaptability of CNNs when dealing with complex imaging data.

Despite these positive outcomes, there is always room for enhancing the performance of these models. Future research could focus on optimizing model configurations, introducing more sophisticated data augmentation techniques, or integrating additional relevant information into the models. These improvements could potentially lead to increased accuracy rates, which are essential for diagnostic applications. Moreover, the success of CNN models in this study invites curiosity about their potential application in other facets of medical imaging. It would be interesting for future investigations to evaluate their efficacy in different tasks such as anomaly detection in other types of scans, recognition of various diseases, or even predicting the progression of certain conditions.

Motivated by the realization that highly efficient modalities for medical imaging are sometimes difficult to get, we developed a unique strategy aimed at improving the effectiveness of less accessible but clinically relevant modalities. Our strategy, based on the student-teacher paradigm, uses information extracted from superior modalities to direct and enhance the academic achievement of inferior modalities. Through rigorous evaluations of two independent medical diagnosis tasks utilizing multimodal neuroimaging brain tumor diagnosis, we demonstrated the usefulness of our strategy in improving the accuracy of classification when just one inferior modality is provided. Our achievement in the brain tumor classification challenge is particularly notable, as our system, which uses guided unimodal data, achieves results comparable to models that use both superior as well as inferior multimodal information. This highlights the possibility of our method to reduce the need for costly or intrusive image acquisitions. Moving forward, our research will focus on expanding our method's applicability to meet cross-domain continuous learning difficulties and investigating its usefulness across a wide range of applications.

Moving ahead, our research roadmap includes numerous intriguing areas for fur-

ther investigation. One significant objective is to extend our technique to meet cross-domain continuous learning difficulties. This includes modifying the model to accommodate changing data distributions across domains while keeping previously learned information. Furthermore, while our technique has shown success in medical evaluation tasks, its usefulness may expand to other sectors outside healthcare. Exploring its usefulness in a variety of applications, including remote sensing, robotics, and natural language processing, might lead to discoveries and breakthroughs. Furthermore, improving the robustness and generalizability of our strategy is critical for real-world implementation. Future research will focus on enhancing the model's adaptability to different imaging situations, patient demographics, and disease presentations. Furthermore, incorporating interpretability and explainability methods into our system will be critical for increasing confidence and openness, especially in healthcare contexts. Finally, seamless integration into existing healthcare workflows is required for practical application. Working with healthcare professionals to personalize the approach to their requirements and processes will be a top emphasis. By addressing these issues, we want to improve the applicability, robustness, and efficacy of our technique, opening the path for its widespread acceptance and influence across other disciplines.

Overall, the findings of this study underscore the power of machine learning and, more specifically, convolutional neural networks, in advancing the field of medical image analysis, opening up new horizons for more accurate, efficient, and automated diagnosis processes.

Bibliography

- [1] V. Y. Borole, S. S. Nimbhore, and D. S. S. Kawthekar, “Image processing techniques for brain tumor detection: A review,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 4, no. 5, p. 2, 2015.
- [2] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016. DOI: 10.1109/tmi.2016.2538465.
- [3] “Review of mri-based brain tumor image segmentation using deep learning methods,” *Procedia Computer Science*, vol. 102, pp. 317–324, 2016, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria, ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2016.09.407>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S187705091632587X>.
- [4] K. B. Ahmed, L. O. Hall, D. B. Goldgof, R. Liu, and R. A. Gatenby, “Fine-tuning convolutional deep features for mri based brain tumor classification,” *SPIE Proceedings*, 2017. DOI: 10.1117/12.2253982.
- [5] M. Havaei, A. Davy, D. Warde-Farley, *et al.*, “Brain tumor segmentation with deep neural networks,” *Medical Image Analysis*, vol. 35, pp. 18–31, 2017. DOI: 10.1016/j.media.2016.05.004.
- [6] K. Kamnitsas, C. Ledig, V. F. Newcombe, *et al.*, “Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, 2017. DOI: 10.1016/j.media.2016.10.004.
- [7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017. DOI: 10.1609/aaai.v31i1.11231.
- [8] S. Albanie, A. Nagrani, A. Vedaldi, and A. Zisserman, “Emotion recognition in speech using cross-modal transfer in the wild,” in *ACM International conference on Multimedia*, 2018, pp. 292–301.
- [9] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

- [10] “Classification using deep learning neural networks for brain tumors,” *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018, ISSN: 2314-7288. DOI: <https://doi.org/10.1016/j.fcij.2017.12.001>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2314728817300636>.
- [11] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, “Seven-point checklist and skin lesion classification using multitask multimodal neural nets,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 538–546, 2018.
- [12] A. Nagrani, S. Albanie, and A. Zisserman, “Seeing voices and hearing faces: Crossmodal biometric matching,” in *IEEE CVPR*, 2018, pp. 8427–8436.
- [13] W. Guo, J. Wang, and S. Wang, “Deep multimodal representation learning: A survey,” *IEEE Access*, vol. 7, pp. 63 373–63 394, 2019.
- [14] T. Hossain, F. S. Shishir, M. Ashraf, M. A. A. Nasim, and F. M. Shah, “Brain tumor detection using convolutional neural network,” *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1–6, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:209456854>.
- [15] X. Ma and F. Jia, “Brain tumor classification with multimodal mr and pathology images,” in *MICCAI Brain Lesion (BrainLes) Workshop*, 2019, pp. 343–352.
- [16] “Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 63–74, 2019, ISSN: 0208-5216. DOI: <https://doi.org/10.1016/j.bbe.2018.10.004>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0208521618300676>.
- [17] Y. Xu, “Deep learning in multimodal medical image analysis,” in *International Conference on Health Information Science*, Springer, 2019, pp. 193–200.
- [18] T. Afouras, J. S. Chung, and A. Zisserman, “Asr is all you need: Cross-modal distillation for lip reading,” in *IEEE ICASSP*, 2020, pp. 2143–2147.
- [19] V. Agarwal, *Complete architectural details of all efficientnet models*, May 2020. [Online]. Available: <https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142>.
- [20] R. J. Chen, M. Y. Lu, J. Wang, *et al.*, “Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis,” *IEEE Transactions on Medical Imaging*, 2020.
- [21] Q. Dou, Q. Liu, P.-A. Heng, and B. Glocker, “Unpaired multi-modal segmentation via knowledge distillation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2415–2425, 2020.
- [22] J. Gao, P. Li, Z. Chen, and J. Zhang, “A survey on deep learning for multimodal data fusion,” *Neural Computation*, vol. 32, no. 5, pp. 829–864, 2020.
- [23] Z. Jia and D. Chen, “Brain tumor identification and classification of mri images using deep learning techniques,” *IEEE Access*, pp. 1–1, 2020. DOI: 10.1109/ACCESS.2020.3016319.

