Contents

[1 Chapter 1 Introduction 2](#_Toc174263256)

[1.1 Background of the Study 2](#_Toc174263257)

[1.2 Problem Statement 3](#_Toc174263258)

[1.3 Significance of the Study 4](#_Toc174263259)

[1.4 Scope of Study 4](#_Toc174263260)

[2 Chapter 02 Literature Review 6](#_Toc174263261)

[3 Chapter 03 Research Methodology 8](#_Toc174263262)

[3.1 Proposed Methodology 8](#_Toc174263263)

[3.2 Data Collection 8](#_Toc174263264)

[3.3 Data Analysis 8](#_Toc174263265)

[3.4 Data Preprocessing 10](#_Toc174263266)

[3.5 Feature Engineering 10](#_Toc174263267)

[3.6 Model Selection 10](#_Toc174263268)

[3.6.1 VGG16 and VGG19 11](#_Toc174263269)

[3.6.2 InceptionV3 11](#_Toc174263270)

[3.6.3 Custom CNN 12](#_Toc174263271)

[3.7 Model Implementation 13](#_Toc174263272)

[3.8 Training and Evaluation 13](#_Toc174263273)

[3.9 Performance Metrics 13](#_Toc174263274)

[3.9.1 Precision 13](#_Toc174263275)

[3.9.2 Recall 14](#_Toc174263276)

[3.9.3 F1-Score 14](#_Toc174263277)

[3.10 Experimental Setup 15](#_Toc174263278)

[4 Chapter 04 Result and Discussion 16](#_Toc174263279)

[4.1 Class wise results of Applied deep learning model 16](#_Toc174263280)

[4.2 Overall Results of Applied Deep Learning Models 17](#_Toc174263281)

[4.3 Confusion Matrix and History graph 18](#_Toc174263282)

[4.4 Discussion and Future Work 21](#_Toc174263283)

[5 References 22](#_Toc174263284)

**List of Figures**

[Figure 3.1 Diagram of our proposed methodology 9](#_Toc174263381)

[Figure 3.2 Data Distribution across Each class 10](#_Toc174263382)

[Figure 3.3 Selected sample of MRI images 11](#_Toc174263383)

[Figure 4.1 Visual Presentation of the results 19](#_Toc174263384)

[Figure 4.2 History graph of Deep learning models 20](#_Toc174263385)

[Figure 4.3 Confusion Matrix of Deep learning models 21](#_Toc174263386)

**List of Tables**

[Table 4.1 class wise result of deep learning model 18](#_Toc174263457)

[Table 4.2 overall result of Deep learning models 19](#_Toc174263458)

# Chapter 1 Introduction

## Background of the Study

Most of the patients suffering from brain tumor conditions die within five years after the detection of their tumors. Brain tumors are at times referred to as intracranial neoplasms which are abnormal growths in the cranial cavity that originate from various types of cells like neurons, glia or support cells, blood vessels, leptomembranes and even ectopic tissues. On the other hand, brain metastases are those tumors in the brain that have developed from cancer in other organs like lung cancer or breast cancer. Brain tumors could either be benign or malignant, meaning non-life threatening or life threatening respectively. While benign brain tumors may exist forever without creating any problem for its owner other than psychological discomfort, malignant types still pose a serious risk to survival due to the aggressive nature in which they spread.

Brain tumors are far one of the most severe and deadly conditions that affect our central nervous systems alongside other system functions. These tumors have been known to have different behaviors that range from fast-growing and fatal ones to slow-growing and non-harmful ones respectively . The classification of the brain neoplasms is a crucial one in deciding on what sort of treatment course is required, which includes surgery, irradiation, cytotoxic agents, combined chemotherapy-and-radiotherapy regimens, etc. Early diagnosis for the patients in this condition has to be made by the doctors specialized in it, as the type of the tumor must influence its treatment and prognosis.

Precise magnetic resonance imaging, as one of the latest non-invasive techniques of imaging, helps in guiding diagnosis of the brain tumor to paint a picture that lucidly depicts the degree of damage caused by such conditions. Images produced by iMRI give relevant information to different parts of the brain, enabling doctors to see something abnormal, for example, the existence of tumors in those areas, just before surgery. This includes information of their location and size. However, at times it can get quite difficult to interpret these scans manually due to the complexity involved when one looks at the brain structures together with subtle differences that exist between various tumor types, leading to the need for the development of automated systems aimed at achieving precise classifications of human brain cancer cells.

