Brain Tumor Classification Based on Deep Convolutional Neural Network using MRI images

Momen Albawarid and Alaa Alhabashna

Al Hussein Technical University, Amman 11831, Jordan
21110399@htu.edu.jo, Alla.alhabaashna@htu.edu.jo

Abstract

Brain tumor classification using MRI images is a vital step in improving early detection and diagnosis. This study explored how well deep learning models can tackle this task, focusing on how data preprocessing, such as image enhancement and augmentation, influences model performance. The big question we wanted to answer: Can these models work accurately enough to be trusted in real-world medical settings?

To find out, the methodology involved training and testing five pre-trained architectures (Xception, ResNet50, InceptionV3, VGG16, and MobileNet) on an extensive MRI dataset. Initially, the models were trained on the original dataset without preprocessing, followed by experiments involving image enhancement and augmentation. The best-performing models were then further tested using a separate validation dataset to assess their robustness. Interviews with experts in AI and bioinformatics were also conducted to provide additional insights into data preparation and model generalization.

The results demonstrated that applying image enhancement significantly improved model accuracy and stability, especially with MobileNet. However, data augmentation negatively affected performance when applied to the training dataset. MobileNet achieved 77% accuracy on the external validation dataset, highlighting the need for high-quality, diverse datasets for further improvements. The study also identified that relying on limited computing resources, such as free cloud-based GPUs, presented challenges during the training process.

In conclusion, this research highlights the importance of data preprocessing and proper evaluation in building reliable AI models for brain tumor classification. Future work should explore advanced architectures, larger datasets, and 3D MRI data to further enhance accuracy and applicability in real-world medical scenarios.

Keywords:

Brain Tumer Detection and Classification, Deep Learning, CNN, Artificial Intelligence, Magnetic Resonance Imaging (MRI).

Introduction

Artificial Intelligence is a technology that allows computers to simulate human thinking, solve problems, and make decisions. Recently, the concept of generative AI has emerged, where AI can create original content. The rapid development of AI capabilities provides opportunities to apply it in various important sectors to support human progress. As a result, there is growing interest in using AI to improve healthcare services [1].

AI is becoming an important tool in healthcare, especially in medical imaging. AI systems are being used to analyze images like CT scans, x-rays, and MRIs to detect diseases and other conditions that might be missed by human specialists. Research shows that AI, especially when powered by advanced neural networks, can identify signs of diseases such as cancer with accuracy like human radiologists. By helping doctors catch early signs of illness and manage large amounts of medical images, AI is making healthcare more efficient and effective [2].

The brain is a complex and vital organ that controls many essential functions and is made up of billions of neurons, synapses, and nerve cells. Like other organs, it can develop abnormalities called brain tumors. Brain tumors are abnormal cell growths that vary in size and location within the brain. They can be life-threatening by destroying healthy cells [3]. Brain tumors are divided into two types: benign and malignant. Benign tumors do not spread to other tissues and are usually not life-threatening. Malignant tumors, on the other hand, grow quickly and can spread to other tissues, significantly impacting the patient's life [4].

Brain tumors include several common types, such as gliomas, meningiomas, and pituitary tumors, which can sometimes be fatal. Gliomas develop in the brain's glial cells, which provide support to nerve cells. They make up about 30% of all brain tumors and are often malignant. Meningiomas are usually benign and grow slowly, forming in the membranes around the brain. Although they do not originate inside the brain, they can cause symptoms by pressing on it as they grow. Pituitary tumors form in the pituitary gland at the base of the brain behind the nose. While most pituitary tumors are benign, some can cause the gland to produce excessive hormones, disrupting essential bodily functions [5].

Medical imaging methods, such as MRI and CT scans, are commonly used to detect cancerous cells and tissues in the body. MRI uses magnetic fields and radio waves to create high-resolution images, making it especially useful for visualizing the brain, spinal cord, joints, and internal organs. It is an essential tool for diagnosing and monitoring diseases like cancer. Compared to other imaging methods, MRI provides more precise and detailed results, making it highly effective for detecting and diagnosing brain tumors [4-5].

Experts analyze MRI images to determine if a brain tumor is present and identify its type. However, manually examining MRI images can be time-consuming, costly, and prone to errors, especially with large volumes of data. This can lead to delays in diagnosis and increased risks for patients. To address these challenges, computer-aided diagnosis systems have been developed, utilizing machine learning (ML) and deep learning (DL) techniques to classify, segment, and detect brain tumors effectively.

In this paper, I will use a convolutional neural network (CNN) to detect brain tumors from MRI images. I will work with various publicly available datasets and aim to develop a CNN model that can effectively and accurately identify brain tumors.

Related Work

In the study described in (Aamir et al., 2022), the authors conducted research using data collected from 233 patients treated at two state-owned hospitals in Guangzhou and Tianjin, China, between 2005 and 2010. The dataset consists of 3064 T1-weighted enhanced contrast brain MRI scans with a resolution of 512 × 512 pixels per image and a voxel spacing of 0.49 × 0.49 mm². It includes three types of brain tumors: Pituitary (930 images), Meningioma (708 images), and Glioma (1426 images), captured in three planes: axial, coronal, and sagittal. The dataset is provided in MATLAB format (.mat) which can be found at (https://figshare.com/articles/braintumordataset/1512427).

The study proposed a method to enhance the images before using them for training. However, the authors did not evaluate the model's performance on the original, unenhanced images, leaving the impact of this enhancement step unclear. They also performed feature extraction to train the model, but the effect of skipping this step was not explored. While the study achieved significant results in automated brain tumor classification, these limitations highlight areas for further research to assess the importance of preprocessing and feature extraction in improving model performance.

In the study presented in (Aurna et al., 2022) a comprehensive approach to brain tumor classification using MRI images, integrating multiple datasets and advanced deep learning techniques to achieve high accuracy and robustness. The datasets used in this study included three unique datasets—Brain Tumor Dataset 1, Dataset 2, and Dataset 3—along with a merged dataset combining all three, comprising a total of 10,620 MRI images. The dataset covered four categories: meningioma, glioma, pituitary tumors, and normal brain tissue, with all images resized to a resolution of 256 × 256 pixels. To enhance the dataset and mitigate overfitting, six augmentation techniques were applied, including horizontal flipping, rotation, zoom, height shift, width shift, and scaling.

The methodology centered on leveraging both pre-trained and custom-built Convolutional Neural Networks (CNNs) for feature extraction. Five pre-trained models—VGG-19, EfficientNet-B0, Inception-V3, ResNet-50, and Xception—were evaluated, alongside a custom CNN model. Among these, EfficientNet-B0, ResNet-50, and the custom CNN were identified as the best feature extractors and subsequently used to build an ensemble model. The extracted features underwent dimensionality reduction through Principal Component Analysis (PCA), significantly reducing the number of features while retaining nearly 100% variance with just 110 components.

The classification process employed a variety of classifiers, including Softmax, Random Forest, K-Nearest Neighbors (KNN), and AdaBoost, to determine the optimal model configuration. The proposed ensemble model demonstrated superior performance, achieving an average accuracy of 99.13%. The robustness of the model was further validated across all datasets using metrics such as accuracy, precision, recall, and F1 score.

To facilitate real-time application, the authors developed a user-friendly interface, enabling practical use in clinical settings. This interface supports the real-time validation of the proposed two-stage ensemble model, emphasizing the practical applicability of the research.

The study by (Muhammed ÇELİK and Özkan İnik, 2024) addresses the challenge of accurately classifying brain MRI images into four categories: glioma, meningioma, pituitary tumors, and no tumor. The authors used a publicly available dataset combining images from Figshare, SARTAJ, and Br35H, containing 7023 images. The dataset was split into training and testing sets in an 80-20 ratio, with images resized to 224 × 224 pixels and converted to RGB format.

The authors proposed a hybrid approach combining deep learning (DL) and machine learning (ML) techniques. They used nine state-of-the-art CNN models (e.g., DenseNet201, EfficientNetB0, ResNet50) and a newly designed CNN model to extract features from the images. These features were classified using ML algorithms, including Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbor (KNN). Hyperparameters for the ML classifiers were optimized using Bayesian optimization.

The results showed that the proposed hybrid approach achieved high classification accuracy. Among the models, the EfficientNetB0-SVM structure performed best, achieving an accuracy of 97.93%, while the proposed CNN-KNN model achieved 97.15%. The proposed CNN model alone achieved 95.66% accuracy. The study also highlighted the computational efficiency of the proposed models, with the CNN-KNN model being the most time-efficient.

