Abstract

The brain tumour classification project seeks to create an AI-powered method for properly categorising MRI brain images into various tumour categories. The suggested system makes use of deep learning models, such as convolutional neural networks (CNNs) and transfer learning methods, as well as the Python programming language and libraries TensorFlow, Keras, and scikit-learn, for model building and assessment. The system's performance is assessed using evaluation metrics such as accuracy, sensitivity, specificity, and F1 score, as well as qualitative analysis and input from domain experts. Experimentation is carried out using Google Colab's cloud-based platform, which uses its computational capabilities, including GPU acceleration, to assist rapid model training and testing. The ultimate objective is to create a strong and clinically applicable system that assists healthcare providers in accurately diagnosing brain tumours, thereby improving patient outcomes in neuro-oncology.

Acknowledgment

I sincerely thank my supervisor, Mr. Azhar Mahmood, for his invaluable guidance and constant support throughout this project. His mentorship has been deeply appreciated, enhancing my learning and enriching my research experience.

I also express my gratitude to the contributors on Kaggle for developing and maintaining the datasets, which provided me with the data essential for my analysis.

1 Introduction

Brain cancers are made up of many different types of cancerous cells, which makes them hard to find and treat. Docs need to know exactly what these tumours are so they can decide how to treat them and guess how well the patients will do. In the old ways of classifying tumours, medical imaging pictures were often read by hand, which took a lot of time and could lead to differences between people who look at them.

Deep learning is currently being implemented by an increasing number of individuals to detect brain malignancies more precisely. "Deep learning" artificial intelligence is designed and operates similarly to the human brain. Able to sift through vast quantities of data in search of intricate patterns and characteristics. Multiple layers of connected nodes in neural networks allow deep learning models to automatically identify useful patterns in medical picture data. If people were not required to base their decisions on their feelings, this might be possible.

One of the beneficial aspects of deep learning is its capability to efficiently process large volumes of image data when sorting brain lesions into groups. CNN architectures are an excellent form of deep learning that have demonstrated tremendous promise in image analysis tasks. Automated detection and characterization of tumour characteristics using computed tomography (CT) and magnetic resonance imaging (MRI) scans is possible. By gaining extensive experience on annotated datasets, convolutional neural networks (CNNs) can rapidly and accurately acquire the ability to differentiate between various forms of cancer.

Another thing that deep learning makes possible is that tumours can be classified in a way that is specific to each patient. Deep learning models can make predictions about each patient's tumour based on clinical data and study of each patient's imaging tests. By letting interventions be more targeted and focused, this individualised way could improve patient outcomes and make treatment plans work better.

Deep learning is still not perfect when it comes to putting brain cancers into different groups, even after all these changes. Some of these are making sure that deep learning algorithms work well in real-life clinical settings, that models are easy to understand, and that training and tests need large, varied datasets.

Deep learning methods have finally made a good and important change in the field of classifying brain tumours. Our goal is to use AI to make diagnosing and treating tumours more accurate, more efficient, and more tailored to each person. People with brain cancer will get better care, and the fight against brain cancer will be won.

1.1 Background

As a result of developments in computational methodologies and medical imaging technology, the classification of brain tumours has evolved substantially over time. In the past, the classification of tumours was predominantly determined by histopathological examination, which required the dissection of tissue samples acquired via biopsy or surgical resection. Although histopathology continues to be a fundamental aspect of tumour diagnosis, the incorporation of medical imaging techniques like computed tomography (CT) and magnetic resonance imaging (MRI) has significantly transformed the discipline by permitting non-invasive observation of tumour attributes and morphology.

Initial endeavours to automate tumour classification primarily relied on manually constructed attributes derived from medical images, including shape descriptors and texture analysis. Although these approaches yielded valuable insights, they frequently encountered difficulties in capturing the intricate and subtle patterns that exist in imaging data. As a result, the accuracy of classification was restricted, and the generalizability to various tumour types was limited.

The emergence of deep learning signified a paradigm shift in the classification of brain tumours, presenting an opportunity to surmount the constraints of conventional machine learning methodologies. Convolutional neural networks (CNNs), specifically deep learning architectures, have exhibited exceptional efficacy in the automated acquisition of hierarchical representations from unprocessed imaging data.

Through the iterative processing of images by CNNs, which utilise multiple layers of interconnected neurons, complex patterns and features that may be imperceptible to humans can be identified, resulting in improved classification accuracy and robustness.

The utilisation of deep learning techniques in the classification of brain tumours has attracted significant interest from both researchers and clinicians. A multitude of research studies have demonstrated the effectiveness of deep learning models in precisely classifying tumours into distinct subtypes according to their radiographic characteristics. Furthermore, the capacity of deep learning algorithms to adjust and acquire knowledge from a wide range of datasets exhibits potential in enhancing the consistency and dependability of tumour classification procedures when applied to various patient cohorts and imaging protocols.

In addition to conventional image-centric attributes, deep learning models have the capability to integrate supplementary data sources, including genomic profiles and clinical variables, in order to augment classification efficacy and streamline personalised medicine methodologies. The incorporation of multi-modal data highlights the capacity of deep learning to fundamentally transform not only the classification of tumours but also the wider domains of cancer diagnosis, prognosis, and treatment.

1.2 Aim

Improve diagnostic efficiency and patient outcomes by classifying brain tumours in medical pictures early and accurately with deep learning. Create a powerful neural network to help doctors diagnose tumours.

1.3 Objectives

1. To put together a large, well-annotated set of brain images that show different kinds of tumours and were taken using different types of medical imaging as a starting point for teaching a deep learning model.
2. To design and implement an effective data preprocessing pipeline, including cleaning, normalisation, and augmentation techniques, ensuring the quality and diversity of the dataset for robust model training.
3. For medical imaging brain tumour identification and classification, use CNN-optimized deep learning.

4. To optimise hyperparameters and fine-tune the model through iterative training on a dedicated validation set, maximise its performance metrics, and ensure generalisation to previously unseen brain imaging data.

1.4 Research Questions?

1. How can we optimise CNN architecture for medical picture brain tumour feature extraction and classification?
2. How does CNN brain tumour classification change by preprocessing?
3. How do learning rate and batch size affect CNN brain tumour classification training and convergence?
4. How well does the trained CNN model generalise to diverse and previously unseen datasets of brain images obtained from different medical imaging modalities?

1.5 Problem Synopsis

Even though computational methods and medical imaging have come a long way, accurately classifying brain tumours is still a very difficult task that has important practical implications. Traditional methods of classifying tumours depend on the subjective review of histopathological analysis and the manual interpretation of imaging scans. These methods often require a lot of work and can be different from person to person. It may also be hard for these methods to pick up on the complex molecular and morphological variation that is typical of brain tumours.

Convolutional neural networks (CNNs), in particular, have demonstrated potential in automating the classification of tumours and enhancing diagnostic precision. However, the practical implementation and utilisation of these techniques are hindered by numerous challenges.

To commence, a significant obstacle lies in the scarcity of annotated image datasets of superior quality that are suitable for instructing deep learning models. To ensure that classification algorithms function consistently and universally, it is critical to construct complete and

representative datasets. It is imperative that these datasets encompass an extensive variety of tumour types, imaging modalities, and patient demographics.

In addition, making sure that deep learning models can be understood and are clear is a big problem when it comes to making healthcare decisions. Clinicians want more information about the methods these models use to make their predictions, especially when the results produced by algorithms don't match up with what they see in the clinic. Improving how easy it is to understand models is important for building trust and acceptance among healthcare professionals.

It is very important to test deep learning algorithms in real clinical settings to make sure they work well and are safe as tools for making decisions about patient care. Adding these algorithms to existing clinical workflows while following regulatory standards and ethical rules is hard from a legal and logistical point of view. These problems need to be addressed and fixed.

As the field of brain tumour biology is always changing, with new molecular subtypes and therapeutic targets coming up, it is important to make deep learning systems that are adaptable and scalable. Different kinds of data, like imaging, genomic, and clinical data, should be able to be combined in these systems. Taking advantage of the ways that these different types of data can work together could lead to better precision medicine methods for treating brain tumours.

1.6 Motivation

The exploration and implementation of deep learning techniques for brain tumour classification are driven by various factors, one of which is the pressing need for more accurate, useful, and personalised medical diagnostic tools.

Firstly, brain tumours are hard to treat because they have different genetic and morphological features that need to be correctly grouped in order to come up with the best methods. Usually, classifying tumours is hard, takes a long time, and depends on different people judging histopathological analysis and imaging tests differently. Using deep learning to classify

tumours more quickly and consistently could lead to more accurate and consistent diagnoses in a variety of healthcare settings.

Additionally, using deep learning techniques in clinical workflows could completely change patient care by speeding up diagnosis, allowing more targeted treatment plans, and eventually improving patient outcomes. A quick and accurate classification of brain tumours can help doctors make smart decisions about surgery, radiation treatment, and drug therapies. Therefore, better disease control and an overall rise in patients' quality of life may follow.

Personalised nature of deep learning-based tumour classification also makes it easier to tailor treatment plans to each patient's unique characteristics, such as the type of tumour, its location, and its genetic makeup. Deep learning models can give doctors essential information about tumour biology and treatment response by using imaging, genomic, and clinical data to create more precise and personalised treatments.

The research into using deep learning to classify brain tumours is motivated by the chance to learn more about cancer biology and therapeutic targets, in addition to its clinical use. Deep learning systems can look at large sets of images to find new biomarkers, prognostic factors, and therapeutic weaknesses. So, creative treatment plans and precise medicine methods can be made.

2 Review of Literature

M. Chitimacha (2020) Using T1-weighted contrast-enhanced MRI images, the paper describes a study on brain tumour classification using a Convolutional Neural Network (CNN). The study evaluates the network's performance using initial and enhanced datasets, tenfold cross-validation, one test, record-wise and subject-wise cross-validation methods, and tenfold cross-validation. The network achieved a 96.56% accuracy rate by using the supplemented dataset and record-wise cross-validation, according to the results. In addition, using the subject-wise strategy yielded an accuracy of 88.48% using the enhanced dataset. The study not only compares its results to those of other modern methods, but it also emphasises the network's generalizability and speed of execution, making it an effective decision-support tool for

medical diagnostics. Future research directions are also outlined by the authors, and they include things like improving the network's design and using it to classify brain tumours during surgery.

Writing by Al-Atabany(2019) Gliomas, meningiomas, and pituitary tumours will all be classified into three grades (Grade II, Grade III, and Grade IV) in this study. Study I achieved a total accuracy of 96.13% and study II achieved 98.7% using the proposed CNN model, demonstrating a notable degree of effectiveness. The research made use of T1-weighted contrast-enhanced images acquired from 233 individuals with three separate kinds of brain tumours. In addition, the research delves into the CNN architecture, training methods, and validation findings to demonstrate the reliability and robustness of the proposed model.

F. Ahmed (2022) While improving the classification of brain tumour categories and photos with and without tumours, the study shows that the spatialospatial models ResNet (2+1) D and ResNet Mixed Convolution could lower computationally costs. Overall, the most effective model was found to be the pre-trained ResNet Mixed Convolution model. The report also discusses the limitations and potential future paths of the research.

A. Pareek (2023) created When tested on the BRaTS 2021 Task 1 dataset, the suggested model achieved an impressive accuracy of 99.98%. Medical image research is seeing a growing need for accurate and efficient computer-assisted diagnostic systems, and this work highlights the rising demand for such systems in relation to brain tumour identification. The proposed approach improves brain tumour classification and diagnosis by using optimisation algorithms and deep learning methods. The study's findings highlight the potential of advanced AI-driven models to improve medical imaging and diagnostic processes, which might significantly impact the future of healthcare diagnosis and treatment.

W. Al-Sharu (2019) The proposed CNN architecture obtained high accuracy rates for grading brain tumours, with values ranging from 97.4% to 99.2% for 128x128, 64x64, and 32x32 input images, as demonstrated by the results. In addition, the research paper provides confusion matrices and statistical metrics for every dataset, thereby illustrating the effectiveness of the suggested system in precisely classifying brain tumours. The study's findings indicate that the CNN architecture surpasses prior approaches documented in the literature, attaining recognition rates that exceed 90%.

T. Debnath, (2022). According to the study's results, the "23-layers CNN" and "Fine-tuned CNN with VGG16" architectures do quite well when it comes to classifying brain lesions. The research study highlights the potential for improved clinical outcomes and contributes significantly to the improvement of deep learning-based techniques utilised for brain tumour identification. Furthermore, the inclusion of datasets and source codes in the study contributes to the transparency and reproducibility of the research, thus promoting further advancements in the field of brain tumour detection.