Deep learning refers to that part of artificial intelligence concerned with the construction and training of models to perform particular tasks, for example, visual recognition or speech recognition. Thus, the latest development in AI was beneficial for the automation of brain tumors identification processes on MRI. When experts see humanity and AI together, it just seems that all of them strive for excellence, but rather, their methods differ considerably. We will examine some important machine learning papers related to brain tumor imaging by assessing deep learning models used in the classification process.

It has been realized that breast cancer leads to a high rate of mortality among women, mainly in least developed countries. This has led to great interest among researchers in its diagnosis using mammography. The automated filing establishes what is referred to as CAD systems, which are designed particularly for diagnosis.

The current study is the application of deep learning models in the classification of brain tumors using MRI images. It points at the efficiency of a number of existing deep learning architectures, including VGG19, CNN, Inception, and VGG16, when applied to identify and classify brain tumors into four different classes include glioma, meningioma, pituitary tumors, and no presence of a tumor. Different performance benchmarking on these models will be done, with the aim of improving the reliability and accuracy of the tools developed in the future, for the diagnosis of brain tumors and hence clinical outcomes for patients.

It is, therefore, the aim of this chapter to address the critical role that brain tumor classification plays in the management of patients, the challenges associated with manual diagnosis, and the potential of deep learning models in enhancing diagnostic accuracy. The next few sections will go specifically into the research questions, objectives, and significance of the study in greater detail, developing the background for the detailed analysis and evaluation of the deep learning models applied in this research.

## Problem Statement

The objective of classification of brain tumors from MRI images has been one of the most paramount challenges of medical diagnosis, where current traditional ways are often slow, subjective, and error-prone. Despite outstanding progress in deep learning, there still remains a huge gap between knowledge of the comparative effectiveness of different models within this domain. The available diagnostic methods, however, not been very effective in the differentiation of the tumor types, which in the end could cause the making of false diagnoses and less-than-ideal treatment outcomes. The purpose of the present study is to indicate the way forward in improving the reliability and validity of brain tumor classification through an assessment of deep learning models such as VGG19, CNN, Inception, and VGG16, so as to suggest the most efficacious way to achieve an improvement in diagnostic accuracy and quality in patient care.

**1.3 Study’s Aims and Objectives**

The major aim is to construct deep learning models for precise stem classification of brain tumors from MRI images. Furthermore, this research should focus on following objectives:

1. Collecting and pre-processing MRI images associated with brain tumors to guarantee that high-quality input data will be available for model training.

2. Carrying out feature engineering on the image dataset so as to improve the models’ discrimination ability with regard to different brain tumor types.

3. Recognize a brain tumor in an MRI image through training and testing various deep learning models such as CNNs, VGG19, Inception and VGG16.

4. Compare all models with respect to accuracy, precision, recall and F1 score in order to determine which is most reliable when it comes to having accurate classification of brain tumors.

**1.4 Research Questions**

Perhaps these are the main research questions that this study aims to answer:

1. Will MRI images be classified as brain tumors with the use of deep learning models?

2. How correct and reliable will the brain tumor detection from MRI images become after using the deep learning models?

3. How accurate, precise, and reliable are these different deep learning models like VGG19, CNN, Inception, and VGG16 in detecting the brain tumors by their recall and F1-score?

4. Which of the proposed deep models has the highest performance and thus could be recommended for accurate and robust classification of brain tumors?

## Significance of the Study

The importance of the present study is that deep learning models can be successfully used to advance the field of medical imaging, in particular, the diagnosis of brain tumors by accurate and efficient classification. Brain tumors, if detected early and classified with accuracy, would significantly help in enhancing treatment planning, patient outcome, and overall survival rates. Through the comparison of different deep learning models, this study will add to the present body of academic knowledge regarding model capabilities and provide practical insights for clinicians and radiologists in deciding which one is the best tool in diagnosing brain tumors. Such findings may result in a more reliable diagnosis, hence decreasing the misdiagnosis rate so that intervention on patients diagnosed with brain tumors is timely, with the avoidance of ineffective or harmful treatments.