This study addresses gaps in previous research, such as the lack of hyperparameter optimization for ML algorithms and limited use of hybrid DL-ML approaches. However, it has some limitations. The dataset may not fully represent real-world variability, and the model's performance on unseen datasets remains unexplored. Additionally, while the study avoided data augmentation, this might reduce the robustness of the models.

The study utilized the Brain Tumor MRI Dataset, a publicly available dataset on Kaggle, containing 7023 MRI images categorized into four classes: no tumor (2000 images), glioma (1621 images), meningioma (1645 images), and pituitary tumors (1757 images). All images were resized to 128 × 128 pixels and converted from grayscale to RGB format for richer feature extraction. The dataset lacks diversity as it only includes axial plane images, excluding coronal and sagittal views, which are critical for capturing complete spatial features of brain tumors. Additionally, the dataset size is relatively small, and the absence of augmentation techniques limits the model's robustness and generalizability.

The methodology involved using a fully convolutional neural network (FCNN) architecture to perform three sequential classification tasks:

Binary classification: Tumorous vs. non-tumorous brain. Multi-class classification (3 classes): Glioma, Meningioma, and Pituitary tumors. Multi-class classification (4 classes): Glioma, Meningioma, Pituitary tumors, and No tumor.

The performance of four optimizers—Adam, Nesterov momentum, AdaGrad, and RMSprop—was evaluated. The model included convolutional layers (32, 64, 128, and 256 filters with a kernel size of (3 × 3) and used the ReLU activation function for intermediate layers and Softmax or Sigmoid activation for output layers. Images were normalized using min-max scaling, and the data was evenly split into training, validation, and testing sets.

The results showed high accuracy across all tasks:

Binary classification: Nesterov momentum achieved 98% testing accuracy, excelling at distinguishing tumorous from non-tumorous brains. Three-class classification: Nesterov achieved 92% accuracy, with challenges in distinguishing glioma and meningioma. Four-class classification: Nesterov performed best, achieving 95% accuracy, with the lowest performance in identifying meningioma.

Despite the promising results, the study has notable limitations. The dataset size is not sufficiently large, which may impact the generalizability of the results. Furthermore, the exclusion of coronal and sagittal plane images reduces the model's ability to capture comprehensive tumor features. The absence of data augmentation also limits the robustness of the model against overfitting and variability in real-world scenarios.

The study (Saleh, Sukaik and Abu-Naser, 2020) used the Brain Tumor MRI Dataset [10] from Kaggle, consisting of 4480 images categorized into four classes: glioma, meningioma, pituitary tumors, and no tumor. The dataset was divided into training (2880 images), validation (800 images), and testing (800 images) sets. Each class in the training set contained 520 images, while the validation and testing sets had 200 images per class. All images were resized to 256 × 256 pixels for processing. To address the issue of limited data, the authors applied data augmentation, particularly for the underrepresented meningioma class, generating additional images to mitigate overfitting.

The methodology employed five pre-trained CNN architectures: Xception, ResNet50, InceptionV3, VGG16, and MobileNet, which were retrained with the dataset. The original classifiers of these architectures were replaced with a custom classifier tailored for brain tumor classification. Each model was trained for 50 epochs, and performance was evaluated using F1-score accuracy.

The results demonstrated high performance across the models: Xception achieved the best accuracy at 98.75%. ResNet50 followed with 98.50%. Other models like InceptionV3, VGG16, and MobileNet achieved accuracies of 98.00%, 97.50%, and 97.25%, respectively.

Despite the promising results, the study has significant limitations, while the dataset is relatively small, with only 4480 images, which may limit the generalizability of the results. A larger dataset is needed to ensure the robustness and reliability of the models. And the dataset only includes images in the axial plane, excluding coronal and sagittal views. This limits the model's ability to capture comprehensive spatial features of brain tumors.

While data augmentation was used, the small dataset size may still lead to overfitting, especially with complex architectures like Xception and ResNet50. On the other hand, the models were tested only on the provided dataset, and their performance on real-world clinical data remains unexplored.

The study (Dikande et al., 2024) utilized the Brain Tumor MRI Dataset [12], comprising 7023 images divided into four categories: no tumor (2000 images), glioma (1621 images), meningioma (1645 images), and pituitary tumors (1757 images). All images were resized to 128 × 128 pixels and converted from grayscale to RGB to enhance feature extraction. The dataset was split into training, validation, and testing sets for the classification tasks. Although this dataset is widely used and freely accessible, it has certain limitations. It contains only axial-plane images, excluding coronal and sagittal views, and its relatively small size may limit the robustness of the results and the generalizability of the models.

The methodology focused on developing a fully convolutional neural network (FCNN) for sequential classification tasks. The first task involved binary classification to differentiate between tumorous and non-tumorous brains. The second task expanded to three-class classification, identifying glioma, meningioma, and pituitary tumors. Finally, the third task included a four-class classification by adding the "no tumor" category. Four optimizers, Adam, Nesterov momentum, AdaGrad, and RMSprop, were tested to determine the best-performing model. Among these, Nesterov momentum demonstrated superior performance, especially in multi-class classification tasks. It achieved 98% accuracy on validation and test sets for binary classification, 92% for three-class classification, and 95% for four-class classification.

While the results highlight the effectiveness of the Nesterov-based FCNN model, achieving high precision, recall, and F1 scores across all tasks, the study has notable limitations. The dataset size is small and lacks diversity, as it does not include coronal and sagittal views, which are crucial for capturing the full spatial features of brain tumors. Additionally, the lack of data augmentation increases the risk of overfitting, particularly given the limited size of the dataset. The model’s performance on unseen or real-world clinical data was not evaluated, raising concerns about its generalizability. Future research should aim to use larger, more diverse datasets and incorporate augmentation techniques to improve the robustness and applicability of the models in clinical settings.

The research gap lies in the limitations of previous studies that used smaller datasets, often with a maximum of around 10,000 images, which is not sufficient for robust model training and testing. Many of these datasets lack diversity, as they include only axial-plane images and exclude coronal and sagittal views, limiting the ability to capture full spatial features of brain tumors.

Additionally, the impact of image enhancement steps was not evaluated, leaving it unclear how these enhancements affect model performance. Similarly, feature extraction was performed in some studies, but the consequences of skipping this step were not explored. While data augmentation was applied, the effectiveness of this technique was not adequately addressed.

Our research will fill this gap by using a larger and more comprehensive dataset to improve classification accuracy. We will address the limitations of prior work by evaluating the impact of enhancement, feature extraction, and augmentation. Furthermore, we will explore different CNN architectures to compare and analyze their performance, providing a more thorough understanding of the factors that influence classification accuracy.

Data collection and Description

In this section, we will discuss the data collected for this study, which includes both primary and secondary data. Primary data refers to information collected directly from original sources. For this study, primary data will be gathered through interviews with two experts in the field, Dr. Rami AlOran and Dr. Murad A. Yaghi, to gain insights into brain tumor diagnosis and classification processes.

Secondary data, on the other hand, refers to data that has already been collected and published by other sources. In this study, the secondary data is the dataset used to train and evaluate the deep learning models. This dataset will provide the MRI images necessary for building and testing the classification models.

By combining primary data from experts and secondary data from existing datasets, this study ensures a comprehensive approach to understanding and improving brain tumor classification. The section will also define the differences between primary and secondary data and explain how each contributes to the research objectives.

Primary Data

The primary data for this study was collected through interviews with two experts in the fields of Artificial Intelligence, Data Science, and Bioinformatics: Dr. Rami AlOuran and Dr. Murad A. Yaghi. Their expertise provides valuable insights into the challenges and advancements in brain tumor classification using AI-based methods.

Dr. Rami AlOuran is an Assistant Professor in the Department of Data Science and Artificial Intelligence at Al Hussein Technical University. He has extensive experience in bioinformatics, machine learning, and computational regulatory genomics. Prior to his academic role, Dr. Rami was a senior bioinformatics and data analyst at Baylor College of Medicine, where he worked in the Laboratory for Integrative Functional Genomics. He holds a Ph.D. and M.S. in Electrical Engineering and Computer Science from Ohio University and a B.S. in Computer Engineering from Mutah University, Jordan.