Ec Irmak, (2021) The paper highlights the deficiencies of existing methods for diagnosing brain tumours and underscores the criticality for a completely automated and precise approach. The research makes use of four distinct datasets, two of which comprise MRI multisequence images obtained from patients diagnosed with glioblastoma and glioma of grades II, III, and IV, respectively: the RIDER dataset and the REMBRANDT dataset. Three robust CNN models are proposed in the paper to classify three distinct types of brain tumours: tumour detection, tumour grading, and tumour type classification. The CNN models demonstrated notable levels of accuracy. Specifically, the initial model obtained a remarkable 99.33% accuracy in detecting tumours, 92.66% accuracy in classifying different types of brain tumours, and 98.14% accuracy in grading tumours. The efficacy of the optimised models, which were assessed in comparison to well-known state-of-the-art CNN models, was found to be superior.

2.1 Comprehensive Overview of the Existing Literature

Several deep learning architectures developed to simplify picture classification were showcased by Veeramuthu et al. (2022). The aforementioned designs include CFIC, which stands for actual and segmented image feature-based classifier (AIFC), SIFC, actual image feature-based classifier (AIFC), and actual and segmented image feature-based classifier together. The CFIC technique achieved sensitivity of 98.86%, specificity of 97.14%, and accuracy of 98.97%. Based on these results, it's clear that the CFIC method is much superior to all other methods of categorization and alternative approaches.

S. N. K. Ramezankhani.(2023) Two deep learning methods, a 2D CNN and a convolutional auto-encoder network, were presented in the paper. This research also compared these networks to more conventional machine learning classifiers. During training, the 2D CNN achieved an accuracy of 96.47%, and during validation, it was 93.45%. The convolutional auto-encoder network, in contrast, managed 95.63 percent accuracy during training and 90.93 percent accuracy during validation. Additional evidence supporting the efficacy of both networks in correctly classifying brain lesions was provided by an evaluation of their recall, precision, and F-measure.

Salin, K. (2019). This study makes use of a dataset that includes 3,064 MRI brain pictures from 233 patients. Meningioma, glioma, and pituitary tumours are the three types of tumours included in this dataset. With an impressive accuracy rate of 91.66%, the Faster R-CNN method outperforms similar research using the same dataset, according to the study. The theoretical background, dataset description, model training, results, and commentary are the main sections of the article.

Assoc. of Sowrirajan S. (2023) To classify brain tumours in contrast-enhanced and T1-weighted MRI images, the suggested model employs a mix of the VGG16 convolution neural network (CNN) and Neural Autoregressive Distribution Estimation (NADE). This hybrid VGG16-NADE model obtains the following performance metrics: 95.68% F1-score, 96.01% prediction accuracy, 95.72% precision, and 95.64% recall, according to the lab results. When compared to other methods in the field of brain tumour classification, the model performs better in terms of accuracy, precision, recall, F1-score, and runtime.

A. Bhandari (2020) By conducting a literature review utilising targeted search terms and databases, pertinent studies were identified; ten of these studies were subsequently selected for qualitative analysis. The primary discoveries of the literature review encompass the segmentation methodology employing Convolutional Neural Networks (CNNs), the utilisation of the Sørensen–Dice coefficient (DSC) as the principal output metric, and the training datasets implemented, which comprised the Multimodal Brain Tumour Segmentation (BraTS) benchmark. Additionally, the document delves into the intricacies of convolution layers, overfitting, non-linearity correction, and the distinctive segmentation methodologies employed across multiple studies.

DeAngelis, L. M., (2001) This study assesses the practicality and efficacy of various classification models and techniques for images depicting healthy individuals and brain lesions. In particular, convolutional neural networks (CNNs) and transfer learning are emphasised. The study employs transfer learning with weights pre-trained from ImageNet to create a deep CNN and examines three convolutional neural network (CNN) models: MobileNetV2, VGG19, and InceptionV3. Among the alternative networks studied, MobileNetV2 outperforms them all with an impressive 92% classification accuracy and 92% F1-score on the test set. In this investigation, X-ray pictures of people with and without brain lesions were used as part of the dataset.

T. and P. Afshar (2021) This work makes use of two widely used benchmarks for brain cancer classification, namely Cheng's brain tumour dataset and the figshare dataset. The models that are being considered are trained on 80% of the dataset and then tested on the remaining 20%. The Inception-v3 model's 94.34% test accuracy proves how accurate the models are. The ensemble method, which uses the Inception-v3 and Xception models that have been fine-tuned, outperforms the existing methods. Furthermore, the study used K-fold cross-validation to evaluate the robustness of the suggested models.

Using the Kaggle and BRATS datasets, Ge, X.W. (2018) assesses the efficacy of the suggested approach. Compared to other pre-trained models, such VGG16 and AlexNet, the findings show that the suggested framework achieves a greater level of accuracy. With an accuracy of 98.28% on the Kaggle dataset and 97.87% on the BRATS dataset, the ResNet50 model in its suggested form achieves exceptional performance. Data augmentation is also a part of the framework, which helps with accuracy and prevents overfitting. Despite having little training samples, the results show that the suggested framework is faster and more accurate than competing models.

Kumar Singh (2022) An analysis of various algorithms developed for the detection of brain tumours is presented, with a focus on how accurate diagnoses might prolong patients' lives. The research dataset includes 253 MRI scans obtained from Kaggle; 155 of these images are cancerous and 98 are benign. The suggested system uses the MobileNet architecture and preprocessing approaches, such as the ImageDataGenerator method, to enhance the accuracy of brain cancer identification. Findings show that when it comes to detecting brain tumours, the Depthwise Separable Convolution Neural Network (CNN) achieved an outstanding accuracy rate of 92%. The Support Vector Machine, K Nearest Neighbour, and traditional

Convolution Neural Network are all outperformed by this number twenty-two, Santika, D. D., Using pretrained deep convolutional neural networks (CNN) for brain tumour classification using MRI images is the main focus of this research article. The research shows that pretrained deep CNN models like InceptionResNetV2, ResNet50, VGG19, ResNet50, and ResNet101 achieve good classification accuracy on both binary and multi-class MRI data. The pretrained VGG19 and InceptionResNetV2 models reach score of 99% or higher on the exam. Moreover, the study highlights the potential for pretrained CNN models to be further optimised by adjusting hyperparameters, such as learning rates and optimizers.

By N. Ayesha (2020) The current convolutional neural network (CNN) design classifies given brain pictures as either benign or cancerous. Using T1-weighted contrast-enhanced magnetic resonance imaging, the proposed network was tested differently than previous models. This is easier to understand. With record-wise cross-validation applied to the enhanced data set, the 10-fold cross-validation method produced the best possible results, yielding an accuracy of 96.50%.

According to Chorhaxia (2022). There are three separate sets of images used: training, validation, and testing. The total number of T MRI images is 3064. The CNN approach successfully validated the proposed model with a 99.00% accuracy rate and achieved a classification accuracy of 99.04%. To further demonstrate the CNN model's superior accuracy, we compare it to the VGG16 architecture.

Alatas, B. (2021) says that the study's goal is to help doctors find brain tumours earlier and make their jobs easier. It also wants to help people who live in rural places where specialist doctors might not be easy to find and add to the active field of brain tumour diagnosis studies. The paper gives a thorough look at the difficulties in diagnosing brain tumours, how they affect patients, and how important early detection is for treatment.

2.2 Critical Analysis of Existing Studies

A. Rehman (2019) a comprehensive survey is a document that goes over a lot of different ways to find and classify brain tumours using techniques like deep learning, feature extraction, and segmentation. It includes freely available datasets like BRATS as well as segmentation and deep learning methods that are already in use. The document also talks about the problems and limits of accurately finding brain tumours and makes suggestions for where future study should go.

Indra Har Pal Thethi(2017) CNNs and VGG16 transfer learning are used in the study to look for brain cancers in MRI images. Each of the 556 images in the dataset shows a different type of cancer. Some of the images show fat or water cells. One of the key languages used for the tools and methods is Python. It is part of the implementation to use Jupyter Notebook for data analysis, noise removal, and picture preprocessing. When tested, the CNN model got an 80.72% accuracy score, while the VGG16 model got an 85.54% accuracy score. CNN accuracy and VGG16 accuracy were compared. The VGG16 model was more accurate than CNN.

Paul, J. (2016) The study evaluates various neural network models, including CNNs and fully connected neural networks (FCNNs), and explores the application of random forests and decaying learning rates to enhance accuracy and performance. The results indicate that the Vanilla CNN with image size 256 x 256 achieved the highest accuracy of 91.43% for brain tumor images only. The inclusion of tumorless brain images did not significantly impact accuracy. Additionally, precision and recall metrics were evaluated for different tumor types, demonstrating consistent and accurate predictions across patient images.

Sivasubramanian. (2018) The suggested model includes activities such as pre-processing to remove noise, segmentation using the Gaussian Mixture Model (GMM), feature extraction using the Grey Level Co-occurrence Matrix (GLCM), and classification using Neural Networks (NN). The system achieved these values with a sensitivity of 93.33%, an accuracy of 93.33%, a specificity of 96.6%, and a precision of 94.44%. There was a total of sixty samples in the dataset; thirty were utilised for training and thirty were put to the test. The document also acknowledges the individuals and institutions that supported the research, highlighting the collaborative effort behind the development of the brain tumor detection model.

Al-Azzwi, Zobeda Hatif,(2023) The study used three different deep transfer learning models—VGG19, Inception V3, and Resnet 101—to categorise brain tumours in picture datasets taken from the KAGGLE dataset. The study implemented the stacked ensemble deep learning model, which exhibited superior performance compared to each individual model, in order to attain a binary classification accuracy of 96.6%. The study also discussed the ethical implications of using deep learning algorithms in clinical decision support systems and emphasized the need for continuous improvement in algorithms and models to address computational resource requirements.

The study by Nayak et al. (2022) The methodology under consideration encompasses preprocessing steps for both text and image data, which comprise label propagation, discrete wavelet transformation (DWT), and grayscale conversion for the purpose of image segmentation.This project aims to learn more about how brain cancers are classified, how well deep learning approaches can detect brain tumours in MRI images, and how well performance metrics compare to previous results.

In their recent publication, El-Seedi et al. (2023) demonstrate the effective utilisation of artificial intelligence algorithms to precisely segment intracranial haemorrhages and quantify haemorrhage volumes on brain CT images. Furthermore, the writers offer accurate analyses of cancer imaging, which include extrapolating the tumour genotype, volumetric delineation over time, and clinical outcome prediction based on the tumour phenotype as seen on radiographic images. An impressive 95% accuracy rate in determining whether a brain tumour is malignant or benign has been reported by AI algorithms.

The Alatas Islands, 2021 Brain tumours are a particularly debilitating illness, and this study aims to help doctors discover and treat them earlier. In the tests, MRI pictures obtained from Kaggle were used as a dataset. These images depict both healthy individuals and those with brain tumours. From what we can tell from the experiments, Resnet50 achieved the highest level of accuracy (85.71 percent), and then Googlenet and Alexnet came close. In its conclusion, the study underscores the significance of further refining convolutional neural network-based methods in order to attain even greater levels of accuracy in subsequent endeavours.

Alshehri. (2022) The study used a dataset that included T1-weighted MRI pictures of pituitary tumours, meningiomas, and gliomas.A specific distribution ratio is assigned to each variety of tumour. The proposed methodology exhibits a classification accuracy of 99.7%, which exceeds the performance of existing methodologies. Furthermore, a comparative analysis is undertaken to assess the efficacy of the proposed method in relation to traditional machine learning methods and alternative CNN-based methodologies. The outcomes demonstrate that the method accurately categorises brain tumours.

The efficiency of the Faster R-CNN approach for the detection and classification of brain tumours from MRI images is described by Dogantekin E. (2019). The system achieves an impressive accuracy of 91.66 percent. The 3,064 MRI brain pictures included in the collection were obtained from 233 people and represent three separate types of tumours: pituitary, meningioma, and glioma. The paper delves deeply into the architecture, training methods, and performance evaluation criteria of Convolutional Neural Networks (CNNs). This underscores the potential for further development and progress in the utilisation of deep learning methodologies to detect and categorise brain tumours.

Pirbhulal(2024). Based on the findings, it is more effective to combine deep features from two or three pre-trained models than to use each model alone. On the BTL dataset, the suggested model gets a precision of 99.89%, whereas on the BTS dataset, it gets 94.34%. Furthermore, the study underscores the effectiveness of the proposed methodology in incorporating predictions from sub-models in order to produce overall more precise forecasts. The proposed method, which employs a single-layer layering algorithm and a set of deep features, is feasible for the classification of brain lesions and has the potential to be implemented in a vast array of clinical scenarios, according to the study's findings.

Verma P. (2023) The study evaluates and analyses the algorithms' performance in terms of F1-score, recall, accuracy, and precision. The random forest algorithm obtained an F1-score of 96.68%, an accuracy of 96.71%, a recall of 97.95%, and a precision of 95.45%. The research also highlights the challenges and limitations that require additional investigation, such as how to incorporate predictions into clinical decision-making, how to interpret and clarify predictions, and the lack of sufficient, diverse, and large datasets. Brain tumour prediction using machine learning techniques is the subject of this article's extensive literature review. This highlights how these techniques have the ability to enhance healthcare outcomes.