## Scope of Study

It focuses on the classification of brain tumors by deep learning models that are applied to MRI images. In particular, this study will focus on the collection, preprocessing, and feature engineering of MRI datasets and then apply and evaluate four deep learning models such as VGG19, CNN, Inception, and VGG16. The study focus on classifying four brain tumor types including glioma, meningioma, no tumor, and pituitary tumors. It does not cover the other imaging modalities such as CT scans, and neither has it talked about any other types of brain abnormalities and tumors except for the ones in the four groups. Clinical application of the models or the real-time implementation was also not the part of the study, but rather the comparison is made in the dataset environment only.

**1.7 Thesis Structure**

The thesis is divided into chapters, which represent the timeline of research done. Chapter 1 presents the background of the study, states the problem, objectives, research questions, significance, scope, and overview of the thesis structure. Literature Review, Chapter 2, reviews the literature in place on brain tumor classification, machine learning, and deep learning models, showing the gaps that this work is trying to fill. Specifically, on the methodology, Chapter 3 elaborates on the research design, which goes further to explain processes involved in data collection, preprocessing, feature engineering, and implementation of deep learning models used in this study. Chapter 4 presents the experimental results obtained from applying VGG19, CNN, Inception, and VGG16 models to the MRI dataset, majorly tending on metrics such as accuracy, precision, recall, and F1-scores. Chapter four also discuss Future Work, summary of key findings of the research, the contributions made, and future investigation areas.

# Chapter 02 Literature Review

A challenge to accurately diagnosing brain tumor types with human mistakes being common and often taking time, as stated in this analysis (Abiwinanda et al., 2019). An attempt to identify the three main types of brain tumors; Glioma, Meningioma and Pituitary is what this research emphasizes on using Convolutional Neural Network (CNN). The way it was structured in terms of architecture was simple with one convolution layer; max pooling layer flattens out before entering a fully connected hidden layer. Training was undertaken utilizing publicly available 3064 T-1 weighted CE-MRI images from figshare. Despite its simplicity, its achieved training accuracy stood at 98.51% while that for validation attained is 84.19%. These results compete with region-based segmentation algorithms that are more complex which report accuracies ranging between 71.39% and 94.68% for the same dataset.

At advanced stages brain tumors are one of the most violent diseases hence making treatment planning an essential to help improve patient’s quality of life because they greatly reduce their lifespan. Commonly CIMGH|UScle1 gets utilized in evaluating different types tumors found within distinct organs. This kind of research (Seetha & Selvakumar Raja, 2018) centered upon detecting brain tumors via MRI images whereupon numerous issues emanate from gigantic amounts of data realized through MRI scans.

Magnetic resonance imaging (MRI) is applied in detecting brain tumors, though manual analysis of big data and the diversity of tumor types are time-consuming and lead to errors. The aim of the current research work, hence, requires an automated computer-aided diagnosis application. There have however been observed strong improvements in recent image classification notably with deep convolutional neural networks (CNNs) though. This paper simply presents the new CNN model, dealing with the brain tumor classification with the basic support of deep layers in the classification of MRI brain images. We have tested our model on three datasets, and from the experimental results, it can be concluded that our approach outperforms many existing methods.

This paper utilizes the automated procedure for detecting brain tumors in MRI proposed in Aamir et al. (2022). First, the MRI image is pre-processed for better visualization. Two types of pre-trained feature extraction models are applied to the images. The extracted features, using a feature extraction process, produced a hybrid feature vector using a partial least squares (PLS) method. Lastly, agglomerative clustering identifies the top locations that contain a tumor. These locations are later resized based on this resizing; the final classification is done using a head network. The classification accuracy obtained on the dataset from the proposed method was 98.95%.

This work enhances the classification for brain tumors from MRI by AI algorithms, in particular by Convolutional Neural Networks and deep learning. In this approach, a brain tumor dataset has been trained with five fine-tuned models: Xception, ResNet50, InceptionV3, VGG16, and MobileNet. The highest F1-score reached was 98.75% with the Xception model. These high accuracies increase early tumor detection by a great margin and help in effective treatment planning, thus preventing severe physical side effects.