Dr. Murad A. Yaghi is also an Assistant Professor in the Department of Data Science and Artificial Intelligence at Al Hussein Technical University. His research focuses on intelligent systems, robotics, and deep reinforcement learning, with additional interests in optimization and AI applications in climate modeling and simulation. Dr. Murad holds a Ph.D. in Computer Engineering from Hacettepe University in Turkey, where his thesis explored fractional-order feedback control of nonlinear aerial systems. He also holds an M.Sc. in Computer Engineering with a specialization in Embedded Systems from Yarmouk University, Jordan, and a B.Sc. in Electrical Engineering from Hashemite University, Jordan.

The interviews aimed to gather their perspectives on the application of AI and machine learning in medical imaging, specifically in brain tumor detection and classification. Their responses provide a foundation for understanding current practices, limitations, and potential advancements in the field.

Questions and Responses

In your opinion, what are the most important factors for building an accurate AI model for medical image classification?

Dr. Rami AlOuran: "Having accurate data, having enough data since you are employing deep learning models, and having a doctor as part of the team for advice."
Dr. Murad A. Yaghi: "It is crucial to integrate domain expertise with technical knowledge. Accurate labeling, a balanced dataset, and iterative testing are critical for achieving reliable results."

What qualities make a medical imaging dataset suitable for training machine learning models?

Dr. Rami AlOuran: "This depends on the application, but it could include clear images without noise, correct labeling, and concise annotations."

Dr. Murad A. Yaghi: "Datasets should not only be noise-free and well-labeled but should also include multiple classes and variations to reflect real-world complexities."

How important is having diverse datasets (e.g., different planes like axial, coronal, sagittal) for improving model performance?

Dr. Rami AlOuran: "This helps in creating a rich dataset, which improves the model's ability to generalize, ultimately enhancing testing performance."

Dr. Murad A. Yaghi: "Having datasets with multiple planes allows the model to learn spatial relationships, improving its capacity to generalize across unseen data."

What role do data preprocessing techniques, like enhancement or normalization, play in improving classification accuracy?

Dr. Rami AlOuran: "Preprocessing techniques play a very important role. However, for each model trained and employed, the proper processing technique should be used."

Dr. Murad A. Yaghi: "Preprocessing techniques such as data enhancement and normalization are foundational for consistency in training. Testing multiple preprocessing methods can help determine the best approach."

What do you think about using feature extraction techniques in deep learning models? Are they necessary, or can they be skipped?

Dr. Rami AlOuran: "As the 'no free lunch' theorem explains, both approaches should be tried and compared to determine which gives the best performance."

Dr. Murad A. Yaghi: "Deep learning models often learn features automatically, but feature extraction can be beneficial in cases of smaller datasets or when domain-specific features are essential."

How can researchers ensure their models are generalizable and not overly dependent on a specific dataset?

Dr. Rami AlOuran: "Augmentation can help with such problems, as well as collecting more diverse data."

Dr. Murad A. Yaghi: "Using a combination of cross-validation, data augmentation, and regularization techniques is essential to make models robust and generalizable."

Secondary Data

The secondary data for this study consists of two datasets: one for training and testing the models and another for validation.

The first dataset used for training and testing is the **Brain MRI ND-5 Dataset** (Alvarado, 2024), sourced from IEEE DataPort. This dataset contains a total of 17,891 MRI images, with 13,929 images for training and 3,962 images for testing, divided into four categories: glioma, meningioma, pituitary tumors, and no tumor. A key strength of this dataset is its inclusion of images from all three planes: axial, coronal, and sagittal, providing a comprehensive spatial representation of brain tumors (Figure 1). The dataset was created by merging five publicly available datasets, enhancing its diversity and robustness, which is essential for training deep learning models for brain tumor classification and detection.

For validation, the **Figshare Brain Tumor Dataset** is used. This dataset contains 3064 T1-weighted contrast-enhanced images from 233 patients, divided into three tumor categories: meningioma (708 images), glioma (1426 images), and pituitary tumors

(930 images). The dataset is split into four subsets, each provided in a zip file, with 5-fold cross-validation indices included to facilitate model evaluation.

By using the Brain MRI ND-5 Dataset for training and testing and the Figshare dataset for validation, this study leverages diverse and comprehensive data sources to ensure robust training and accurate performance evaluation of the classification models.

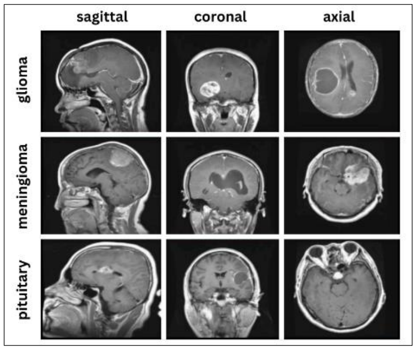

Figure 1 - Three distinct subtypes of brain tumors.

Research Approach and Methodologies

This section provides a comprehensive overview of the research approach and methodologies used in the study. It begins with an explanation of the Onion Research Model, which outlines the layers of the research process. Each layer, including Philosophy, Theory Development Approach, Methodological Choice, Research Strategy, Time Horizons, and Techniques and Procedures, is discussed in detail, with justification provided for the choices made. Additionally, the section includes a chart that illustrates the research methodology, breaking down the key steps taken in the study and critically evaluating each part to explain and justify the methods and analyses used.

Onion Research Model

The Onion Research Model is a structured framework used to guide researchers in developing their research methodology. Created by Saunders et al. in 2007, it visually resembles an onion, with multiple layers representing the different stages of the research process. The model emphasizes starting from the outermost layer and working inward systematically to address key decisions required for conducting research (Saunders et al., 2007).

This approach ensures that every step, from philosophical foundations to data collection and analysis, is carefully planned and aligned with the research objectives. The Onion Research Model is widely used because it provides a clear, logical, and comprehensive structure for designing effective research studies.

Philosophy

The philosophy layer in the research onion focuses on the beliefs and assumptions that guide how knowledge is developed and understood in the research. It determines how researchers view reality and approach their study. Different philosophical approaches include:

Positivism: Positivism assumes that reality is objective and can be observed and measured through scientific methods. Researchers using this approach rely on quantitative data and aim to find universal truths.

Interpretivism: Interpretivism focuses on subjective understanding and emphasizes that reality is shaped by individuals' perceptions. This approach often uses qualitative methods to explore meanings and experiences.

Critical Realism: This approach assumes that while there is a reality independent of human perception, our understanding of it is influenced by social, cultural, and contextual factors. It combines elements of both qualitative and quantitative methods.

Postmodernism: Postmodernism challenges the idea of a single truth or universal knowledge. It emphasizes diversity, context, and the idea that knowledge is socially constructed and fragmented.

Pragmatism: Pragmatism focuses on practical outcomes and recognizes that both subjective and objective approaches can be useful. Researchers using pragmatism adopt methods based on what works best to achieve their objectives, often mixing qualitative and quantitative methods.

In my research, I have chosen to adopt **pragmatism** methodology. This approach is particularly suitable because my study involves testing and evaluating multiple deep learning methods to classify brain tumors. Pragmatism allows for flexibility, enabling me to use different techniques and approaches based on their effectiveness rather than being constrained by a single philosophical view. As we experiment with various deep learning models and methods, the results and their impact on classification accuracy will guide our decisions. In this way, experience becomes the ultimate judge of the methods' success, aligning perfectly with the exploratory and results-driven nature of this study. This approach ensures that the research remains adaptive, open to innovation, and focused on achieving practical solutions.

Approach to theory Development

This layer focuses on how researchers design their study and move from theory to data or from data to theory. There are three main research approaches:

Deduction: This approach starts with a theory or existing knowledge. Researchers create hypotheses based on this theory and test them using data. Deductive research moves from general ideas to specific conclusions and is often used in quantitative studies.

Induction: Inductive research starts with data collection. Researchers observe patterns or trends in the data and then develop a theory based on these observations. It moves from specific data to general conclusions and is commonly used in qualitative research.

Abduction: This approach begins with an observation that cannot be explained by existing theories. Researchers propose new hypotheses or explanations and test them to develop a theory. Abduction is often used for exploring new ideas or solving unusual problems.

In my research, I have chosen a **deductive** approach because it aligns with the nature of my study. I have reviewed numerous studies on brain tumor classification using deep learning and machine learning, which have shown promising results. Based on this existing knowledge, I aim to test these approaches on a larger and more diverse dataset. I will build on the established theories and techniques, applying them to different resources and testing various methods. This allows me to evaluate how well these techniques perform on a broader dataset, identify potential improvements, and confirm the findings of previous studies.