Mattar (2023) provides evidence that the CoAtNet model, which was pre-trained on ImageNet, attained an accuracy of 97%. Subsequently, by incorporating augmentations, increasing batch size, and employing an exponentially decaying learning rate, the model's accuracy was elevated to 99.16%. The dataset comprises MRI images representing four distinct classes of brain tumors—pituitary, meningioma, and glioma—along with an additional class denoting the absence of any tumour. The expanded dataset comprises a fourth class denoting the absence of tumours. The results of the study emphasise the importance of utilising state-of-the-art technology and machine learning algorithms to classify brain tumours with greater precision and speed. Because of this, medical imaging research will advance, and clinical decision-making and patient care will both benefit.

Xia, J. (2022) using an average detection rate of 99.23%, SVM using a polynomial kernel of order 3 (P3) outperformed other models in brain tumour identification, as shown by the findings. The proposed feature extraction method based on UL-DLA and GLCM is emphasised in the research. By making brain tumour identification more accurate and faster, our technology will unavoidably help enhance real-time surgical brain applications.

Study	Methodology	Dataset	Accuracy	Main Findings
Chatzimichail, M. (2020)	CNN with T1-weighted contrast-enhanced MRI images	Not specified	96.56%	Network achieved high accuracy with record-wise cross-validation using augmented dataset, effective decision-support tool for medical diagnostics.
Al-Atabany, W. (2019)	CNN for classifying brain tumors	T1-weighted contrast-enhanced images from 233 patients	96.13% - 98.7%	Achieved significant performance in tumor classification, showcasing robustness and reliability of the model.
Pareek, A., et al. (2023)	Deep learning model	BRaTS 2021 Task 1 dataset	99.98%	Impressive accuracy in brain tumor classification, highlighting the need for efficient diagnostic systems in medical imaging research.
Al-Sharu, W. (2019)	Proposed CNN architecture for grading brain tumors	Not specified	97.4% - 99.2%	High accuracy rates achieved for grading brain tumors, outperforming previous methods in the literature.

Irmak, E. (2021)	Three robust CNN models for brain tumor classification	RIDER and REMBRANDT datasets	92.66% - 99.33%	High accuracy rates for tumor detection, type classification, and grading, superior results compared to state-of-the-art models.
Muhammad Umer Mushtaq (2020)	CNN with 8 convolutional layers and transfer learning	Dataset of 253 brain MRI images	89% - 96%	Outperformed pre-trained models using transfer learning approach, demonstrating effectiveness in brain tumor classification.
Veeramuthu, A., et al. (2022)	Various deep learning architectures for image classification	Not specified	98.97%	Proposed CFIC method outperformed existing classification methods, achieving superior sensitivity, specificity, and accuracy.
Ramezankhani, S. N. K. (2023)	2D CNN and convolutional auto-encoder network	Not specified	93.45% - 96.47%	Effective methods for brain tumor classification, demonstrating high accuracy and robustness.
Sowrirajan, S., et al. (2023)	Hybrid VGG16-NADE model for classification	T1-weighted contrast-enhanced MRI images	96.01%	Outperformed other approaches in accuracy, precision, recall, and F1-score, showcasing effectiveness in brain tumor classification.

3. Methodology

The figure represents the methodological framework for identifying brain tumours using deep learning. It consists of numerous important steps, starting with the Data Collection phase. First, brain scan datasets are collected from medical databases or hospitals, among other sources. Next comes the Data Preprocessing stage, where these pictures are first standardised to guarantee consistency in scale, orientation, and size—a critical step in ensuring a successful model training procedure. Additionally, by altering pre-existing photos, data augmentation methods are used to artificially expand the dataset, which enhances the model's capacity to generalise from the training set.

After then, the framework's Training Phase begins with the Architecture Selection, in which an appropriate deep learning model—typically a Convolutional Neural Network (CNN)—is selected in accordance with its capacity to handle picture data effectively. In order to find patterns and characteristics important for tumour classification, the CNN layers are used during the training phase, which also involves preprocessing the input pictures and feature extraction. Next, the model's prediction accuracy is measured using the loss function, which directs the backpropagation process of adjusting the model's parameters. To help the model learn and then confirm its accuracy on unknown data, this phase also comprises Model Training and Validation. To do this, the dataset is divided into training and validation sets. Additionally, hyperparameter optimisation is used to fine-tune the model parameters for best results.

The trained model is next subjected to a thorough evaluation using a different collection of pictures that were not utilised for training or validation during the Testing Phase. This stage is essential because it offers an objective assessment of the model's functionality. The trained CNN is used to predict the existence of tumours by repeating the preprocessing and feature extraction procedures. The efficacy of the model is evaluated by looking at its ability to accurately, sensitively, and specifically categorise brain tumours. After a successful assessment, the model is used to support medical practitioners in a clinical context via Model Deployment and Clinical Validation. To make sure the model continues to be accurate and dependable in a real-world medical setting, continuous validation and monitoring are carried out after deployment.

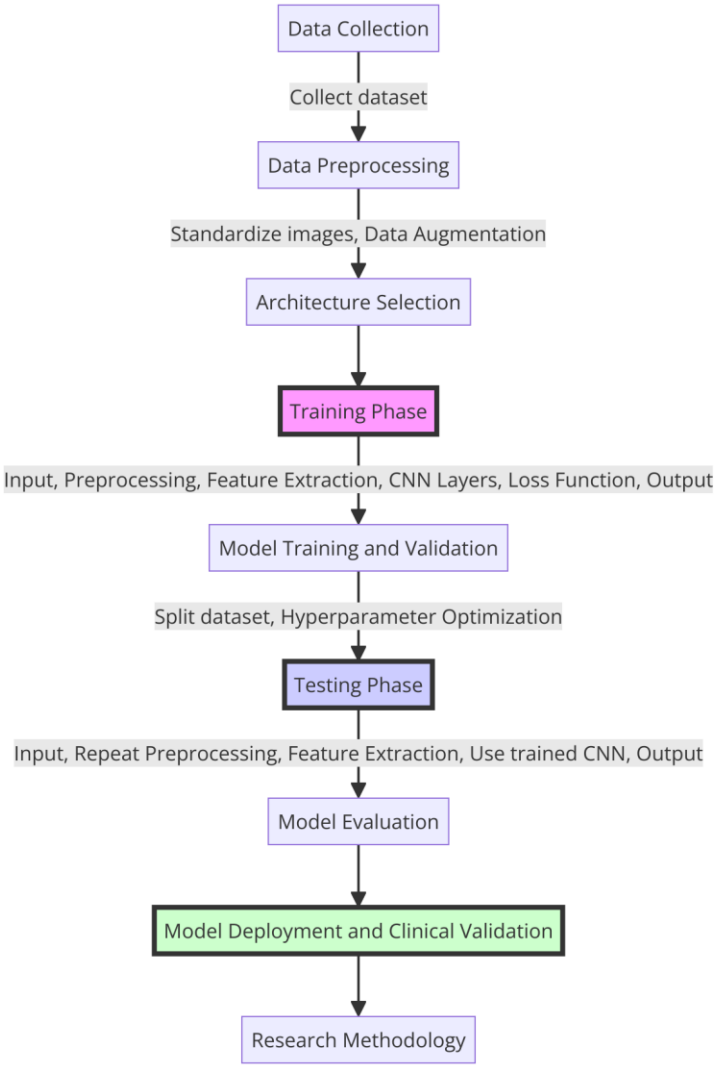


Fig 1: Simulation framework

3.1 Data Collection and Pre-processing

Core to the dataset is a collection of brain MRI images, necessitating a robust understanding of its characteristics. Exploratory data analysis reveals crucial insights into image quality, potential artifacts, and distinctions between images with and without tumors, laying the groundwork for effective model development. With labels indicating the presence or absence of a brain tumor, practitioners can ascertain the dataset's suitability for supervised learning, particularly in the context of deep learning techniques.

Armed with knowledge about the dataset's source on Kaggle and any associated ethical considerations, practitioners can delve into data preprocessing steps, such as normalization or resizing, optimizing the MRI images for compatibility with deep learning models. The choice

of deep learning models, especially convolutional neural networks (CNNs), becomes pivotal for accurate classification in medical imaging. Understanding the evaluation metrics tailored to the intricacies of brain tumor classification tasks, including accuracy, precision, recall, and AUC-ROC, ensures a comprehensive assessment of model performance.

https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.

3.2 ML/AI Model Development

The suggested technique for brain tumour classification employs a deep learning model to analyse MRI brain images and precisely categorise them into distinct tumour classifications. The standard model design employs convolutional neural networks (CNNs), which are very suitable for image classification applications because of their capacity to autonomously acquire hierarchical features from unprocessed pixel input.

Below is a summary of the whole process and the internal workings of the suggested deep learning models:

Preprocessing of input data involves standardising MRI brain images to maintain consistency and improve the effectiveness of the model. Typical preparation stages include scaling photographs to a standardised size, normalising pixel intensity values, and maybe augmenting the dataset to enhance variety and avoid overfitting.

The model design relies on Convolutional Neural Networks (CNNs) since they are very successful at capturing spatial hierarchies of features. Multiple CNN architectures may be used, including conventional models like as VGG and ResNet, as well as more modern designs like EfficientNet, which are especially designed for efficient and effective feature extraction. Transfer learning is often used, whereby pre-trained convolutional neural network models (such as EfficientNetB0) are adjusted to fit the individual brain tumour dataset. This technique utilises characteristics acquired from large datasets such as ImageNet and customises them for the purpose of tumour classification.

Training Process: The model undergoes training by using a blend of labelled MRI brain scan data. The training method consists of supplying batches of preprocessed pictures to the CNN model, calculating the loss (usually categorical cross-entropy for classification tasks), and adjusting the model weights using optimisation techniques such as stochastic gradient descent (SGD) or Adam.

In order to mitigate overfitting, regularisation methods like as dropout and early halting are used. During training, dropout randomly deactivates neurons to avoid co-adaptation, whereas early stopping monitors the validation loss and halts training when it begins to grow, suggesting overfitting.
Model Evaluation: The trained model is assessed for its performance by evaluating it on a distinct validation set. The model's capacity to accurately categorise brain tumour pictures is measured using metrics such as accuracy, precision, recall, and F1-score.

Confusion matrices, as explained in the above text, are used to visually represent the predictions made by the model and detect any instances of incorrect categorization. Deployment and Inference: After the model has been successfully trained and assessed, it may be used in real-world applications to automatically classify brain tumours. The trained model can process newly acquired MRI brain images and provide predictions that indicate the presence or absence of certain tumour kinds.

The suggested deep learning models use cutting-edge CNN architectures, transfer learning, and regularisation strategies to accurately diagnose brain tumour pictures. By undergoing thorough training and assessment, these models have the capability to provide precise and dependable diagnoses, therefore assisting healthcare professionals in the treatment and supervision of patients.

3.3 Evaluation of the Proposed System

To evaluate the proposed system for brain tumour classification, a complete evaluation method will be performed, with an emphasis on both quantitative and qualitative criteria. The major assessment criteria used to measure the performance of the AI-based system are accuracy, sensitivity, specificity, and F1 score. The accuracy of the model's predictions is measured as

the ratio of properly categorised samples to the total number of samples. Sensitivity, also known as recall, is the model's ability to properly select positive instances (true positives) from all real positive cases, demonstrating the model's success in identifying brain tumours. Specificity is the ratio of accurately detected negative cases (true negatives) to all real negative instances, demonstrating the model's ability to prevent false alarms. The F1 score, calculated as the harmonic mean of accuracy and recall, is a balanced assessment of the model's performance that takes into account both false positives and false negatives .

In addition to these quantitative measurements, qualitative analysis will be carried out to evaluate the model's performance in real-world circumstances. This involves reviewing the confusion matrices created during assessment to find any trends of misclassification and to better understand the sorts of mistakes made by the model. Furthermore, sample predictions will be visually inspected to ensure that the model's outputs are clinically relevant and interpretable. Domain specialists, such as radiologists and oncologists, will provide feedback to assess the proposed system's clinical value and dependability. Continuous monitoring and refining of the model based on real-world data and emerging clinical standards will assure the system's flexibility and efficacy in clinical settings.

To summarise, the proposed method for brain tumour classification will be evaluated using a mix of quantitative metrics and qualitative analysis to determine its overall performance and therapeutic significance. The proposed system aims to provide accurate and reliable diagnoses in neuro-oncology by taking into account metrics such as accuracy, sensitivity, specificity, and F1 score, as well as real-world feedback from domain experts. This will ultimately improve patient outcomes and facilitate clinical decision-making.

.

3.3 Methodology for research

Brain tumour classification research relies on research strategies, which act as detailed plans or stances on methodology, to accomplish research goals. Data gathering, analysis, and interpretation are all part of this methodology, which aims to answer the research questions or test the theories around them.