This research is therefore motivated by the need for an effective Brain tumor detection and classification strategy since it has a direct influence on the outcome of the treatment. DeepTumorNet is introduced as a hybrid deep learning model proposed for classifying three types of BTs: glioma, meningioma, and pituitary tumor. It is based on the GoogLeNet CNN architecture, modified to drop the last five layers before adding fifteen new layers. Leaky ReLU was added as an activation function to improve feature extraction. The model went well on a freely available dataset: its accuracy stood at 99.67%, precision at 99.6%, recall at 100%, with an F1-score of 99.66%. These results show improved performance over the rest of the state-of-the-art models, including AlexNet, ResNet50, Darknet53, ShuffleNet, GoogLeNet, SqueezeNet, ResNet101, XceptionNet, and MobileNetv2.

# Chapter 03 Research Methodology

## Proposed Methodology

The research design will be on the classification of brain tumors using deep learning models on Magnetic Resource Imaging images. This comprises the collection and preprocessing of MRI datasets, development, training, and finally, evaluation of these four deep learning models: VGG16, VGG19, CNN, and InceptionV3. These models are trained against a labeled dataset for validating their accuracy in classifying four classes of brain tumors such as glioma, meningioma, no tumor, and pituitary tumor. This means that the general objective of this work is to establish the most accurate model with regard to brain tumor classification as present in figure 3.1.

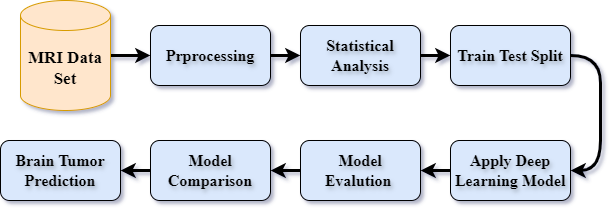


Figure . Diagram of our proposed methodology

## Data Collection

In this work, the dataset to be used is the "Brain Tumor MRI Dataset" from Kaggle (*Brain Tumor MRI Dataset*, n.d.).The dataset is split into training and testing directories, with images categorized into four classes such as glioma, meningioma, no tumor, and pituitary.

## Data Analysis

In this section the dataset was thoroughly analyzed to tell the way images were distributed out across the training set and the test set. The training set consists of 5,712 images, while the evaluation set has 1,311. A class-wise distribution was also performed to understand the representation of each tumor type in the dataset. The glioma class consists of 1,621 images, pituitary tumor class has 1,757 photos, meningioma class includes 1,645 and belonging to nontumor includes 2,000 pieces. This analysis showcases how balanced representations among different tumor types are important. Figure 3.2 shows image distributions among different classes which emphasizes shapes within the dataset and accentuates need for balanced training aimed at burning stamina into machine learning performance of deep convolutional networks.