Methodological Choice

The methodological layer focuses on how the research is conducted and the methods used to collect and analyze data. There are three main methodological approaches:

Mono Method: This approach uses a single method, either qualitative or quantitative, to collect and analyze data. It is straightforward but limited in scope.

Mixed Method: This combines both qualitative and quantitative approaches to provide a more comprehensive view of the research problem.

Multi-Method: This uses multiple methods within either the qualitative or quantitative approach. It allows for deeper analysis while staying focused on one type of data.

I have chosen the **multi-method quantitative approach**. The primary focus of my study is on a large dataset of MRI images, which will be analyzed using various deep learning techniques. This requires multiple quantitative methods to thoroughly evaluate and test the dataset. Additionally, I conducted interviews with Dr. Murad and Dr. Rami to gather expert insights, which support the research process but are secondary to the dataset analysis. By focusing on multi-method quantitative techniques, my research ensures robust and detailed results while leveraging expert opinions to enhance the study's depth.

Research Strategy

It determines how the researcher collects and interprets data. Common research strategies include:

Experiment: Controlled testing to establish cause-and-effect relationships. Survey: Collecting data from a group of people through questionnaires or interviews. Archival Research: Using existing records or datasets to conduct the study. Case Study: In-depth exploration of a specific instance or situation. Ethnography: Immersive observation of a cultural or social group. Action Research: Solving practical problems while actively engaging with the research process. Grounded Theory: Developing a theory based on the analysis of collected data. Narrative Inquiry: Exploring personal stories and experiences to understand a phenomenon.

In my research, I have chosen a combination of **archival research**, **action research**, and **narrative inquiry**.

Archival research is appropriate because I am working with a large pre-existing dataset of MRI images, which forms the foundation of my study. **Action research** fits as I am actively applying deep learning techniques to the dataset, testing different approaches to improve brain tumor classification accuracy. **Narrative inquiry** is included to incorporate the expert insights gathered from my interviews with Dr. Murad and Dr. Rami, which provide valuable context and depth to the research process. Together, these strategies ensure a comprehensive and practical approach to the study.

Time Horizon

The time horizons layer focuses on the time frame within which the research is conducted. There are two main approaches:

Cross-Sectional: This approach is independent of time, as it focuses on collecting and analyzing data at a single point in time. It is suitable for studies that do not require tracking changes over a period.

Longitudinal: This approach is time-related and involves studying the same variables or subjects over an extended period to observe changes and trends.

In my research, I have chosen a **cross-sectional approach**. Since my study focuses on analyzing a large dataset of MRI images and testing various deep learning methods at a single point in time, a cross-sectional method is appropriate. This allows me to evaluate and compare the performance of different techniques without the need to track changes over time.

Techniques and Procedures

The techniques and procedures layer focuses on the specific methods used to collect, process, and analyze data. This includes selecting appropriate tools, frameworks, and processes to ensure the research objectives are met.

In my research, the main data source is a large dataset of MRI images, which will be processed using deep learning techniques. Data collection includes using pre-existing archival data and interviews with experts to gain additional insights. Preprocessing methods, such as resizing and normalization, will be applied to the dataset to ensure consistency.

For analysis, various deep learning models will be implemented to classify brain tumors. The results will be evaluated using standard metrics like accuracy, precision, and recall. Validation will include testing the models on a separate validation dataset to ensure their generalizability and robustness. This approach ensures a comprehensive and structured process to achieve reliable and meaningful results.

Research Methodology

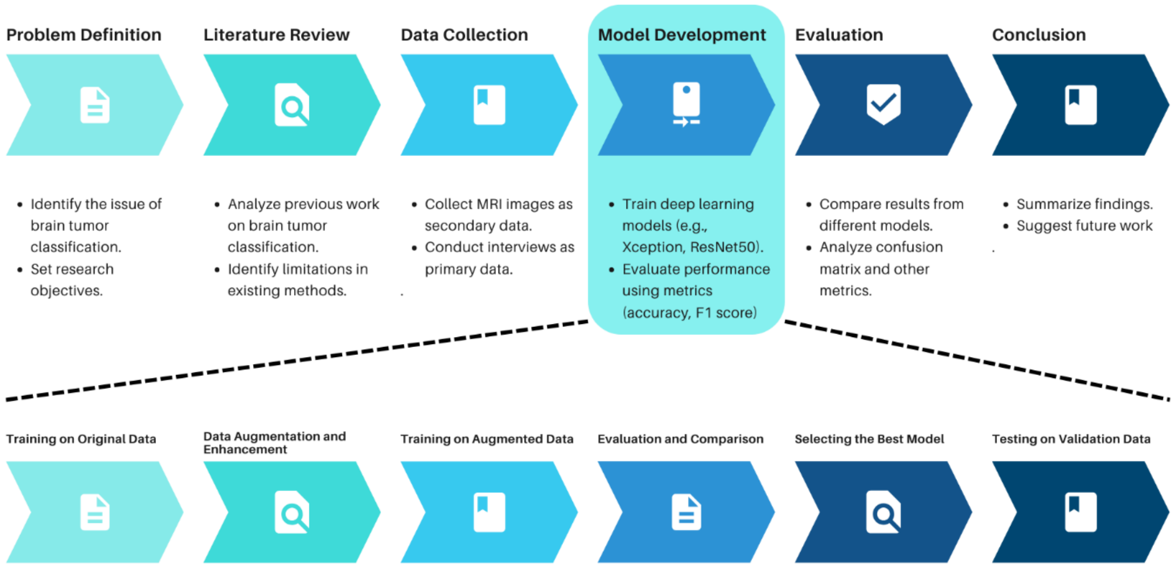

Figure 2 - The structure of the research approach

Problem Definition: Brain tumor classification is a critical problem in medical imaging. The goal is to accurately classify brain tumors using MRI scans to assist in diagnosis and treatment. In this research, the objective is to use advanced deep learning techniques to improve classification accuracy, addressing limitations in existing methods.

Literature Review: A detailed review of previous studies on brain tumor classification was conducted. This included analyzing existing deep learning approaches like Xception, ResNet50, and others. Limitations in these studies were identified, such as smaller datasets, lack of data augmentation, and limited evaluation on unseen data.

Data Collection:
Secondary Data: MRI images were collected from publicly available datasets. These images represent four classes: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. **Primary Data:** Interviews were conducted with experts, Dr. Rami AlOuran and Dr. Murad Yaghi, to gain insights into challenges and advancements in AI-based medical imaging.

Model Development
Step 1: Five pre-trained deep learning architectures (Xception, ResNet50, InceptionV3, VGG16, and MobileNet) were trained on the original dataset without applying data augmentation or enhancement.
Step 2: The dataset was then enhanced and augmented with techniques like rotation, flipping, and brightness adjustments to create a more diverse and robust dataset. You can see the impact of the enhancement step over the images in figure 3.
Step 3: The best two of the five architectures were retrained on the augmented dataset to evaluate improvements in performance.
Step 4: The results from the models were compared using metrics like accuracy, precision, recall, and F1 score. The best-performing architecture was selected.
Step 5: The selected model was tested on unseen validation data to ensure its generalizability and robustness.

Evaluation: The results from each model were compared in detail. A confusion matrix was used to visualize classification accuracy for each class. Metrics like F1 score, precision, and recall provided deeper insights into the performance of the models. This step ensured that the best model met the research objectives effectively.

Conclusion: The findings were summarized, highlighting the performance of the best model and its ability to classify brain tumors accurately. Future recommendations included using larger and more diverse datasets, testing new deep learning architectures, and applying further enhancements to improve model performance.