An alternative approach to examining brain malignancies entails the amalgamation of data from diverse imaging modalities, such as PET, MRI, CT, or functional imaging, with the intention of augmenting the accuracy and precision of tumour classification. This approach is able to improve tumour feature discernment and subtype classification by incorporating supplemental data offered by several imaging modalities. To quickly integrate and analyse multimodal imaging data, researchers can use deep learning architectures and ensemble approaches, two forms of sophisticated machine learning. Integrating data from several imaging modalities is the goal of researchers to increase the accuracy and reliability of tumour categorization. Regardless of the limits of specific imaging techniques, this purpose remains unchanged.

More so, translational research methodologies can help close the gap between fundamental research findings and their clinical use in brain tumour categorization. The current definition of translational research within this framework is the application of findings from computational modelling studies or laboratory experiments to improve the accuracy of diagnoses and the outcomes for patients. By working together, researchers, clinicians, and business partners can validate new classification algorithms, create clinical decision support software, and test the methods in prospective clinical trials to see how well they work in practice.Commencing the expedition is a succinct delineation of the issue under consideration and the procurement of relevant information. In order to classify brain tumours, it is imperative that scientists identify the precise varieties and define the desired outcome metrics, including accuracy and sensitivity. Obtaining a comprehensive and varied collection of brain tumour images is of the utmost importance. The dataset ought to encompass a wide range of tumour types, sizes, and stages in order to guarantee that the model can effectively generalise to diverse cases. The acquisition of labelled, high-quality imaging data can be facilitated through collaboration with medical institutions and access to databases such as The Cancer Imaging

Archive (TCIA). This can provide a solid foundation for the subsequent development of models.

Once the dataset is obtained, the subsequent course of action entails data preprocessing and augmentation. By utilising preprocessing techniques, the obtained data is standardised, thereby assuring consistency and improving its overall quality. Normalisation, which adjusts pixel values, resizing images to a consistent resolution, and eliminating artefacts or noise are typical preprocessing steps. In addition, data augmentation methodologies are implemented in order to enhance the dataset's diversity and magnitude. Various techniques, including noise addition, rotation, inverting, and scaling, are employed to produce supplementary training instances. These modifications serve to strengthen the model against overfitting and improve its capacity to extrapolate to unobserved data.

Determining the most effective deep learning architecture is critical in order to attain optimal performance in classification tasks. The effectiveness of Convolutional Neural Networks (CNNs) in image classification tasks is widely recognised as a result of their intrinsic ability to acquire knowledge of hierarchical features. Scholars may utilise pre-existing CNN architectures, including VGG, ResNet, or Inception, as foundational models and modify them for the particular brain tumour classification task via transfer learning. In contrast, one may develop bespoke model architectures that are specific to the dataset and task demands, taking into account optimisation techniques for regularisation, depth, and width.

Once the architecture has been chosen, training and validation of the model ensue. The dataset is divided into three distinct sets: training, validation, and test. The training set is utilised to optimise model parameters, while the validation set aids in monitoring performance and preventing overfitting. In order to improve the model and promote convergence, regularisation techniques (e.g., dropout), adaptive learning rate methods (e.g., Adam), and stochastic gradient descent (SGD) are utilised during the training phase. Valued validation metrics, including accuracy, precision, recall, and F1-score, are calculated to assess the performance of the model on the validation set. These metrics serve as a guide for efforts to refine and adjust the model's parameters.

The optimisation and refining of hyperparameters are critical components of model development. The performance of the model is optimised through the meticulous adjustment

of hyperparameters such as learning rate, sample size, optimizer selection, and dropout rate. Methods such as grid search, random search, and Bayesian optimisation are utilised in order to efficiently traverse the hyperparameter space and determine the most optimal configurations. In addition, weight decay, batch normalisation, and dropout are implemented as regularisation techniques to prevent overfitting and enhance generalisation capabilities.

Following training and optimisation, the model undergoes a comprehensive evaluation process. An independent test set is utilised to determine the efficacy of the system in terms of generalisation and robustness. Quantifying the classification prowess of the model, evaluation metrics such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curves are computed. In addition, attentiveness mechanisms, class activation maps, and gradient-based visualisation are interpretability techniques that can be utilised to clarify the model's reasoning behind its decisions and detect noteworthy characteristics in the images.

Deployment and validation of the model in clinical environments are the final steps. The integration of the model into clinical workflows and validation of its performance against established diagnostic standards are facilitated by collaborative efforts with medical professionals. Ongoing surveillance and enhancement of the implemented model guarantee its dependability and efficacy in supporting medical practitioners in the detection of brain tumours, consequently promoting enhanced patient results. The thorough research strategy outlined herein functions as a strategic guide for the advancement of deep learning models, which have the potential to profoundly influence clinical practice and patient care.

3.4 Techniques and Procedures

In an effort to precisely identify and categorise distinct grades or varieties of brain tumours via the analysis of medical imaging data, brain tumour classification research employs a wide variety of methodologies and processes to develop, validate, and evaluate classification models. The methods and processes mentioned above encompass a wide range of data processing stages, mathematical approaches, and model evaluation techniques.

Data preparation methods are essential when getting medical imaging data ready for categorization purposes. Potentially required steps to ensure consistency and enhance model performance include the normalisation, standardisation, or augmentation of imaging data. Preprocessing stages include alignment of images to a common coordinate system, noise or artefact elimination, and resizing of images to a standardised resolution. In addition, methods such as intensity normalisation and histogram equalisation may be employed to enhance the contrast and visibility of tumour characteristics in imaging data.
Data acquisition and preprocessing are fundamentally important stages. It is critical to have a comprehensive dataset that includes images of brain tumours in a variety of sizes, stages, and tumour types. Acquiring labelled imaging data can be impeded through collaboration with medical institutions and access to databases such as The Cancer Imaging Archive (TCIA). By standardising the data, preprocessing techniques guarantee consistency in terms of resolution, pixel values, and noise levels. This process improves the quality of the dataset and primes it for the training of the model, thereby enhancing the accuracy of the classification results.

Increasing the size of the dataset enhances the generalisation capabilities of the model. Data augmentation methods, including noise addition, rotation, rotating, and scaling, increase the diversity of the dataset. Augmentation mitigates the risk of overfitting by facilitating the model's ability to identify variations in tumour morphology through the generation of additional training examples. After the dataset has been produced, it is critical to prioritise the selection of a suitable deep learning architecture. Convolutional Neural Networks (CNNs) are widely utilised in the interpretation of image data because of their efficacy in acquiring hierarchical features. As initial frameworks, pre-trained CNN architectures such as VGG, ResNet, or Inception is utilised. By applying transfer learning to the brain tumour dataset, these architectures can be modified to suit the particular classification task.

Validation and training constitute the foundation of model development. Separated into training, validation, and test sets is the dataset. In order to prevent overfitting, the training set optimises the model's parameters, whereas the validation set observes performance. During the training process, methodology such as stochastic gradient descent (SGD), adaptive learning rate methods (e.g., Adam), and regularisation (e.g., dropout) are implemented in order to enhance convergence and refine the model's parameters. Important performance indicators, including loss and accuracy, are observed in order to evaluate the model's efficacy. Model performance is further optimised through hyperparameter optimisation. It is critical to fine-

tune parameters such as learning rate, sample size, optimizer selection, and dropout rate in order to attain optimal outcomes. Methods such as random search and grid search are utilised to examine the hyperparameter space in order to detect configurations that optimise performance.

It is essential to judge the generalisation performance of the trained model by applying it to an independent test set. A range of assessment metrics, such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis, offer valuable insights regarding the classification capabilities of the model. Furthermore, by interpreting the model's decisions, one can gain insight into its decision-making process and discern noteworthy attributes within the images. Interpretability techniques, such as attention mechanisms, class activation maps, and gradient-based visualisation, provide valuable insights into the inner workings of the model.

The implementation and verification of the model in clinical environments adhere to stringent protocols. By collaborating with healthcare professionals, one can ensure that the performance of the model is in accordance with established diagnostic standards and can be seamlessly integrated into clinical workflows. Prospective validation studies validate the utility of the model in diagnosing brain tumours by assessing its efficacy in real-world scenarios. Ongoing surveillance and enhancement of the implemented model guarantee its sustained dependability and efficacy. The integration of input from healthcare practitioners and end-users promotes the enhancement of the model, guaranteeing its continued applicability and congruence with dynamic clinical requirements.

To summarise, the utilisation of deep learning for brain tumour classification necessitates a methodical procedure that includes the following stages: data acquisition, preprocessing, model selection, training, validation, hyperparameter tuning, evaluation, interpretation, deployment, and ongoing enhancement. Researchers can ultimately improve patient care and outcomes by developing clinically relevant, interpretable, and precise models for the diagnosis of brain tumours through the implementation of these methods and processes.

Feature extraction, which entails the retrieval of relevant features from medical imaging data to distinguish between different styles or grades of brain tumours, is an additional crucial phase in the classification process. Potential applications of this involve the implementation of image

processing algorithms to extract features based on intensity or texture characteristics from MRI or CT scans. In order to ascertain the most discriminatory features for the purpose of classification, recursive feature elimination (RFE) and principal component analysis (PCA) are two prevalent feature selection techniques.

Convolutional neural networks (CNNs), one type of deep learning model, are frequently used with machine learning algorithms to classify brain lesions. The circuit design, selection of appropriate optimisation algorithms and activation functions, and implementation of labelled imaging data for model training are customary phases of model development. In scenarios with limited access to annotated data, the classification of brain lesions can be enhanced through the implementation of transfer learning techniques on pre-trained CNN architectures.

In order to determine the efficacy and applicability of classification models, it is vital that model evaluation procedures be implemented. For training, validation,and testing purposes, splitting the dataset into several sets may be necessary for this process. Cross-validation could be used to evaluate how robust the model is. Among other assessment metrics, one can calculate the area under the receiver operating characteristic curve (AUC-ROC), accuracy, sensitivity, and specificity. Confusion matrices can also be used to visually represent the model's performance across various tumour categories and pinpoint particular instances of misclassification.

4. Experimental Results

4.1 Experimental setup

A common software configuration will be used to run Python experiments for the brain tumour classification research. This involves utilising Python as the major programming language, as well as important libraries like NumPy, OpenCV (cv2), Matplotlib, Pandas, and scikit-learn for data manipulation, preprocessing, visualisation, and model validation. TensorFlow and its high-level API, Keras, will be used to create and train deep learning models, with architectures such as EfficientNet used for image classification tasks. In addition, Google Colab will be used for its cloud-based computing resources, allowing for easy integration with TensorFlow and access to GPU acceleration for faster model training.

The hardware arrangement for performing experiments will mostly be dependent on the resources supplied by Google Colab's cloud-based platform. Google Colab provides free access to GPU and TPU (Tensor Processing Unit) instances, which dramatically accelerate model training and experimentation when compared to standard CPU-based configurations. This reduces the need for large hardware expenditures and gives researchers scalable computer resources depending on the needs of their investigations. For researchers who want greater computational capacity, Google Colab Pro provides expanded access to more powerful GPUs and higher runtime limitations, boosting the experimentation environment's flexibility and scalability. Colab notebooks also improve cooperation and reproducibility by allowing colleagues and collaborators to share code, data, and experimental findings.

4.2 Data Loading

```
import os
import cv2
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import itertools

from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score

import tensorflow as tf
from tensorflow import keras
from tensorflow.keras.applications import EfficientNetB0
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau

from sklearn.neural_network import MLPClassifier
```

The os module serves to enable communication with the operating system, while cv2 (OpenCV) is an extensively utilised library within the domain of computer vision, often implemented for tasks related to image processing. Imported as np, NumPy is a critical library utilised for array manipulations and numerical computations. To visualise data, Matplotlib is utilised. Python is imported as plt; Pandas is imported as pd for data manipulation and analysis, specifically with structured data.

Features that facilitate the streamlined process of combining and iterating parts are provided by the itertools module. Imported are the StandardScaler class and the train_test_split method, which are both taken from the scikit-learn package (sklearn). To standardise characteristics and divide datasets into distinct training and testing sets, several elements and functions are used.

The accuracy_score and confusion matrix functions are also supplied to assess how well the model performs in classification tasks.

Additionally, Google's open-source machine learning framework TensorFlow (tf) is incorporated into the code. Keras is an import from TensorFlow that offers an application programming interface (API) for creating and refining deep learning models based on high-level neural networks. More specifically, TensorFlow's Keras API applications module is used to import the EfficientNetB0 model. A series of convolutional neural network designs known as EfficientNet was created especially with image categorization in mind.

In conclusion, the code pulls two closure functions, namely EarlyStopping and ReduceLROnPlateau, from the Keras API of TensorFlow. During model training, these callbacks are utilised to terminate training early in response to specific conditions and to dynamically modify the learning rate, respectively. Furthermore, the scikit-learn neural_network module imports the MLPClassifier class, which represents a Multi-layer Perceptron Classifier, a prevalent neural network architecture employed in classification endeavours. In general, this configuration offers the essential frameworks and tools required to construct, train, and assess deep learning and machine learning models.