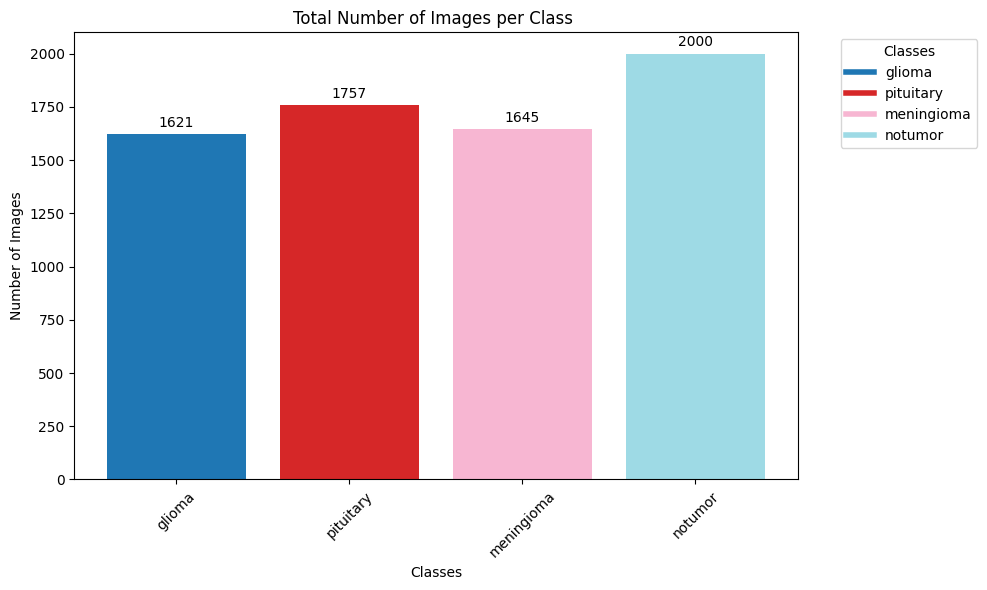


Figure . Data Distribution across Each class

For better clarification on the MRI images, selected samples were drawn from each class and then examined. This process has been vital to dispel any confusion regarding early picture identification characteristics of various types of brain tumors and nontumor category. By plotting these images, researchers can observe the visual distinctions between glioma, pituitary tumors, meningioma, and nontumor MRI scans as shown in the Figure . Consequently, this step not only makes one familiarize with the dataset but also plays an important role in the initial assessment of the challenges that the classification models will face. Understanding the visual features of the images helps in refining preprocessing steps and model design to better capture the nuances in the data, ultimately leading to more accurate and reliable classification outcomes.

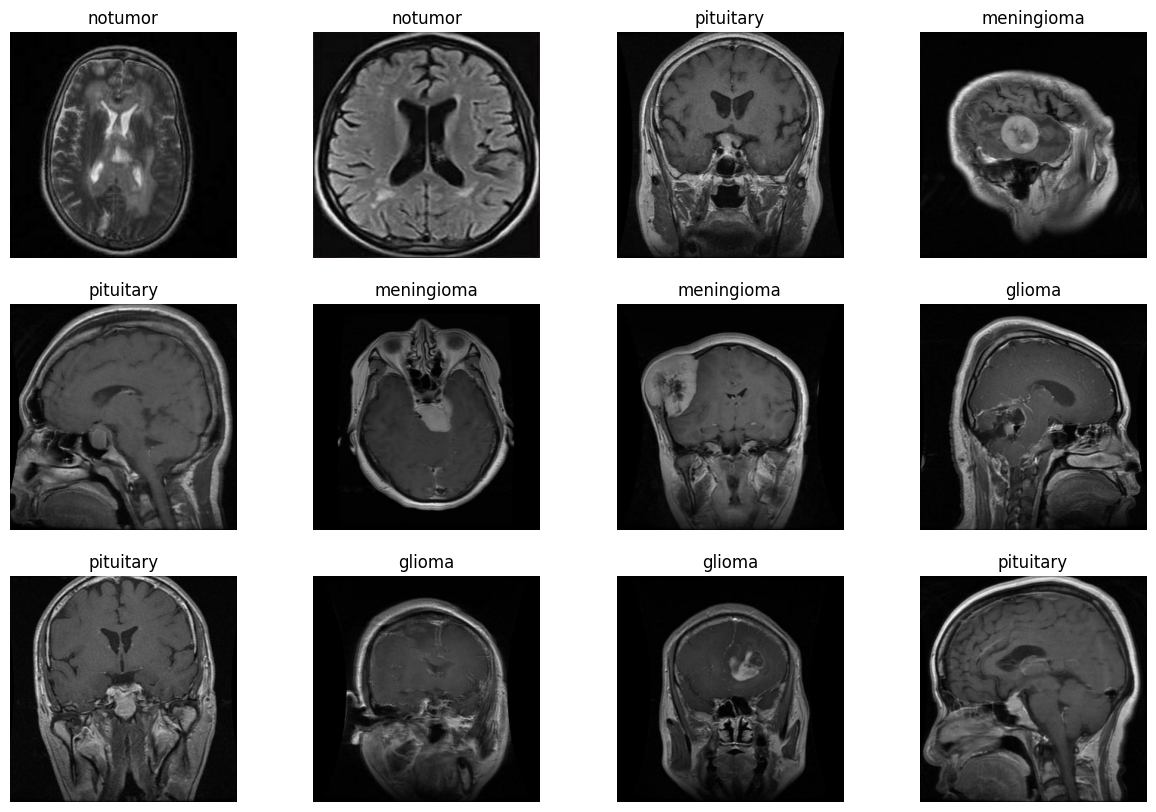


Figure . Selected sample of MRI images

## Data Preprocessing

The preprocessing step is very important to have the MRI images properly formatted and ready for input into the model. All the images were resized to 224×224 pixels to be consistent across all models.I applied data augmentation techniques using the ImageDataGenerator class to artificially increase the size of the training dataset by generating variants of images. The augmentation process involved random rotation, width and height shifts, shearing, zooming, and horizontal flipping. These techniques help in prevention from overfitting and provide better generalization ability in models.