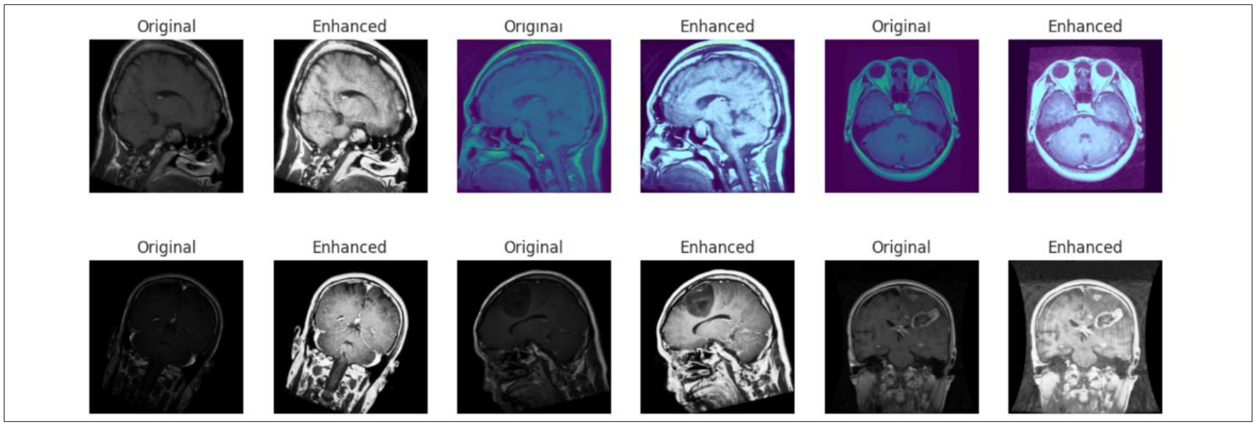

Figure 3 - The impact of the Enhancement method over the images

Results and Discussion

The first step in the Model Development phase involved training five pre-trained deep learning architectures (Xception, ResNet50, InceptionV3, VGG16, and MobileNet) on the original dataset without any data augmentation or enhancement. This step aimed to establish baseline performance metrics for comparison with subsequent steps involving augmented and enhanced data. To streamline the process, the dataset was uploaded from Google Drive to the local directory of the Google Colab environment. This approach reduced training time significantly by enabling faster data access. It can be compared to moving your resources closer to the workspace, eliminating delays associated with accessing them from a remote location repeatedly.

The five architectures were trained with the following hyperparameters: Learning Rate: 0.01, Batch Size: 32, Epochs: 20
Below, we discuss the results of this step, comparing the performance of each model and analyzing their strengths and limitations.

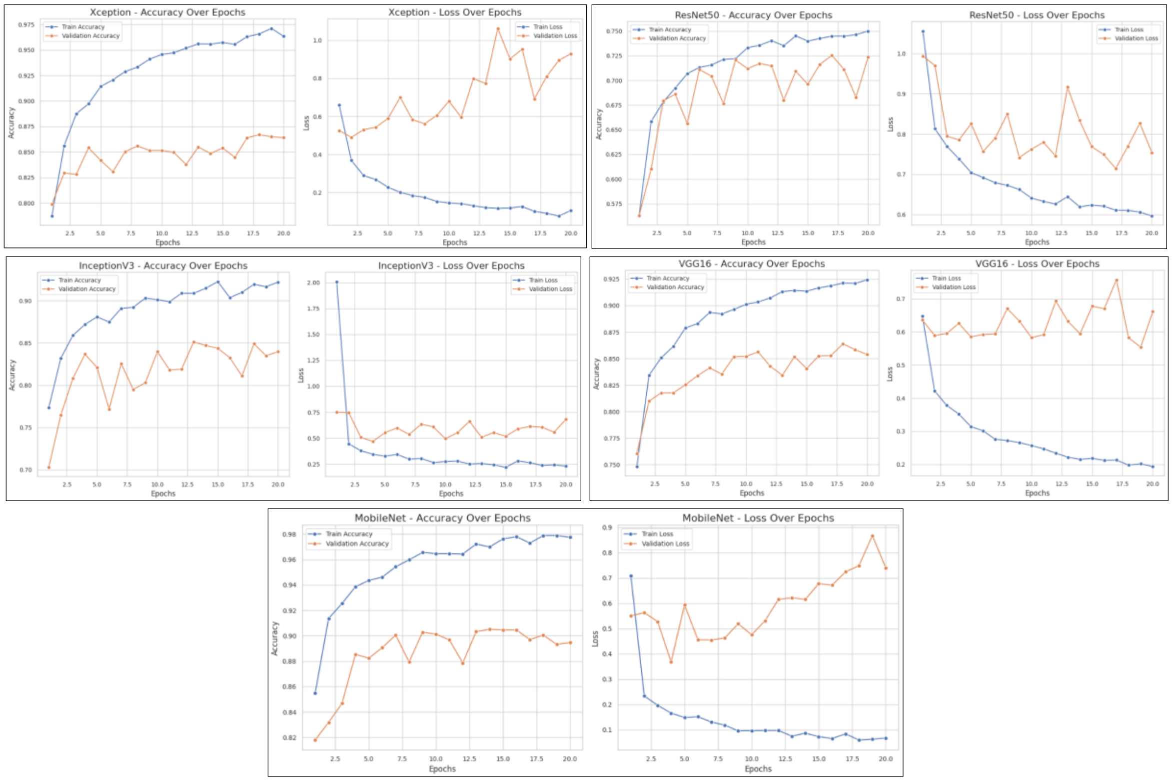

Figure 4 - The results of the five architectures before Image enhancement and Data augmentation

This section presents an evaluation of the performance of five pre-trained deep learning architectures (Xception, ResNet50, InceptionV3, VGG16, and MobileNet) trained on the original dataset without data augmentation or enhancement. The models' performance is analyzed based on their training and validation accuracy, as well as their training and validation loss over 20 epochs.

The Xception model achieved a steady improvement in training accuracy, reaching approximately 97%, while the validation accuracy stabilized at around 92%. The training loss decreased consistently, indicating effective learning. However, the validation loss fluctuated and remained higher than the training loss, suggesting potential overfitting or instability during validation. This indicates that Xception performs well but could benefit from further techniques to improve generalization.

The ResNet50 model demonstrated rapid improvement in training accuracy, stabilizing at 85%, while the validation accuracy fluctuated and remained close to 78%. The training loss decreased significantly, but the validation loss exhibited notable fluctuations. This performance indicates that while ResNet50 captures patterns effectively during training, it struggles to generalize as well as Xception.

InceptionV3 showed strong performance, with training accuracy steadily increasing to 94% and validation accuracy stabilizing around 90%. The training loss decreased consistently, and the validation loss remained relatively stable compared to other models. These results suggest that InceptionV3 achieves a good balance between training and validation performance, making it a strong candidate for further evaluation.

The VGG16 model achieved a training accuracy of approximately 92%, while the validation accuracy stabilized at 88%. Training loss decreased consistently over the epochs, while validation loss fluctuated slightly. Although the model shows good generalization, the fluctuations in validation loss suggest minor instability during training.

MobileNet achieved training accuracy of over 95%, with validation accuracy stabilizing at around 91%. The training loss decreased sharply, and the validation loss remained low and relatively stable compared to the other models. These results highlight MobileNet's effectiveness in achieving high accuracy and stability during training and validation.

All models effectively reduced training loss, confirming successful learning. However, the fluctuations in validation loss across some models highlight the need for further refinement, such as applying data augmentation or additional regularization techniques.

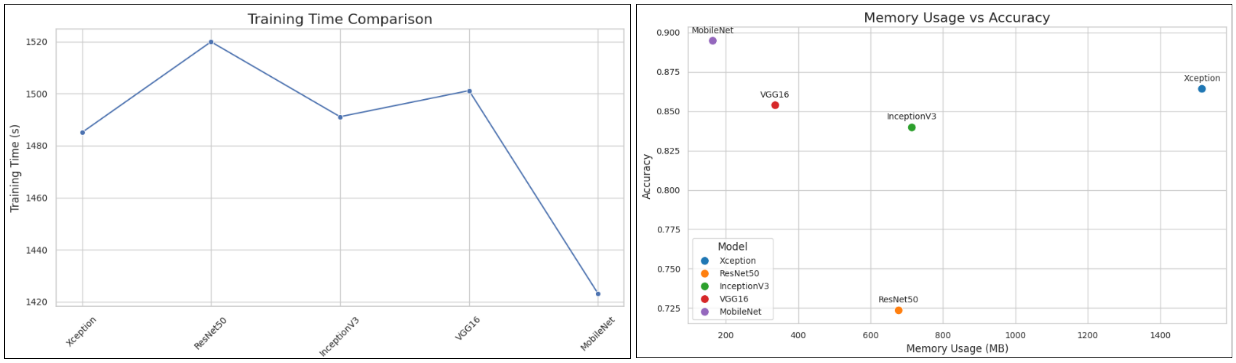

Figure 5 - Training time comparison and Memory usage vs Accuracy

The first plot compares the training time (in seconds) for each model. ResNet50 took the longest time, approximately 1520 seconds, likely due to its deeper architecture and higher computational complexity. And MobileNet had the shortest training time at around 1420 seconds, showcasing its lightweight design optimized for efficiency. Other models like Xception, InceptionV3, and VGG16 fell between 1480 to 1510 seconds, indicating moderate computational demand compared to ResNet50.