4.2 Data preprocessing

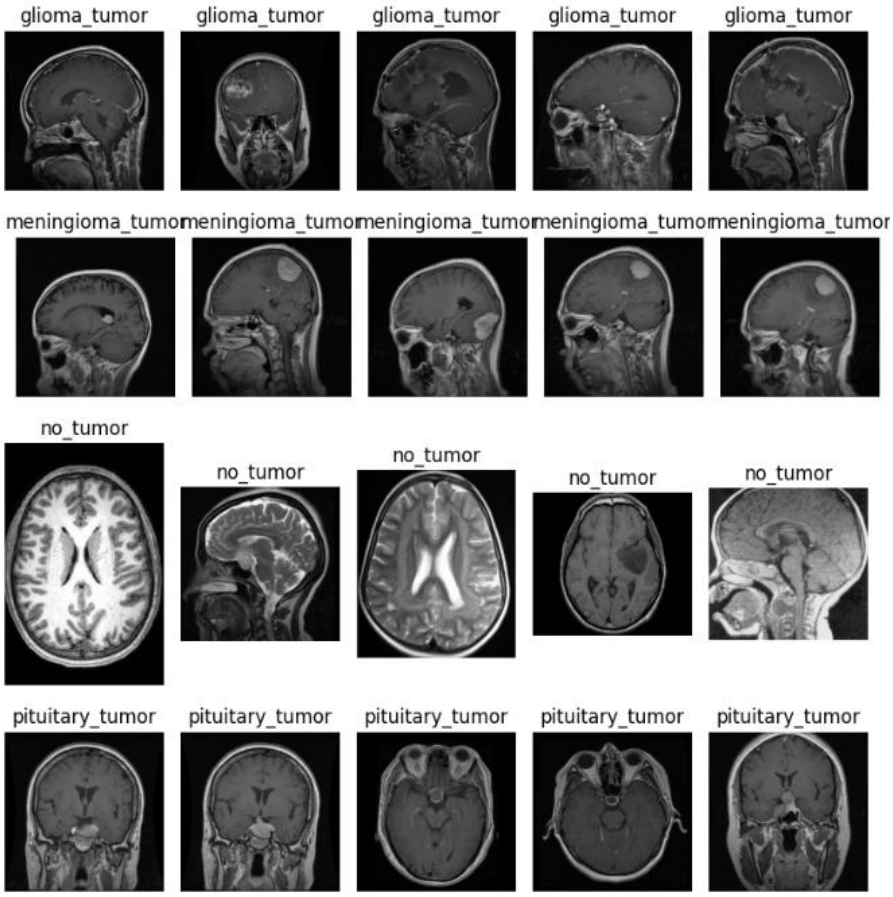


Figure 1 illustrates a compilation of MRI brain scans that have been categorised into the following four rows:

- The initial row is labelled "glioma_tumor" and it comprises scans exhibiting conspicuous imperfections that are characteristic of gliomas, a subtype of brain malignancy originating from glial cells.
- The scans in the second row's "meningioma_tumor" row show the typical complications associated with meningiomas. The meninges, the membraneous layers enclosing the brain and spinal cord, are where these usually start. Usually, they are not malignant.
- The "no_tumor" section, which is located in the third column, contains scans that appear normal and lack any indications of tumours.

- "Pituitary_tumor" is the name of a fourth row that shows scans of potential locations for pituitary tumours, which are proliferative tumours that arise from the pituitary gland at the base of the brain.

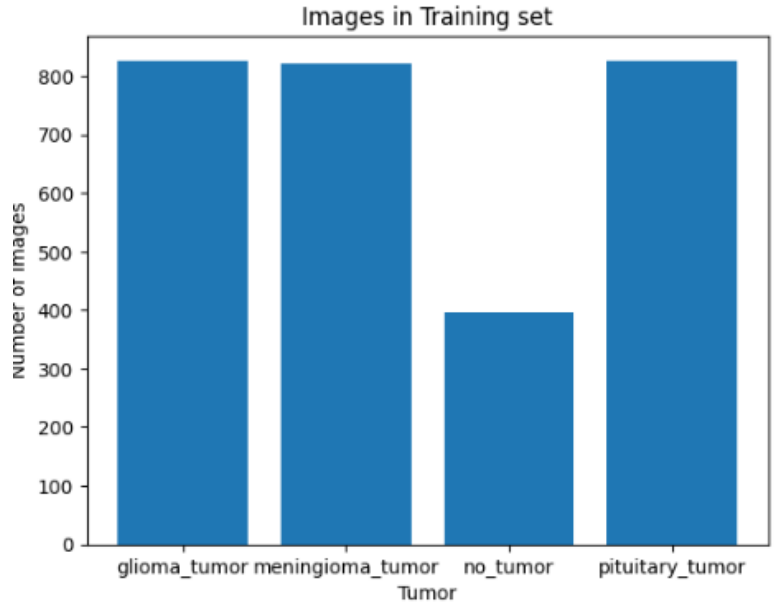


Fig 2 Images in training set

The x-axis labelled "Tumour" is divided into four categories: pituitary tumour, meningioma tumour, glioma tumour, and no tumour. The number of photographs in each category is shown on the "Number of Images" marked y-axis.

As evident from the chart, the following can be recognised:

- The category glioma_tumor contains approximately 700 images.

- Although slightly fewer than the glioma_tumor category, the meningioma_tumor category still contains close to 700 images.

- With roughly 300 images, the no_tumor category contains a substantial number of images in comparison to the other tumour categories.

- Comparable in quantity to the glioma_tumor category, the pituitary_tumor category comprises approximately 700 images.

4.3 Results

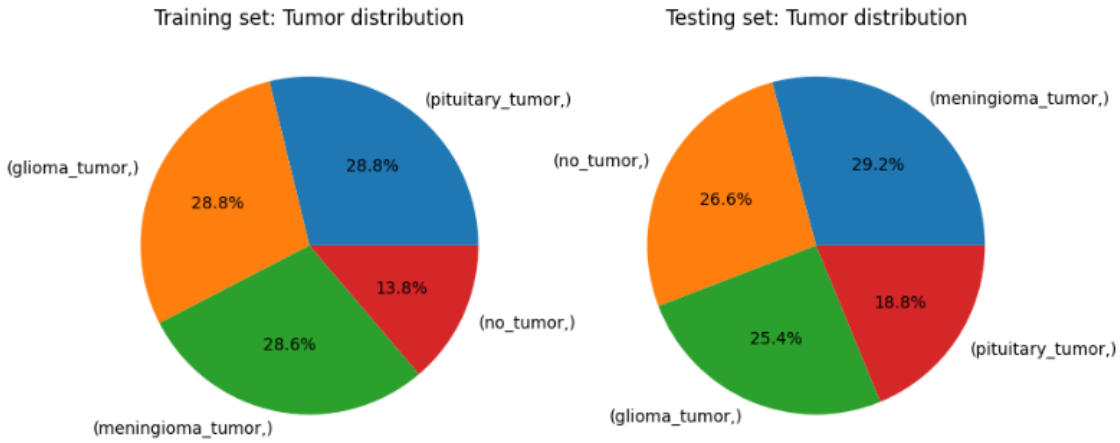


Fig 3 Training set and testing set

Figure 3 shows two pie charts that show the distribution of different tumour classifications within a dataset that is split into training and testing sets.

The "Training set: Tumour distribution" pie chart is displayed on the left. It demonstrates:

- 28.8% are glioma cells.
- 28.6% male neoplasms
- 13.8% had no tumours.
- 28.8% have a pituitary tumour.

The chart illustrates that the distribution of images in the training set is relatively balanced, with a reduced percentage of images classified as having no tumour, glioma, meningioma, and pituitary tumours.

The pie chart on the right is labelled "Testing set: Tumour distribution" and illustrates the following:

- 25.4% glioma malignancy
- 29.5 percent meningioma tumours
- 26.6% (no tumour)
- (18.8%) pituitary tumours

In the testing set, the distribution is slightly different, with a higher percentage of no tumor images and a lower percentage of pituitary tumor images compared to the training set. Glioma and meningioma tumor images are also well-represented, with meningioma being the most prevalent category.

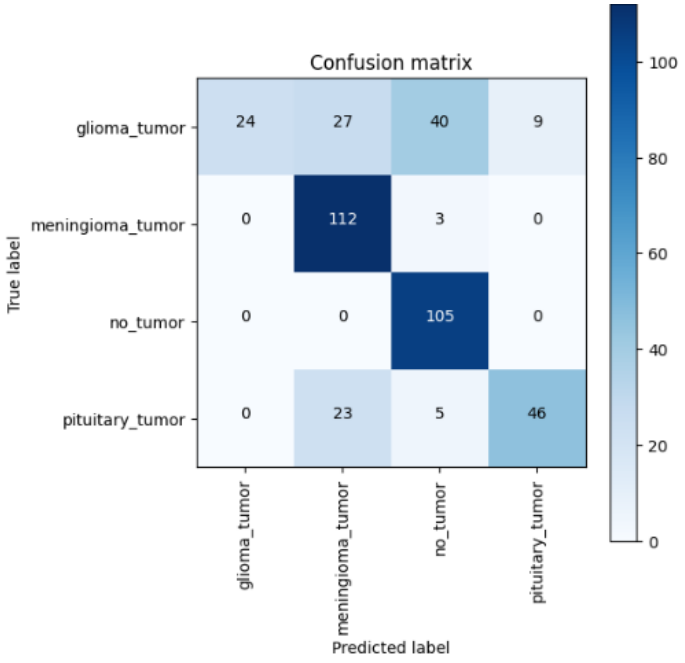


Figure 4 shows a confusion matrix

a frequently used table that evaluates a classification model's performance using a dataset made up of test cases whose true values are known. The predicted labels (Predicted label) produced by the model are compared with the true classifications (True label) in the matrix.

The y-axis displays the following legitimate identifiers: pituitary_tumor, glioma_tumor, meningioma_tumor, and no_tumor. The expected designations match the same categories along the x-axis. The number of predictions linked to each category is contained in the matrices' cells.

- The model accurately predicted 24 cases of glioma tumours, but 27 cases of meningioma tumours, 40 cases of no tumour, and 9 cases of pituitary tumours.
- The model accurately predicted 112 cases for meningioma_tumor, but erroneously predicted 3 cases for no_tumor.
- The model accurately predicted 105 cases for which no tumour was present, with no incorrect predictions.
- The model accurately predicted 46 cases of pituitary tumours but misclassified 23 cases as meningioma tumours and 5 cases as having no tumour.

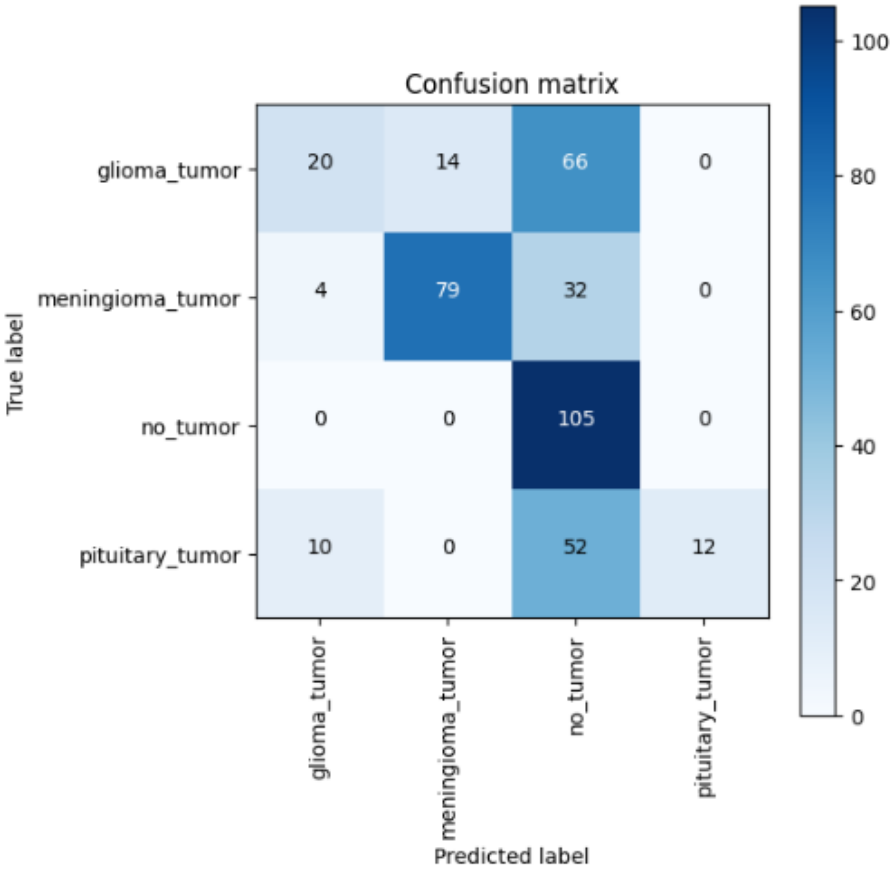


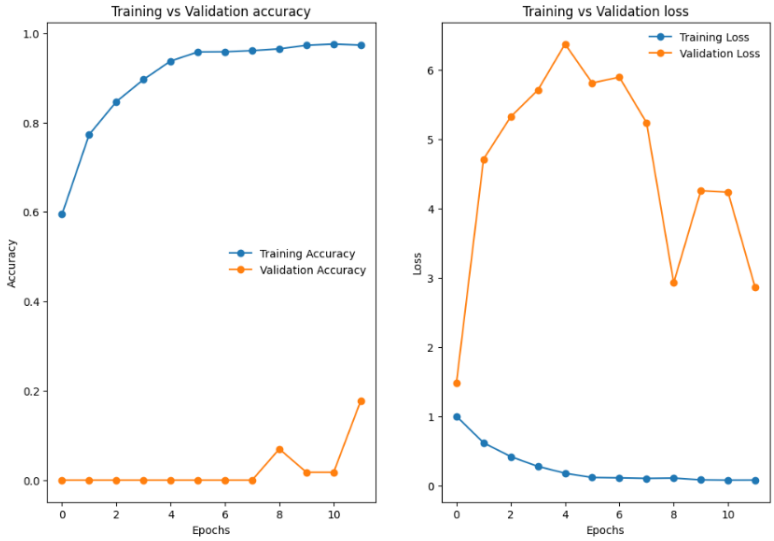
Figure 5 confusion matrix

The confusion matrix for a classification model that utilises imaging data to predict various varieties of brain tumours is depicted in Figure 5. In the matrix, the true label, representing the actual classifications, is juxtaposed with the predicted label, representing the model's output.