## Feature Engineering

In this work, Feature Engineering was conducted using pre-trained models like VGG16, VGG19, and InceptionV3 for brain tumors classification. These models were used as feature extractors, wherein the convolutional layers were frozen, and new fully connected layers were stacked to fine tune the models for the task of brain tumor classification.

## Model Selection

Four deep learning models VGG16, VGG19, InceptionV3, and custom built Convolutional Neural Network are build in this study. These models have been chosen for the study based on prior knowledge about their performance with respect to image classification problems, especially in medical imaging. Each model has its own strengths, and all of their architectures are designed in a way that they can capture the minute features present in the MRI images of brain tumors.

### VGG16 and VGG19

VGG16 and VGG19 are very famous models in the Visual Geometry Group. Characterized deep architectures have multiple convolutional layers combined with fully connected layers at the end. The most important innovation of the VGG model is the use of very small filters across the network. it will then capture the fine details of images while keeping computational efficiency per requirements. Below is the general structure of a VGG model represented by the following equations:

**Convolutional Layers:** Each convolutional layer applies a filter.

where ∗ denotes the convolution operation, 𝑏 is the bias term, and 𝑓 is the activation function, ReLU.

**Max Pooling:** Following a set of convolutional layers, max pooling is used to reduce spatial dimensions:

Where represents the local region in the feature map.

The flattened output from final convolutional layer undergoes one or more fully connected layers:

where and are the weights and biases of the fully connected layer.

**Softmax Output:** The softmax function found at last output layer produces class probabilities.

VGG16 consists of 16 layers, where there are 13 convolutional layers and 3 fully connected layers, while VGG19 has 19 layers, of which 16 are convolutional and 3 are fully connected. Pre-training is done for each model on the ImageNet dataset and fine-tuned on the Brain Tumor MRI data set to borrow advantageous features using transfer learning.

### InceptionV3

InceptionV3 is a deep convolutional network belonging to the family of Inception. The model makes use of an Inception module, which applies different sizes of convolution filters on the image in parallel and concatenates their outputs to efficiently capture the multi-scale features inherent in the image.

Mathematically, an Inception module can be explained as follows:

Parallel Convolutions A number of convolution operations are applied parallel to the same input.

where ​ are filters of different sizes (e.g., 1x1, 3x3, 5x5), and are the corresponding biases.

Concatenation: These parallel convolutions' feature maps are concatenated:

Reduction Layer: The concatenated output may be followed by a pooling or dimensionality reduction layer for controlling the computational cost.

InceptionV3 introduced a number of optimizations, including factorized convolutions, which go a long way in further lightening the computational load but maintain high accuracy. The model has more complicated architecture compared to VGG, but it captures very diverse features from MRI images.

### Custom CNN

The custom CNN will serve as a baseline to which the more complex models like VGG and Inception can be compared. It is comprised of convolutional layers, succeeded by pooling layers and then fully connected layers succeeded by a softmax output layer.

This custom CNN can be represented mathematically through the following equations:

Convolutional Layer: the convolution is performed as:

Pooling Layer: A max-pooling layer reduces the dimension of the feature map.

Dense Layer: The pooled feature map is flattened, and the flattened feature maps pass through dense layers.

Output Layer: The last dense layer applies the softmax function to give the predicted probabilities of four tumor classes.

This model, though simple, still has a good way of capturing the most crucial features and acts as a baseline to check the performance of more advanced models.

## Model Implementation

All the models implemented were done using TensorFlow and Keras. Each model from VGG16 and VGG19 had been made in a manner where pre-trained weights had been loaded while keeping the convolutional layers frozen to retain the learned features. A custom fully connected layer with units equal to 256 was added, followed by a dropout layer to prevent overfitting. It was trained by a CNN implemented from scratch with three convolutional layers, which were followed by max-pooling layers and topped with a dense layer for classification. The third model used was the InceptionV3, which had a pre-trained base with only the top layers replaced with a global average pooling layer, followed by a dense layer and a dropout layer.