The second plot visualizes the relationship between memory usage (in MB) and accuracy for the models. Xception has the highest memory usage (~1400 MB) and achieves the second highest accuracy (~87%), indicating its effectiveness but also its computational demands. On the other hand, MobileNet balances efficiency and accuracy well, achieving high accuracy (~90%) with the lowest memory usage (~200 MB). Which is a really big difference between the two.

InceptionV3 and VGG16 are in the mid-range for both memory usage (~400 MB and ~500 MB, respectively) and accuracy (~84% and ~86%). Also ResNet50 has high memory usage (~800 MB) but lower accuracy (~78%), making it less efficient compared to the other architectures.

MobileNet demonstrates the best trade-off between accuracy and resource usage, making it ideal for environments with limited computational power, which the best model in my case because I work on Google Colab with a mid-performance Laptop. These insights highlight the strengths and trade-offs of each model, supporting the selection of MobileNet and Xception as strong candidates for further analysis.

Results indicated that achieving high accuracy in brain tumor classification with deep learning models was challenging without image enhancement and data augmentation. Therefore, the next stage of this research will involve applying these techniques to the dataset and re-evaluating the performance of the Best two chosen models (Xception and MobileNet) using the same hyperparameter settings. I applied Enhancement which can be described like the image's brightness and contrast are improved using histogram equalization. Specifically, the image is converted to the YUV color space, and histogram equalization is applied to the Y channel (luminance), which represents the brightness. This step enhances the visibility of features in the image by redistributing the intensity values more evenly. And I got the results below, which didn't make sense to me.

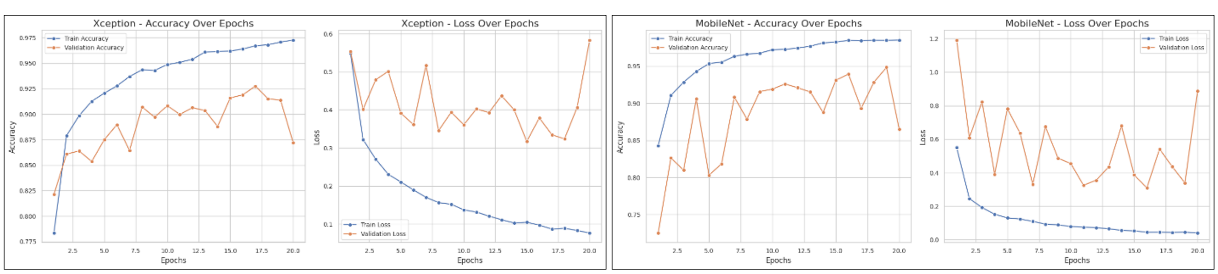

Figure 6 - The results of the best two architectures after Image enhancement and Data augmentation

Following the application of image enhancement to both the training and testing datasets, and data augmentation techniques to the training dataset only, the goal was to expose the model to a wider variety of training images. This approach was intended to improve generalization, under the assumption that augmented training images would help the model handle variations while the testing dataset remained unchanged to simulate real-world scenarios, where MRI images are not flipped or cropped.

The accuracy plots for both models, Xception and MobileNet, indicate good performance on the training dataset, with a steady increase in accuracy over epochs. However, the validation accuracy fluctuates significantly and does not show consistent improvement. This discrepancy between the smooth training accuracy curve and the erratic validation accuracy suggests overfitting. The models are learning to perform well on the augmented training data but fail to generalize effectively to the testing data.

The loss plots further highlight this issue. For both models, training loss decreases consistently, indicating that the models are learning effectively from the augmented training data. Validation loss, however, fluctuates randomly and remains much higher, reinforcing the observation that the models struggle to generalize to the unaltered testing data.

These results suggest that the data augmentation applied only to the training dataset may have introduced inconsistencies between the training and testing datasets. The augmented training images likely diverged too far from the original testing images, leading to poor performance during testing. This highlights a mismatch between the training environment and the real-world application, where testing MRI images remains unaltered.

Given these results, I decided to retry the training process without data augmentation to ensure consistency between the training and testing datasets using MobileNet architecture. This adjustment aims to evaluate whether the removal of augmentation will improve the model's ability to generalize to unaltered testing images. So, we got the results below.

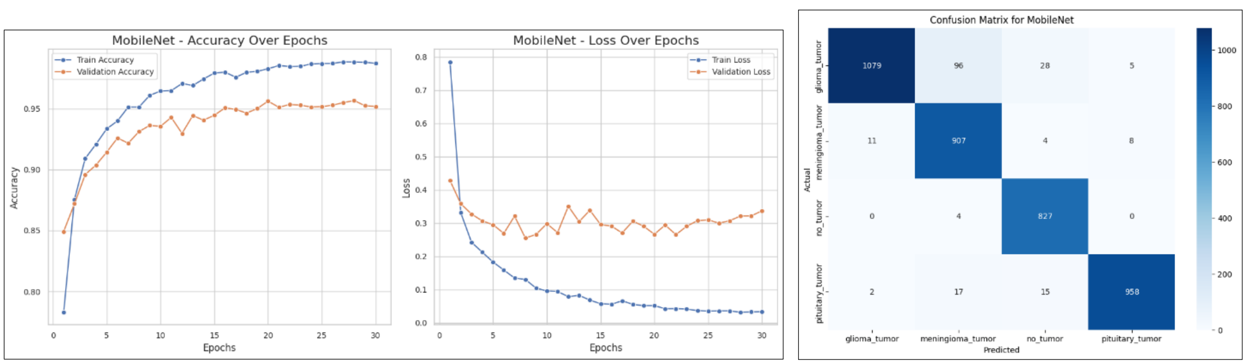

Figure 7 - The results of the MobileNet without data augmentation

The results demonstrate significant improvements in the stability and performance of the MobileNet model after training on the enhanced images without augmentation. The accuracy plot shows that the validation accuracy stabilizes over the epochs and follows a similar upward curve to the training accuracy. This indicates that the model is generalizing better to the testing dataset. The validation accuracy reaches a consistent level close to the training accuracy, demonstrating reduced overfitting compared to previous experiments with augmentation.

The loss plot further supports these findings. Both the training and validation loss decrease steadily over the epochs, with the validation loss maintaining a more stable trajectory. This highlights that the model is learning effectively and that the gap between training and validation performance has narrowed, improving overall model reliability.

The confusion matrix provides deeper insights into the classification performance across the four classes:

The model achieves high accuracy for the glioma tumor class, with 1,079 correct predictions and only a small number of misclassifications. Similarly, the meningioma tumor and pituitary tumor classes are well-classified, with 907 and 958 correct predictions, respectively, and minimal misclassification to other classes. The no tumor class shows a strong performance with 827 correct predictions and no significant confusion with other tumor types. These results suggest that training MobileNet on enhanced images without augmentation was an effective strategy. The approach improved both stability and accuracy, addressing the inconsistencies observed in the earlier experiments. This also confirms that the model performs well across all tumor categories, achieving robust generalization to unaltered testing data.

In addition to this, the MobileNet model was tested on the validation dataset **(Figshare Brain Tumor Dataset)** and achieved an accuracy of **77%**. While this indicates that the model is functioning, it also highlights limitations in its performance on previously unseen data. The confusion matrix provides a more detailed breakdown of the model's predictions (figure 8). This result raises questions about the dataset's quality and labeling accuracy. It is challenging to determine whether the model's lower accuracy is due to its limitations or issues with the dataset itself. To make a definitive assessment, a high-quality dataset with accurate and reliable labels is essential.