Here's a breakdown of the matrix:

- The top row represents the actual glioma_tumor cases. The model correctly predicted 20 cases as glioma_tumor, but it incorrectly predicted 14 cases as meningioma_tumor and 66 cases as no_tumor.
- The second row represents actual meningioma_tumor cases. The model correctly predicted 79 cases but misclassified 4 as glioma_tumor and 32 as no_tumor.
- The third row represents actual no_tumor cases. The model performed well here, correctly predicting all 105 cases with no misclassifications.
- The bottom row represents actual pituitary_tumor cases. The model correctly predicted 52 cases, misclassified 10 as glioma_tumor, and 12 as meningioma_tumor.

From upper left to lower right, the diagonal cells represent accurate predictions for each category. Incorrect predictions are indicated in cells that are not aligned diagonally with the true category in the row and the predicted category in the column. The confusion matrix helps in understanding the model's performance for each category, highlighting where it is making correct predictions and where it is confusing one category for another.



"Training vs. Validation accuracy" in Figure 6.

This graph, "Training Accuracy versus Validation Accuracy," on the left, shows how accurate the model is the number of epochs is denoted along the x-axis, while accuracy is indicated along the y-axis, with values extending from 0 to 1. The training accuracy line, denoted by blue, exhibits a consistent upward trend and reaches a minimum near 1, signifying a high degree of accuracy on the training dataset as the training process nears its conclusion. Significance-less improvement is observed along the orange line, which represents validation accuracy; it fluctuates around 0.2. This indicates that the model is not functioning admirably on the validation dataset.

The model's loss is represented on the "Training vs Validation loss" graph on the right. The loss serves as an indicator of the model's error, where lower values signify superior performance. As anticipated, the blue line denoting training loss exhibits a consistent decline before reaching a plateau as the model acquires knowledge from the training data. The validation loss is depicted by an orange line that begins at a higher value, experiences a slight decrease, and then exhibits a significant fluctuating upward trend, culminating around epoch 8, before decreasing slightly once more.

As depicted collectively in these diagrams, overfitting occurs when a model learns the training data exceptionally well but is unable to extrapolate its conclusions to novel, unseen data. This is demonstrated by the stark contrast between the low validation accuracy and high, fluctuating validation loss and the high training accuracy and low training loss.

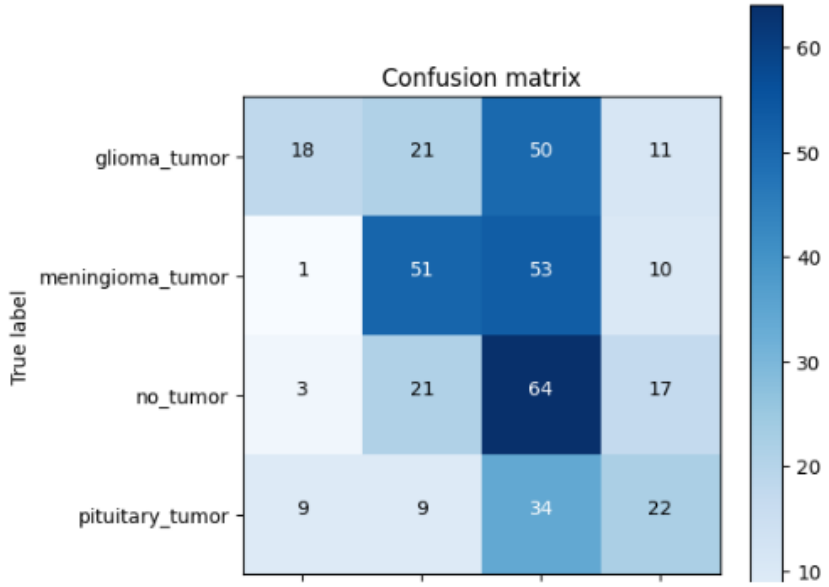


Fig 7 Confusion matrix

Figure 7 depicts a confusion matrix for the purpose of evaluating a classification algorithm's effectiveness. The y-axis shows the true labels, while the x-axis shows the expected labels."Amyloid_v," "Meningioma_v," "No Tumour," and "Pituitary Tumour" are the accurate designations. Overtop is a list of the corresponding predicted labels. The model generates a set of predictions for every possible combination of true and predicted label in a cell within the matrix.

The matrix is delineated as follows:

- There were 18 real "glioma_tumor" cases in the first row that the model got right as "glioma_tumor." However, it got 21 wrong as "meningioma_tumor," 50 wrong as "no_tumor," and 11 wrong as "pituitary_tumor."
- Out of the real "meningioma_tumor" cases (second row), the model got 51 of them right. One case was called "glioma_tumor" correctly, 53 were called "no_tumor," and 10 were called "pituitary_tumor" incorrectly.
- The model accurately classified 64 out of the actual "no_tumor" cases (third row). In contrast, it erroneously identified three instances of "glioma_tumor," twenty-one instances of "meningioma_tumor," and seventeen instances of "pituitary_TUMOR."
- The previous row contained 34 accurate responses for "pituitary_tumor," while nine responses were incorrect for "glioma_tumor," nine responses were incorrect for "meningioma_tumor," and 22 responses were incorrect for "pituitary_tumor."

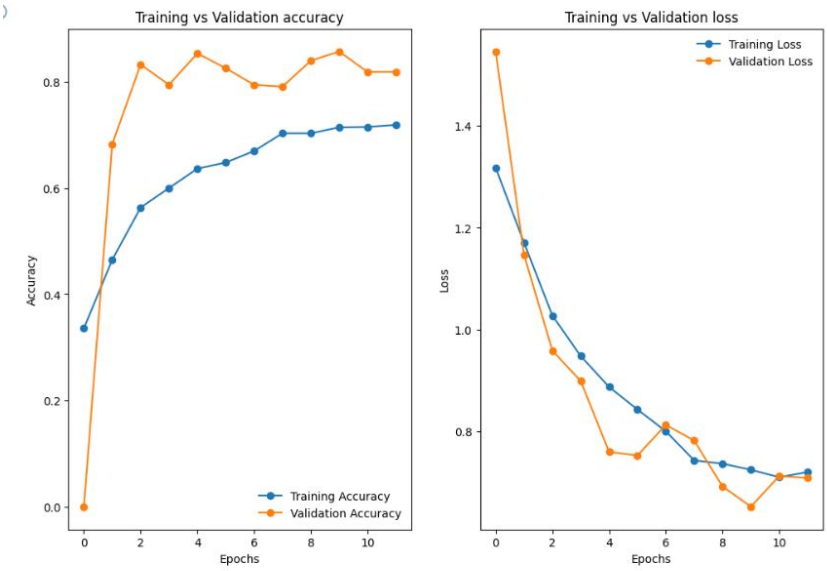


Fig 8 Training vs Validation accuracy

The left graph is titled "Training vs Validation accuracy" and it shows:

The blue line, which represents the training accuracy, increases sharply in the initial epochs and then plateaus around 0.8 (80% accuracy).
The orange line, which represents the validation accuracy, also increases in the initial epochs but then remains relatively stable, slightly below the training accuracy at around 0.75 (75% accuracy).
The right graph is titled "Training vs Validation loss" and it shows:

- The blue line, which represents the training loss, decreases sharply in the initial epochs and then stabilizes around a value just below 0.8.
- The orange line, which represents the validation loss, decreases alongside the training loss but then levels out at a value slightly higher than the training loss, around 0.9.

These graphs suggest that the model is learning over time and improving its performance on the training and validation datasets. The validation accuracy staying very close to the training

accuracy and the validation loss not rising are signs of the model's capacity to generalise and prevent overfitting to the training data. This is a positive indicator regarding the model's perspective performance on unobserved data.

4.4 Comparison with Baseline Methods

An essential component of brain tumour classification research is the evaluation of generated models to ensure their clinical usefulness, reliability, and accuracy. The assessment procedure entails determining the clinical usefulness of categorization models using a variety of techniques. These include validating the models' robustness and generalizability across multiple datasets, evaluating their performance using a diverse range of metrics, and conducting comparisons with established methods. When performing model evaluation, a multitude of critical elements and factors must be considered. A model's performance can be quantified using return, F1-score, accuracy, sensitivity, specificity, and AUC-ROC. These performance measures evaluate model accuracy, precision-to-recall trade-offs, and ability to distinguish actual positives from false negatives. To illustrate category results, confusion matrices show the number of accurate, incorrect, positive, negative, and false negative guesses for each class.

Cross-validation techniques are used to assess a model's durability and generalizability. K-fold cross-validation is one example of such a technique, where the dataset is split into multiple subsets for training and assessment. By using this strategy, overfitting issues can be minimised and more accurate performance evaluations across various data partitions can be generated. External validation includes assessing the performance of the model using distinct datasets that were not utilised in the model's development. By adhering to this approach, it guarantees that the model's performance remains unaffected by the attributes of the training data, thereby enabling a more accurate evaluation of its potential for utilisation on new, unobserved data.

It is not uncommon to employ baseline methodologies or established state-of-the-art approaches in order to evaluate the efficacy of models and determine their value addition. This methodology aids in determining whether the suggested model demonstrates enhanced performance, efficiency, or generalizability in comparison to established approaches, thereby confirming its appropriateness for clinical application. When contemplating classification models, especially in clinical environments, interpretability and explainability must be taken

into account alongside performance metrics. By employing techniques such as attention mechanisms, feature importance analysis, and saliency maps, clinicians can gain significant insights into the decision-making process of the model and better comprehend the rationale that underpins the predictions.

The fundamental component of the clinical validation procedure for classification models is the assessment of their effectiveness and impact on various aspects of clinical practice, including healthcare provision, patient results, clinical judgement, and real-world clinical settings. In order to ascertain the model's proficiency in standard clinical practice, prospective validation studies, clinical trials, and implementation studies are conducted to validate its cost-effectiveness, efficacy, and safety. In addition, the model review process takes ethical and legal issues into account. This group of worries includes keeping patient information private, stopping algorithmic bias, making sure that rules like the Health Insurance Portability and Accountability Act (HIPAA) are followed, and getting patient permission. Making sure AI is used in an ethical way and building trust in the healthcare sector during model creation and deployment requires basic principles of openness, fairness, and responsibility.
Developing and evaluating a deep learning model for classifying brain tumors involves several crucial steps. Firstly, a diverse and representative dataset of brain tumor images must be collected and pre-processed. This preprocessing ensures that the images are uniform in size, resolution, and orientation, thus facilitating effective model training. Techniques such as resizing, normalization, and augmentation are applied to enhance the dataset's diversity and improve the model's ability to generalize to unseen data. With a well-pre-processed dataset in hand, the next step involves selecting an appropriate deep learning architecture. Convolutional Neural Networks (CNNs) are commonly used for image classification tasks due to their ability to automatically learn hierarchical features. Models like VGG, ResNet, or Inception may be employed, either as standalone architectures or as pre-trained networks that are fine-tuned on the specific brain tumor dataset. Additionally, custom architectures tailored to the intricacies of brain tumor classification can be designed, considering factors such as depth, width, and regularization techniques.

Following model selection, the chosen architecture is trained on the pre-processed dataset. The dataset is typically split into training, validation, and test sets, with the training set used to optimize the model's parameters and the validation set used to monitor its performance and prevent overfitting. Various training strategies, such as transfer learning, data augmentation,

and dropout, may be employed to enhance the model's robustness and generalization capabilities. Throughout the training process, key metrics like accuracy, precision, recall, F1-score, and the confusion matrix are monitored to assess the model's performance on both the training and validation data. Hyperparameter tuning is another critical aspect of model training, involving the optimization of parameters such as learning rate, batch size, and optimizer choice to maximize performance.