Every model was compiled with the Adam optimizer, a learning rate of 0.0001, and trained with the categorical cross-entropy loss. They were trained for 50 epochs with batch sizes of 32, and their performance was validated using a separate validation set.

## Training and Evaluation

The trained models used the augmented training dataset, and the performance was tested on the test dataset. During training, performance was monitored on the training and validation accuracy and loss. The performance of the trained models was finally assessed in terms of precision, recall, the F1-score, and the confusion matrix for their classification performance.

Finally, in the process, some predictions over the test dataset will be generated and compared against the hold-out true labels. This would give a detailed account of how the models perform through a confusion matrix and a classification report across various classes. These measures are used in estimating the best model based on accuracy, precision, recall, and F1-score.

## Performance Metrics

To evaluate the performance of the deep learning models applied in this study, we employed several key metrics: Precision, Recall, and F1-Score. These metrics are crucial for assessing the effectiveness of the models in classifying brain tumor images into the correct categories.

### Precision

**The precision measures the accuracy of the positive predictions made by the model. It indicates how many of the instances predicted as positive by the model are actually positive. In situations where high costs are associated with false positives, precision is particularly valuable.**

Mathematically, precision is defined as:

​

where:

* TP are the people who everyone thought were nice until they did something really bad; theseare also examples of correctly predicted positives.
* Falses positives refer specifically to those situations where the computer says it’s a cat but it isn’t a cat at all.

### Recall

Recall, or Sensitivity or True Positive Rate, is a measure of how well the model can find all the relevant instances with a positive value. The ratio of the true positives to the actual number of positives in the dataset is what it is about. This is important when you want to maximize your recall and get as many true positives as possible.

Recall is define as:

where:

* TP (True Positives) is a measure of how many positive instances were predicted correctly.
* FNs (false negatives) refer to the actual positive cases that were wrongly predicted by the model to be negative.

High recall indicates that the model successfully identifies most of the positive instances, with few false negatives.

### F1-Score

The F1-score is the measure that computes the harmonic mean of precision and recall to achieve a balance. This is very important when there is unbalanced class distribution or when precision and recall are both critical for the purpose in question. The F1-Score also considers false positives as well as false negatives in such cases that one should not bias towards either precision or recall in model performance.

The F1-Score is defined as:

The F1-Score ranges from 0 to 1, where 1 indicates perfect precision and recall. A high F1-Score indicates that the model performs well in both precision and recall, making it a robust metric for evaluating the overall performance of classification models.

## Experimental Setup

Experiments were conducted on a system with enough power and having a powerful GPU for better training performance. All models were implemented in Python, using the TensorFlow and Keras libraries. Matplotlib and Seaborn were used for data visualization and analysis.

# Chapter 04 Result and Discussion

In this particular chapter, we shall delve into the findings of deep learning models used in the study. Each model has been subjected to extensive evaluations through plotting training histories and confusion matrices. Insights into different aspects of how accurate they were or not at different training as well as testing stages are provided in these visualizations. These results will be explored more deeply with an aim to provide a critical analysis that would enable us examine how well each individual setup is able to classify brain tumor images accurately. Such analysis will reveal where there are strengths and weaknesses and therefore facilitate future advancements.

## Class wise results of Applied deep learning model

Among the four models developed to categorize brain cancers, different outcomes were noted in terms of performance. The VGG19 model had an exceptional overall performance especially in classifying no-tumor cases which were rewarded with a precision of 0.99, recall 0.99 and F1-score 0.99. It was also good at identifying pituitary tumors (precision 0.94, recall 0.99 and F1-score 0.96). For gliomas VGG19 produced precise measurements that had 0.92 precision, 0.89 recall and F1-score was rated as 0.91 respectively. For the meningioma category it had precision rates of 0.90 with recalls accounting for about 87% resulting to an F1-score of 0.89 as present in Table 4.1.