As part of the research, a graphical user interface (GUI) was developed to deploy the trained model for practical use. The GUI allows users to upload MRI images and receive real-time predictions of brain tumor classes (glioma tumor, meningioma tumor, pituitary tumor, or no tumor). This user-friendly interface is designed to make the model accessible for non-technical users, demonstrating its potential application in real-world medical scenarios. The GUI ensures seamless integration between the model and its end-users, providing an intuitive platform to test and validate the predictions. (figure 9)

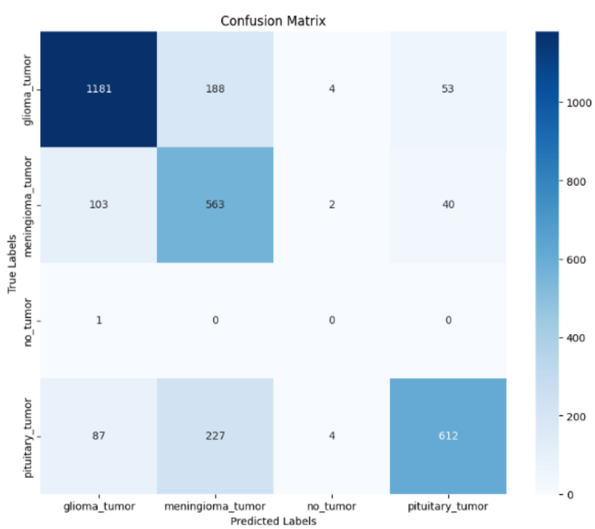

Figure 8 - The confusion Matrix of MobileNet tested on the validation dataset

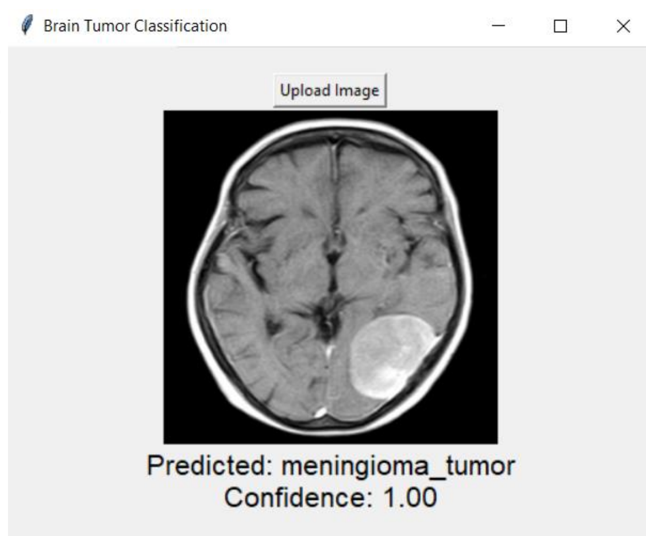

Figure 9 - The GUI window

Conclusion and Recommendations

The results improved clearly throughout the experiments. At first, training the five models without any image enhancement or data augmentation gave inaccurate predictions and poor performance on the testing dataset. To fix this, image enhancement was applied to both the training and testing datasets, and data augmentation was added to the training dataset. However, when testing the MobileNet and Xception models, the results stayed the same or became slightly worse, with unstable validation performance.

Because of this, I decided to remove data augmentation but kept the image enhancement. This change gave much better results, with stable learning curves and more accurate predictions, especially with the MobileNet model. This shows how important it is to have consistency between training and testing datasets to improve the model's performance.

The aim of this study was to investigate the use of deep learning for predicting brain tumors and to evaluate its potential accuracy for application in the medical field. The findings revealed the effects of image enhancement and data augmentation on model performance. Additionally, the use of a large and diverse dataset provided realistic feedback on the models' predictive ability, which is often missing in other studies that rely on smaller datasets. This study highlights the importance of proper data preparation and consistency to achieve reliable results.

Based on the findings, the following **recommendations** can be made:

Experiment with Hyperparameters: Future studies should explore a wider range of hyperparameters, such as learning rate, batch size, and optimizers, to further improve performance.

Utilize Advanced Computing Resources: Using a supercomputer or a dedicated GPU cluster would allow for more extensive experiments and faster training, overcoming the limitations of using Google Colab and limited free GPU resources.

Test Additional Techniques: Incorporating advanced techniques, such as transfer learning with fine-tuning and ensemble methods, could further enhance prediction accuracy.

In the **future**, the scope of this research could be expanded to explore more deep learning architectures, including more complex or novel models. Additional steps could include applying segmentation techniques to detect and outline tumors in MRI images, providing more actionable insights for medical professionals. Developing a deep learning model capable of handling 3D MRI images could also be a promising direction. Such a model could not only predict the presence of a brain tumor but also highlight the tumor's location in three-dimensional space.

Furthermore, continuously feeding the model with new and diverse data could help improve its accuracy and robustness. The ultimate goal would be to create a reliable deep learning system that can be fully trusted and implemented in the medical field for assisting in brain tumor detection and diagnosis.

Reflections

The reflection section critically evaluates the research journey, focusing on the methodologies employed, their effectiveness, and areas for improvement. This part delves into the decisions made during the research process, alternative approaches considered, and lessons learned from the outcomes. It highlights how the selected methodologies shaped the research results and identifies areas where adjustments could have enhanced the study. By reflecting on these aspects, this section aims to provide a balanced perspective on the research experience while offering actionable insights for future work.

Selected Research Methodology

The research journey provided valuable insights into the effectiveness of the selected research methodology, highlighting both its merits and limitations. The literature review was a foundational step in shaping the research process. It involved analyzing previous studies on brain tumor classification and identifying gaps, such as the reliance on smaller datasets, lack of data augmentation, and limited evaluation on unseen data. This critical review provided the rationale for using a larger dataset and testing the effects of enhancement and augmentation techniques. However, a limitation of this step was the limited availability of papers that explored larger datasets or alternative preprocessing techniques, which required more effort to contextualize my work within the broader research landscape.

The review process helped me develop the skill of critically analyzing methodologies and identifying practical challenges in prior studies. By bridging these gaps, the research aimed to offer a more comprehensive exploration of deep learning’s potential in medical imaging.

The data collection process was a key part of this research and included both primary data from expert interviews and secondary data from existing datasets. This combined approach gave the study a strong foundation by including both theoretical insights and practical data for model development.

The primary data came from interviews with two experts: Dr. Rami AlOuran and Dr. Murad A. Yaghi. They shared valuable insights about using Artificial Intelligence (AI) in brain tumor classification, the importance of diverse datasets, and how to ensure generalization in AI models. These interviews helped identify key factors for improving model performance. However, due to time constraints, the insights from these interviews were not fully applied to the model design and experiments. This limited the impact of their valuable knowledge on the study, leaving an area for improvement in future research.

To be on consideration the reliance on public datasets raised concerns about the accuracy of the image labels, which could have impacted the model’s performance. In addition to this, the Figshare dataset’s use of only axial images reduced its ability to fully test the model’s generalization. So future research should focus on ensuring high-quality data and better integration of expert insights to produce stronger and more reliable outcomes.

The research had several strengths and limitations. One key merit was the comprehensive literature review, which identified critical gaps in previous research and guided the selection of methods and datasets. Additionally, structured data handling through the use of secondary datasets and preprocessing techniques demonstrated how data preparation significantly influences model performance. Another strength was the practical implementation of a GUI, which showcased the potential application of the research outcomes in real-world scenarios, bridging the gap between theoretical findings and practical use.

However, there were also limitations in the methodology. The scope of the literature review, while thorough, could have explored a broader range of methodologies and datasets to strengthen the research framework. The reliance on secondary datasets introduced uncertainties regarding labeling accuracy and data diversity, which may have impacted the validity of the results. Limited access to advanced computing resources, such as GPUs, restricted the ability to conduct more complex experiments or test advanced architecture. Additionally, due to time constraints, the insights gathered from expert interviews were not fully utilized, leaving opportunities for better integration of domain knowledge into the research process.

Alternative Research Methodology

Through this research, several challenges and limitations were identified, leading to insights into alternative methodologies that could have improved the outcomes. One significant issue was the reliance on Google Colab's free GPU, which limited the ability to train and test all five pre-trained architectures (Xception, MobileNet, ResNet50, VGG16, and InceptionV3) after applying data enhancement and augmentation. Instead, I tested only the two best-performing architectures, MobileNet and Xception, to evaluate the effects of these preprocessing techniques. While this approach provided some insights, it would have been more effective to test all five architectures systematically.

An alternative methodology could have involved:

Testing Enhancement and Augmentation on One Model First: Instead of applying enhancement and augmentation to all architectures initially, I could have tested these preprocessing steps on a single architecture, evaluated the results, and then extended the approach to other models. This would have saved time and computing resources.

Utilizing More Advanced Computing Resources: If access to a paid or more powerful GPU had been possible, I could have conducted more extensive experiments on all architectures and tested multiple configurations of preprocessing techniques.

This process revealed several strengths and weaknesses in the research methodology. One key strength was the realization that data enhancement alone, without augmentation, improved the performance of the MobileNet architecture. However, this approach also highlighted inefficiencies in how the models were tested. Focusing on all five architectures without first validating the preprocessing methods on one architecture led to a more time-consuming process than necessary, especially given the computing resource limitations.
Another important lesson was the need for efficient data handling. Initially, accessing the dataset from Google Drive caused significant delays during training, but moving the dataset to the Colab environment for direct access significantly reduced training time. This experience taught me the importance of optimizing the training pipeline to maximize available resources.