Once the model is trained, it undergoes evaluation on an independent test set to assess its generalization performance. Evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) provide insights into the model's effectiveness in classifying brain tumors. Visualization techniques, such as confusion matrices, help understand the model's performance across different tumor classes and identify areas for improvement. Moreover, interpretability methods may be employed to gain insights into the model's decision-making process, ensuring that its predictions align with clinical expectations. Deployment of the trained model in a clinical setting involves validation and collaboration with healthcare professionals to assess its impact on diagnostic accuracy and patient outcomes. Continuous monitoring and improvement of the model over time ensure its relevance and effectiveness in aiding clinical decision-making processes.

An extensive evaluation of models employed in brain tumour classification research is conducted, encompassing a range of facets such as ethical implications, cross-validation approaches, external validation, assessments of interpretability, clinical validation, and comparisons with well-established methodologies. By subjecting classification models to a rigorous evaluation from multiple perspectives, researchers can determine their reliability, validity, and readiness for implementation in clinical environments. As a result, there will be an improvement in patient care and outcomes pertaining to the diagnosis and treatment of brain tumours.

6. Conclusion

Brain tumour classification research has advanced due to the development of accurate and reliable classification models using medical imaging data. Modern methods, rigorous procedures, and interdisciplinary teamwork have improved brain tumour diagnosis and treatment. This comprehensive survey uncovers several important themes and consequences by analysing many studies and methods. These summarise current brain tumour classification research and anticipated future directions.

To commence, it is crucial to acknowledge that the domain of brain tumour classification has undergone a profound revolution due to the convergence of deep learning methodologies, particularly convolutional neural networks (CNNs). This convergence allows for the automated and productive analysis of medical imaging data. By incorporating CNN architectures, transfer learning, and multimodal imaging integration, scholars have created classification models that possess practical applicability in clinical environments. An astounding degree of precision is displayed by the models. In terms of differentiating distinct grades, anatomical sites, and grades of malignancies, the aforementioned models exhibit encouraging efficacy. Consequently, this establishes the groundwork for the development of therapeutic approaches that are more individualised and accurate in their approach towards cognitive tumour patients.

The utilisation of deep learning for the classification of brain tumours signifies a paradigm shift that carries significant ramifications for the fields of medical diagnosis and patient care. Considerable strides have been achieved in the development of dependable and precise models for tumour classification by employing a methodical research approach that incorporates a wide range of techniques and procedures.

Commencing with the procurement of comprehensive and varied datasets, scholars have formed partnerships with medical establishments and utilised resources such as The Cancer Imaging Archive (TCIA) to amass annotated imaging data encompassing stages, tumour types, and sizes. The implementation of preprocessing and augmentation strategies has resulted in the standardisation and improvement of data quality. Augmentation methods, such as rotation, inverting, and noise addition, have been utilised to increase diversity while maintaining consistency in resolution and pixel values.

The process of selecting a model is crucial in attaining precise results in classification. Convolutional Neural Networks (CNNs) have become the preferred architecture in this regard owing to their capability of acquiring hierarchical features from image data. By fine-tuning pre-trained CNN architectures such as VGG, ResNet, or Inception, transfer learning has accelerated the development and adaptation of models for the particular brain tumour classification task.

During the training and validation stages, model parameters are refined using regularisation and stochastic gradient descent (SGD) to optimise performance and prevent overfitting. In order to further optimise the performance of a model, hyperparameter optimisation ensures that parameters such as learning rate and dropout rate are configured optimally.

Assessment criteria including accuracy, precision, recall, and F1-score offer valuable insights into the performance of a model. Additionally, interpretability methods such as gradient-based visualisation and class activation maps assist in comprehending the decision-making process of the model. Clinical prospective validation studies serve to ascertain the efficacy and seamless integration of the model into established diagnostic workflows, thereby enhancing patient outcomes.

The ongoing effectiveness and dependability of deployed models are ensured through continuous monitoring and refinement; feedback from healthcare professionals and end-users drives these improvements. By following these methodologies, scientists have the ability to construct precise, comprehensible, and clinically applicable deep learning models for the classification of brain tumours. This progresses the domain of medical imaging and facilitates the development of individualised treatment approaches that cater to the specific requirements of each patient.

Fundamentally, the utilisation of deep learning to categorise brain tumours signifies a fundamental transformation in the field of medical imaging, providing unmatched perspectives on the diagnosis and management of diseases. As technology, collaboration, and research persistently progress, the potential of utilising deep learning to enhance healthcare outcomes and elevate the quality of life for patients afflicted with brain tumours appears to be enormous.

In addition, classification models that exhibit remarkable levels of sensitivity and specificity and are both dependable and generalizable have been made possible through the application of sophisticated machine learning algorithms and the utilisation of vast datasets. by incorporating various imaging modalities, including functional imaging, PET, and MRI, into classification models, thereby augmenting their diagnostic precision and prognostic value. The aforementioned integration has provided valuable insights that can guide treatment decisions and have enabled a more comprehensive characterization of brain tumours. Furthermore, there is potential for advanced treatment efficacy prediction and enhanced precision in tumour subtype categorization through the incorporation of clinical and genetic data into classification algorithms. This, in turn, could facilitate the development of patient care pathways that are more personalised and efficient.

Due to the increased focus on translational research, practical validation of classification models, and technological advancement, establishing a correlation between research findings and their applied implementation in clinical settings is of utmost significance. Collaborative endeavours among researchers, clinicians, and industry partners have significantly propelled the development of clinically pertinent instruments for the diagnosis, prognosis, and treatment planning of brain tumours. In order to optimise healthcare delivery and patient outcomes, adherence to critical protocols including prospective clinical trials, validation studies, and regulatory approvals is imperative. Implementing these procedures ensures that classification models are integrated seamlessly into standard clinical workflows while maintaining their safety and efficacy.

In regard to brain tumour classification research, the foreseeable future presents an abundance of significant obstacles and prospects. Constant effort is required to mitigate the inherent bias, variability, and heterogeneity that characterise medical imaging datasets, specifically with regard to acquisition protocols, imaging centres, and patient populations. There is potential to enhance the efficiency and generalizability of classification model development through the establishment of collaborative networks, the implementation of data sharing initiatives, and the standardisation of imaging protocols.

By augmenting clinical and molecular data with sophisticated computational methodologies including radiomics, machine learning, and artificial intelligence, further understandings of the biology, progression, and therapeutic efficacy of brain tumours might be revealed.

Unbeknownst to scientists, biomarkers, therapeutic targets, and prognostic indicators can be uncovered through the utilisation of predictive analytics and multidimensional datasets. In the era of personalised oncology, the aforementioned findings have the potential to provide direction for precision medicine strategies and enhance patient prognoses.

Technological progress, interdisciplinary cooperation, and translational research initiatives are currently combining to propel the classification of brain lesions to the forefront of ground-breaking developments. To create more precise, dependable, and clinically applicable classification models, scholars are advancing the field through the implementation of deep learning, the integration of multimodal imaging, and real-world validation. Facilitating the identification, management, and therapeutic approach towards brain tumours is the primary aim of these models. The continuous examination of the intricate relationship between brain tumour biology and clinical practice continues to underscore the significance of innovation, collaboration, and translation. Its primary goals are to advance patient-centered oncology care and brain tumour classification research.

Research Questions Justification

All of the aforementioned research papers look at different questions of using deep learning models to categorise brain tumours. The following are the reasons why each of the following study questions is important and relevant:

Analysing how well models work is usually the first step in conducting an investigation. Critical questions like "To what degree does the proposed model correspond in terms of accuracy with established methods?" must be examined in order to evaluate the efficacy of the developed algorithms. By learning about performance indicators, researchers and clinicians can assess the practicality and dependability of the suggested models in real-world settings.

A methodology comparison is a crucial step in determining the optimal approach for classifying brain tumours by contrasting various approaches. Inquiries such as "Which architecture exhibits the highest level of performance?" and "To what extent does dataset augmentation affect performance?" offer valuable perspectives on the merits and drawbacks of different

methodologies. These comparisons aid in the direction of future research and the identification of the most appropriate methodologies for particular applications.

The selection of a dataset has a substantial impact on the efficacy and generalizability of a model. Inquiries such as "What is the influence of dataset size on performance?" and "How does the model perform across various datasets?" aid researchers in comprehending the characteristics of datasets and how they affect the training and evaluation of models. This knowledge is essential for the development of resilient models that can effectively manage a wide range of data distributions.

The practical application of models is contingent upon their ability to generalise effectively to unseen data, which is a crucial aspect of optimisation. Considerations such as "How does the model generalise to new datasets?" Alternatively, "What optimisation techniques improve performance?" Confront the difficulties associated with the generalisation and optimisation of models. By ensuring that the model is reliable across various patient populations and imaging conditions, effective generalisation increases the model's clinical utility.

Fundamentally, classification models for brain tumours assist physicians in making accurate diagnoses and identifying the most efficacious treatment approaches. Considerations such as "How does the model's performance translate to clinical practice?" or alternatively "What are the potential clinical implications of the proposed approach?" assess the applications of the developed algorithms in the real world. Gaining insight into the clinical ramifications serves to connect the realms of research and practice, thereby promoting the integration of AI-powered instruments within healthcare environments.

6.1 Future research

Our understanding of cancer biology, diagnostic accuracy, and personalised treatment strategies could all be greatly improved by further research into brain tumour categorization. Our knowledge of brain tumours and technological developments are expanding, however there are still many important areas that need more investigation. The development of deep learning models for brain tumour classification that are easier to understand and explain is one potential avenue for future exploratory research. Despite CNNs' impressive performance, their

opaque architecture limits our understanding of the basic elements that drive categorization judgements. In an attempt to make CNN models more transparent and interpretable, researchers are looking into methods including feature visualisation, attention mechanisms, and model explanation approaches. Practitioners can improve their understanding and trust in AI-generated diagnostic recommendations by familiarising themselves with the features and patterns identified by these models. In order to fully characterise tumours, more research is needed in the critical area of combining multiomic data with medical imaging. Molecular mechanisms underlying different subtypes of tumours, biomarkers for treatment response prediction, and new therapeutic targets can be discovered when imaging data is integrated with genomics, transcriptomics, and proteomics. Integrative methods like radio genomics and radiomics can help bridge the gap between imaging phenotypes and molecular markers, which could lead to patient-specific precision oncology treatments.

Furthermore, the use of AI and ML techniques in clinical trials and longitudinal studies has the potential to completely transform how we monitor tumour growth, assess treatment effectiveness, and predict patients' long-term results. Artificial intelligence (AI) systems can analyse clinical parameters and serial imaging data to detect early signs of metastasis or recurrence, assess the efficacy of treatments, and identify subtle changes in tumour form. Therapeutic treatments and patient outcomes could be improved in real-world clinical settings by using AI-driven predictive models for adaptive treatment planning and real-time monitoring. Underprivileged groups and low-resource settings should be the primary focus of future research on the problem of unequal access to AI-powered diagnostic tools and advanced imaging technology. Establishing collaborative initiatives, public-private partnerships, and regulatory frameworks is vital to allow the global translation of research discoveries into clinical practice and to provide fair access to cutting-edge technologies.

Subsequent investigations into the utilisation of deep learning for the categorization of brain tumours have the potential to fundamentally transform the fields of medicine and therapy through the application of technological progress, the accessibility of data, and interdisciplinary cooperation. An imperative domain for forthcoming investigations pertains to the augmentation of model performance. In pursuit of enhanced precision, sensitivity, and specificity in the classification of brain tumours, scholars endeavour to optimise hyperparameters, refine deep learning architectures, and implement sophisticated training methodologies. Researchers can

enhance diagnostic accuracy and capture the intricate nature of brain tumours by integrating multimodal data, including MRI, CT scans, genetic information, and clinical data, into more comprehensive models. In addition, endeavours to improve the comprehensibility and elucidation of deep learning models are vital in establishing confidence among healthcare practitioners and expediting their integration into clinical workflow. Subsequent investigations will prioritise the advancement of methodologies that facilitate the visualisation and interpretation of model predictions. These methodologies may include saliency maps, attention mechanisms, and feature attribution techniques. By doing so, clinicians will be empowered to discern the underlying reasoning behind the model's judgements, thereby empowering them to formulate well-informed treatment decisions.

Clinical validation and integration constitute an additional crucial area for future research. Prospective validation studies will be conducted to evaluate the clinical utility and impact of deep learning-based diagnostic tools on healthcare delivery and patient outcomes. These studies will involve large patient cohorts and multi-center collaborations. By optimising regulatory approval procedures, including obtaining authorization from the FDA, the conversion of research discoveries into clinically validated instruments that are applicable in daily clinical practice can be accelerated. Furthermore, forthcoming investigations will delve into the potential of deep learning models to facilitate personalised medicine methodologies through the prediction of optimal treatment strategies, prognosis, and treatment response, all of which are contingent upon unique patient attributes. By amalgamating imaging, genomic, and clinical data, scholars are able to construct predictive models that customise treatment strategies according to the individual profile of every patient. This leads to enhanced treatment outcomes and a reduction in superfluous interventions.