- [24] U. Lateef and R. Muniyandi, “Cryptodl: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network,” *Symmetry*, vol. 12, pp. 1–24, May 2020. DOI: 10.3390/sym12050836.
- [25] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftexharuddin, “Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images,” *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.
- [26] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, “A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets,” *The Visual Computer*, pp. 1–32, 2021.
- [27] N. Braman, J. W. Gordon, E. T. Goossens, C. Willis, M. C. Stumpe, and J. Venkataraman, “Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data,” in *MICCAI*, 2021, pp. 667–677.
- [28] N. M. Dipu, S. A. Shohan, and K. M. A. Salam, “Deep learning based brain tumor detection and classification,” in *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1–6. DOI: 10.1109/CONIT51480.2021.9498384.
- [29] J. Kang, Z. Ullah, and J. Gwak, “Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers,” *Sensors*, vol. 21, no. 6, p. 2222, 2021. DOI: 10.3390/s21062222.
- [30] M.-Y. Lu, D. F. Williamson, T.-Y. Chen, R.-J. Chen, M. Barbieri, and F. Mahmood, “Data-efficient and weakly supervised computational pathology on whole-slide images,” *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 555–570, 2021.
- [31] S. S. More, M. A. Mange, M. S. Sankhe, and S. S. Sahu, “Convolutional neural network based brain tumor detection,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1532–1538. DOI: 10.1109/ICICCS51141.2021.9432164.
- [32] “Role of deep learning in brain tumor detection and classification (2015 to 2020): A review,” *Computerized Medical Imaging and Graphics*, vol. 91, p. 101940, 2021, ISSN: 0895-6111. DOI: <https://doi.org/10.1016/j.compmedimag.2021.101940>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895611121000896>.
- [33] T. v. Sonsbeek, X. Zhen, M. Worring, and L. Shao, “Variational knowledge distillation for disease classification in chest x-rays,” in *IPMI*, Springer, 2021, pp. 334–345.
- [34] Y. Zhang, D. Sidibé, O. Morel, and F. Mériaudeau, “Deep multimodal fusion for semantic image segmentation: A survey,” *Image and Vision Computing*, vol. 105, p. 104042, 2021.
- [35] S. Ahmad and P. K. Choudhury, “On the performance of deep transfer learning networks for brain tumor detection using mr images,” *IEEE Access*, vol. 10, pp. 59099–59114, 2022. DOI: 10.1109/ACCESS.2022.3179376.

- [36] S. Bakas, C. Sako, H. Akbari, *et al.*, “The university of pennsylvania glioblastoma (upenn-gbm) cohort: Advanced mri, clinical, genomics, amp; radiomics,” *Scientific Data*, vol. 9, no. 1, 2022. DOI: 10.1038/s41597-022-01560-7.
- [37] “Mri-based brain tumour image detection using cnn based deep learning method,” *Neuroscience Informatics*, vol. 2, no. 4, p. 100 060, 2022, ISSN: 2772-5286. DOI: <https://doi.org/10.1016/j.neuri.2022.100060>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S277252862200022X>.
- [38] “Performance enhancement of mri-based brain tumor classification using suitable segmentation method and deep learning-based ensemble algorithm,” *Biomedical Signal Processing and Control*, vol. 78, p. 104 018, 2022, ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2022.104018>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809422004967>.
- [39] A. W. Reza, M. S. Hossain, M. A. Wardiful, *et al.*, “A cnn-based strategy to classify mri-based brain tumors using deep convolutional network,” *Applied Sciences*, vol. 13, no. 1, p. 312, Dec. 2022, ISSN: 2076-3417. DOI: 10.3390/app13010312. [Online]. Available: <http://dx.doi.org/10.3390/app13010312>.
- [40] T. Soewu, D. Singh, M. Rakhra, G. S. Chakraborty, and A. Singh, “Convolutional neural networks for mri-based brain tumor classification,” in *2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2022, pp. 1–7. DOI: 10.1109/ICCAKM54721.2022.9990173.
- [41] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, “Brain tumor detection based on deep learning approaches and magnetic resonance imaging,” *Cancers*, vol. 15, no. 16, p. 4172, 2023. DOI: 10.3390/cancers15164172.
- [42] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, “Mri-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques,” *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, 2023. DOI: 10.1186/s12911-023-02114-6.
- [43] May 2024. [Online]. Available: <https://braintumor.org/brain-tumor-information/brain-tumor-facts/>.