In future projects, I would prioritize testing preprocessing steps on a single model before scaling the approach to others. Additionally, investing in more powerful computational resources would enable more comprehensive experimentation and reduce the time constraints imposed by limited GPU availability. These lessons will guide future research and help streamline the process for better results and resource management.

Recommended Action and Future Considerations

Reflecting on the challenges faced and lessons learned during this research, several actionable steps can be recommended to improve similar studies in the future:

Based on the findings, applying image enhancement without augmentation proved to yield better results. Future studies should focus on fine-tuning the enhancement process and experiment with various learning rates and higher epochs to further improve performance. Due to GPU limitations, the epochs used in this study were relatively low. Increasing epochs with adequate resources could enhance model training.

The reliance on Google Colab's free GPU posed significant challenges in running multiple experiments efficiently. It is highly recommended to secure access to powerful GPU resources through cloud-based platforms like AWS, Azure, or paid Colab plans. This would allow researchers to train models faster and explore more extensive configurations.

The dataset used in this study was effective but not without limitations, particularly concerning labeling accuracy. Future research should prioritize acquiring more data from reliable sources to improve model training and validation. A larger, well-labeled dataset would contribute to better generalization and more robust results.

Segmentation can be a valuable addition to future research. Designing models capable of detecting tumors and highlighting their exact locations in the images would significantly improve the practical applications of brain tumor classification in the medical field.

Incorporating 3D MRI data into future models would provide a more detailed and comprehensive understanding of brain tumors. This could allow the model to utilize spatial information more effectively, potentially improving accuracy and usability in real-world medical scenarios.

Recommended Methodology

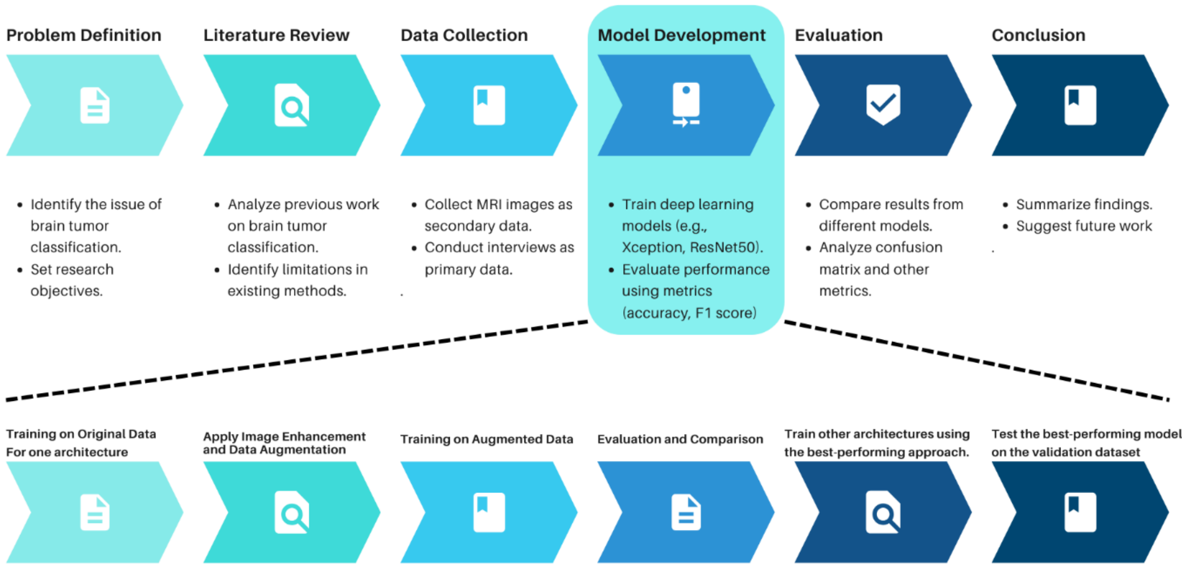

Figure 10 - Updated version of paper Methodology

The updated flowchart reflects a more structured and efficient research methodology. It begins with identifying the problem of brain tumor classification and conducting a literature review to understand prior work and research gaps. The data collection step incorporates both primary data from expert interviews and secondary data from existing MRI datasets, ensuring a comprehensive foundation.

In the Model Development section, the process is refined to train one architecture on the original dataset first, followed by applying image enhancement and data augmentation to evaluate their impact. This step ensures the best practices are identified before expanding to other architectures like MobileNet and Xception. The results are then evaluated using various metrics, such as accuracy and F1 score, and compared to determine the most effective approach.

Finally, the Evaluation and Comparison stage leads to selecting the best-performing model, which is tested on a validation dataset for robustness. The conclusion provides a summary of findings and recommendations for future work, including exploring 3D MRI data and segmentation methods. This updated approach ensures a clear and systematic process for achieving reliable results.

References

[1]A. Melnikovas, "Adapting Research Onion Model for Futures Studies," https: //www.semanticscholar.org/paper/Towards-an-Explicit-Research-Methodology-% 3A-Adapting-Melnikovas/dccf54a5a4312ceb2261e1989cda01f73989d735, 2018, [Online; accessed 10-Jan-2024].

[1] Stryker, C. and Kavlakoglu, E. (2024). What is Artificial Intelligence (AI)? [online] IBM. Available at: https://www.ibm.com/topics/artificial-intelligence.

[2] IBM (2020). *Artificial Intelligence in Medicine | IBM*. [online] www.ibm.com. Available at: https://www.ibm.com/topics/artificial-intelligence-medicine

[3] Abd-Ellah, M.K., Awad, A.I., Khalaf, A.A.M. and Hamed, H.F.A. (2018). Two-phase multi-model automatic brain tumor diagnosis system from magnetic resonance images using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2018(1). doi:https://doi.org/10.1186/s13640-018-0332-4.

[4] Kumar, S. and Mankame, D.P. (2020). Optimization driven Deep Convolution Neural Network for brain tumor classification. *Biocybernetics and Biomedical Engineering*, 40(3), pp.1190–1204. doi:https://doi.org/10.1016/j.bbe.2020.05.009.

[5] Sharif, M.I., Li, J.P., Khan, M.A. and Saleem, M.A. (2020). Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. *Pattern Recognition Letters*, 129, pp.181–189. doi:https://doi.org/10.1016/j.patrec.2019.11.019.

[6] Aamir, M., Rahman, Z., Dayo, Z.A., Abro, W.A., Uddin, M.I., Khan, I., Imran, A.S., Ali, Z., Ishfaq, M., Guan, Y. and Hu, Z. (2022). A deep learning approach for brain tumor classification using MRI images. *Computers and Electrical Engineering*, 101, p.108105. doi: https://doi.org/10.1016/j.compeleceng.2022.108105

[7] Aurna, N.F., Yousuf, M.A., Taher, K.A., Azad, A.K.M. and Moni, M.A. (2022). A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Computers in Biology and Medicine*, 146, p.105539. doi:https://doi.org/10.1016/j.compbiomed.2022.105539.

[8] Muhammed ÇELİK and Özkan İnik (2024). Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification. Expert Systems with Applications, 238, pp.122159–122159. doi:https://doi.org/10.1016/j.eswa.2023.122159.

[9] Saleh, A., Sukaik, R. and Abu-Naser, S.S. (2020). *Brain Tumor Classification Using Deep Learning*. [online] IEEE Xplore. doi:https://doi.org/10.1109/iCareTech49914.2020.00032.

[10] www.kaggle.com. (n.d.). *Brain Tumor Classification (MRI)*. [online] Available at: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

[11] Dikande, M., Aurelle Tchagna Kouanou, Monthe, V., Michael Kameni Nana and Bertrand Moffo Lonla (2024). A deep learning method for brain tumor classification using MRI data. *Informatics in medicine unlocked*, 44, pp.101423–101423. doi:https://doi.org/10.1016/j.imu.2023.101423.

[12] Nickparvar, M. (2021). *Brain Tumor MRI Dataset*. [online] www.kaggle.com. Available at: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

[13] Alvarado, D. (2024). Electromyography of the biceps femoris during the gait cycle Dataset. *IEEE DataPort*. [online] doi: https://doi.org/10.17616/R3H77H

[14] Saunders, M., Lewis, P., & Thornhill, A. (2007). Research Methods for Business Students, (6th ed.) London: Pearson.