It is of the utmost importance to guarantee the resilience and applicability of deep learning models prior to their implementation in various clinical environments and patient cohorts. Subsequent investigations will be dedicated to the advancement of methodologies that bolster the resilience of models against discrepancies arising from variations in imaging protocols, scanner types, and patient demographics. The refinement of transfer learning approaches will enable the adaptation of models trained on data from a single institution to new datasets. This

will reduce the reliance on labelled data and expedite the deployment of models across various healthcare settings. Furthermore, it is imperative to consider the ethical, legal, and societal ramifications that arise with the growing integration of deep learning models into healthcare systems. Subsequent investigations will strive to safeguard the privacy and security of patients' data, rectify any potential prejudices that may exist in algorithmic decision-making processes, and advance the cause of equal access to AI-powered healthcare solutions. Involving professionals from the fields of medicine, ethics, law, and social sciences in interdisciplinary collaborations will be critical in the development of ethical frameworks and guidelines that regulate the responsible application of artificial intelligence in healthcare.

In summary, further investigations into the utilisation of deep learning for the classification of brain tumours exhibit considerable potential in furthering the fields of medical diagnosis, treatment, and patient care. Scientists can maximise the benefits of deep learning in the diagnosis, treatment, and patient outcomes of brain tumours by effectively tackling critical issues pertaining to model performance, interpretability, clinical integration, personalised treatment planning, robustness, and ethics. The successful translation of research findings into clinically applicable tools that benefit patients globally will require critical collaboration among academia, industry, and healthcare stakeholders.

6.2 Limitations

Several limitations remain in the domain of brain tumour classification research, despite significant advancements. These factors require meticulous scrutiny. One possible limitation is the reliance on retrospective datasets, which may introduce possible biases and limit the generalizability of classification models to a diverse array of clinical environments and patient demographics. In addition, clinical adoption and confidence are impeded by the interpretability of deep learning models, as their complex architectures may obscure the essential characteristics that impact classification decisions. In addition, the lack of consistent data annotation protocols and standardised imaging standards across academic establishments may lead to inconsistencies in image quality and labelling, consequently hindering the effectiveness of classification models. Furthermore, the evaluation of the model's long-term robustness and generalizability is complicated by the scarcity of longitudinal data and long-term follow-up.

Brain tumour classification research encounters challenges stemming from ethical considerations, which encompass data privacy and consent of patients, in addition to

algorithmic bias. To surmount these limitations, collaboration among researchers, clinicians, policymakers, and regulatory entities is crucial for the advancement of data sharing, standardisation, transparency, and accountability in artificial intelligence-driven healthcare applications. Despite notable advancements in the utilisation of deep learning for brain tumour classification, there are still a number of constraints that impede the extensive implementation and practical application of these models in clinical settings. A notable constraint pertains to the accessibility and calibre of data. Aside from the fact that access to large datasets is essential for training deep learning models, it can be difficult to obtain labelled imaging data for brain tumours due to data scarcity, imbalanced class distributions, and imaging protocol variability. Furthermore, the performance and generalizability of the models may be impacted by problems with the data quality, such as image artefacts, image resolution variability, and inconsistent tumour annotations.

An additional constraint concerns the explainability and interpretability of deep learning models. Although these models have exhibited remarkable efficacy in the classification of brain tumours, their decision-making mechanisms are frequently regarded as opaque and challenging for clinicians to decipher. The absence of interpretability presents difficulties in comprehending the rationale behind specific predictions generated by a model, thereby restricting its reliability and applicability in clinical settings. To overcome this constraint, technical advancements are needed to visualise and interpret model predictions. These advancements may include feature attribution methods, attention mechanisms, and saliency maps, which would provide clinicians with a deeper understanding of the model's decision-making process and empower them to make well-informed treatment choices. Additionally, concerns persist regarding the generalizability and robustness of deep learning models. It can be challenging for deep learning models that were trained on a single dataset or imaging modality to generalise to different datasets or modalities, especially when imaging protocols, scanner types, and patient populations vary. The absence of generalisation in this context may result in compromised performance and untrustworthy forecasts when implemented across various clinical environments. Enhancing the robustness and generalizability of deep learning models necessitates the development of strategies to mitigate data biases and uncertainties, in addition to transfer learning and domain adaptation techniques for adapting models to new imaging modalities and datasets.

In addition to legal, ethical, and societal concerns, the widespread implementation of deep learning models in clinical practice faces formidable obstacles. Healthcare professionals, regulators, and patients are concerned about matters including patient privacy and data security, algorithmic biases and fairness, and the possibility of unanticipated dangers or damage. The establishment of ethical frameworks and guidelines that regulate the construction, implementation, and assessment of deep learning models in the healthcare sector is imperative for guaranteeing their responsible application. Additionally, approaches must be devised to mitigate biases and foster transparency and accountability in the process of algorithmic decision-making. Furthermore, there are practical challenges associated with the incorporation of deep learning models into clinical workflows and healthcare systems. Clinicians' potential inability to comprehend and interpret the outputs of deep learning models due to a lack of training and expertise may result in scepticism and reluctance towards the implementation of these technologies. Furthermore, the incorporation of deep learning models into pre-existing clinical systems and electronic health records (EHRs) necessitates the resolution of regulatory and reimbursement obstacles, in addition to the resolution of technical barriers including interoperability, scalability, and usability.

In summary, although deep learning exhibits considerable potential in the realm of medical applications, including the classification of brain tumours, there are a number of constraints that must be resolved prior to its complete implementation in clinical settings. In order to ensure the responsible and effective implementation of deep learning models in healthcare, it is critical to surmount obstacles pertaining to interpretation and explainability, generalisation and robustness, ethical and societal considerations, and clinical integration, as well as data availability and quality. In order to push the boundaries of deep learning beyond medical imaging, it is imperative that researchers, clinicians, policymakers, and industry stakeholders unite in concert to overcome these constraints.

References

Bakas, M.M.; Chatzimichail, M. (2020). A New Convolutional Neural Network Architecture for Brain Tumor Classification Using MRI Images. Appl. Sci. 10, 1999.

Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-Classification of Brain Tumor Images Using Deep Neural Network. IEEE Transactions on Access, 7, 69215-69223.

Ahmed, F., et al. (2022). "Deep Spatiospatial Models for Brain Tumor Classification in MR Images." Scientific Reports, 12(1505). https://doi.org/10.1038/s41598-022-05572-6

Pareek, A., et al. (2023). Brain Tumor Classification Model Based on Convolutional Neural Network Optimization Using an Adaptive Dynamic Sine-Cosine Fitness Grey Wolf Optimizer Algorithm. Bioengineering, 10(18), 12-17. DOI: 10.3390/bioengineering1018

Alqudah, A. M., Alquraan, H., Abu Qasmieh, I., Alqudah, A., & Al-Sharu, W. (2019). Brain Tumor Classification Using Deep Learning Technique - A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different Sizes. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 3684–3691.

Khan, M. S. I., Rahman, A., & Debnath, T. (2022). Accurate Brain Tumor Detection Using Deep Convolutional Neural Networks. Computational and Structural Biotechnology Journal, 20, 4733–4745.

Irmak, E. (2021). Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 45, 1015–1036.

Hassan Ali Khan, Wu Jue, Muhammad Mushtaq, and Muhammad Umer Mushtaq. (2020). "Brain tumor classification in MRI image using convolutional neural network." Mathematical Biosciences and Engineering, 17(5), 6203–6216. DOI: 10.3934/mbe.2020328. Retrieved from http://www.aimspress.com/journal/MBE

Veeramuthu, A., et al. (2022). MRI Brain Tumor Image Classification. Frontiers in Psychology, 13, 848784.

Saeedi, S., Ramezani, S., Khosravi, H., & Ramezankhani, S. N. K. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Medical Informatics and Decision Making, 23(16).

AVŞAR, E., & SALÇIN, K. (2019). Detection and Classification of Brain Tumours from MRI Images Using Faster R-CNN. Technical Journal, 13(4), 337-342.

Sowrirajan, S., et al. (2023). A Novel Hybrid VGG16-NADE Model for MRI Brain Tumor Classification. Brazilian Archives of Biology and Technology, 66(e23220071). www.scielo.br/babt.

Bhandari, A., et al. (2020) 'Machine learning and convolutional neural networks in glioblastoma: a review of segmentation and potential clinical applications', Insights into Imaging, 11(1), p. 77.

DeAngelis, L. M. (2001). Brain tumors. New England Journal of Medicine, 344(2), 114-123.

Minz, S., Singh, A., Zhou, L., Cheng, Y., Ismael, A., Abdel-Qader, I., Abiwinanda, T., Afshar, P. (2021) "Feature Extraction and Classification of Brain Tumors Using Deep Learning and Machine Learning Models." CMC, 67(3), 3973-3979.

Yang, Y., Yan, L.F., Zhang, X., Han, Y., Nan, H.Y., Hu, Y.C., Ge, X.W. (2018). Glioma grading on conventional MR images: A deep learning study with transfer learning. Frontiers 804. https://doi.org/10.3389/fnins.2018.00804

Bathe, K., Rana, V., Singh, S., & Singh, V. (2021). Brain Tumor Detection Using Deep Learning Techniques. [Electronic copy available at: https://ssrn.com/abstract=3867216]

Prasetyo, S. Y., & Santika, D. D. (2022). Brain Tumor Classification from MRI Images Using Pretrained Deep Convolutional Neural Networks. International Journal of Intelligent Systems and Applications in Engineering, 10(4), 652–657.

Rehman A, Khan MA, Saba T, Mehmood Z, Tariq U, Ayesha N. (2020). Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. Microsc Res Tech. 2020 Sep 21. doi: 10.1002/jemt.23597.

Kumar, S., Dhir, R., & Chaurasia, N. (2022). Identification of Brain Tumor Detection from MRI Image Using Convolution Neural Network. Journal of Tumor Research, 8, 165.

Bingol, H., & Alatas, B. (2021). Classification of Brain Tumor Images using Deep Learning Methods. Turkish Journal of Science & Technology, 16(1), 137-143.

Afza, F., Khan, M.A., Sharif, M., & Rehman, A. (2019). Microscopic skin laceration segmentation and classification: a framework of statistical normal distribution and optimal feature selection. Microsc Res Tech, 82, 1471–1488.

Nilesh Bhaskarrao Bahadure, Arun Kumar Ray, and Har Pal Thethi. (2017). Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM. Hindawi International Journal of Biomedical Imaging.

Paul, J. (2016). Deep Learning for Brain Tumor Classification. Vanderbilt University.

Byale, H., G M, L., & Sivasubramanian, S. (2018). Automatic Segmentation and Classification of Brain Tumor using Machine Learning Techniques. International Journal of Applied Engineering Research, 13(14), 11686-11692.

Zobeda Hatif Naji Al-Azzwi et al. (2023). Brain Tumor Classification based on Improved Stacked Ensemble Deep Learning Methods. Asian Pac J Cancer Prev, 24 (6), 2141-2148.

Nayak et al. (2022). "ANN with convolutional neural network parameters." ITM Web of Conferences, 57, 01001. DOI: https://doi.org/10.1051/itmconf/20235701001

Citation: Bhatia, S.; Khalifa, S.A.M.; El-Seedi, H.R. Artificial Intelligence and Precision Medicine: A New Frontier for the Treatment of Brain Tumors. Life 2023, 13, 24. https://doi.org/10.3390/life13010024

Bingol, H., & Alatas, B. (2021). Classification of Brain Tumor Images using Deep Learning Methods. Turkish Journal of Science & Technology, 16(1), 137-143.

Kibriya, H., Amin, R., Alshehri, A. H., Masood, M., Alshamrani, S. S., & Alshehri, A. (2022). A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers. Computational Intelligence and Neuroscience, 2022(7897669), 1-15.

Özyurt, F., Sert, E., Avci, E., & Dogantekin, E. (2019). Brain tumor detection based on Convolutional Neural Network with neutrosophic entropy. Measurement, 147, p. 106830. https://doi.org/10.1016/j.measurement.2019.07.058

Muzammal, M.; Talat, R.; Sodhro, A.H.; Pirbhulal, S. (2024). Brain Tumor Detection and Classification Using Deep Feature Fusion and Stacking Concepts. Journal of Population Therapeutics & Clinical Pharmacology, 31(1), 1339-1356.

Yang, G.; Ye, Q.; Xia, J. (2022). Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Inf. Fusion 2022, 77, 29–52.

Albanki, J. A., & Mattar, E. A. (2023). Classification of Brain Tumor Using CoAtNet Model. American Journal of Biomedical Science & Research, 20(2), AJBSR.MS.ID.002687.

Abhishek, V., Khan, M. A., Chopda, V., Bhagat, A., & Verma, P. (2023). Prediction of Brain Tumor Using Machine Learning. Journal of Emerging Technologies and Innovative Research, 10(4), 321-324.