In addition to that, VGG16 model exhibited high levels of performance with regard to precision and recall especially when distinguishing no tumor cases at all which earned it 0.98 precision; 1% less than the first number but almost equal to 99% observed in both recall rate and F-1 score as well respectively. When compared with other models such as in diagnosing pituitary tumors’ category this engine retained good characteristics such as countenance pointing out to values around 0.91 for maximum correctness along with even metrics represented through recall system equivalents.

In general, overall effectiveness-wise the best performance was displayed by the VGG19 model mostly when identifying gliomas where correctness was measured at best at 0.92 and misrepresentation rate was seen at its greatest extent as 0.89.

Table . class wise result of deep learning model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Target Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| VGG19 | glioma | 0.92 | 0.89 | 0.91 | 300 |
| meningioma | 0.90 | 0.87 | 0.89 | 306 |
| notumar | 0.99 | 0.99 | 0.99 | 405 |
| pituitary | 0.94 | 0.99 | 0.96 | 300 |
| CNN | glioma | 0.98 | 0.69 | 0.81 | 300 |
| meningioma | 0.79 | 0.57 | 0.66 | 306 |
| notumar | 0.90 | 0.96 | 0.93 | 405 |
| pituitary | 0.67 | 1.00 | 0.80 | 300 |
| Inception Model | glioma | 0.81 | 0.39 | 0.52 | 300 |
| meningioma | 0.56 | 0.25 | 0.34 | 306 |
| notumar | 0.72 | 0.80 | 0.76 | 405 |
| pituitary | 0.47 | 0.92 | 0.62 | 300 |
| VGG16 | glioma | 0.97 | 0.83 | 0.89 | 300 |
| meningioma | 0.84 | 0.88 | 0.86 | 306 |
| notumar | 0.98 | 0.99 | 0.98 | 405 |
| pituitary | 0.91 | 0.99 | 0.95 | 300 |

## Overall Results of Applied Deep Learning Models

The overall performance of the used deep learning models in the classification of brain tumors reveals their effectiveness variances. Among them, VGG19 attained highest accuracy as 0.94, with precision, recall and F1-score all being 0.94. This consistency across metrics indicates VGG19’s robustness and dependability for classifying brain tumors.

The VGG16 model also did well attaining an accuracy of 0.93 and precision, recall and F1-score of 0.92 indicating strong predictive ability as show in Table 4.2.

In comparison, the custom CNN model was less effective, with test accuracy at 0.81 precisely it had 0.83 in share of true positive rates while its sensitivity and F1-score were both 0.80. It showed some reasonable reliability despite having lower performance as show in Figure 4.1.

In contrast, Inception had poorest results; an accuracy of 0.60 precision at 0.64; sensitivity at 0.59; F1-score being 0.56. The implication here is that it has serious limitations in diagnosis indicating a high likelihood of missing a diagnosis altogether.

The VGG19 and VGG16 outperform others in terms of performance they become choices when it comes to effective brain tumor classification

Table . overall result of Deep learning models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| VGG19 | 0.94 | 0.94 | 0.94 | 0.94 |
| CNN | 0.81 | 0.83 | 0.80 | 0.80 |
| Inception Model | 0.60 | 0.64 | 0.59 | 0.56 |
| VGG16 | 0.93 | 0.92 | 0.92 | 0.92 |

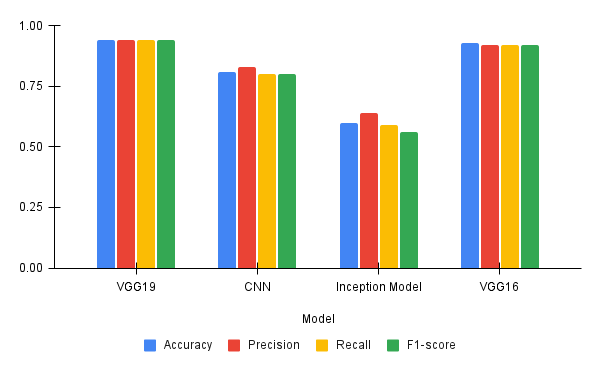


Figure . Visual Presentation of the results

## Confusion Matrix and History graph

In this section, we will analyze the confusion matrix and the training history graphs generated from the results of the deep learning models. These insights into model performance help us understand how well they are able to tell classes apart as well as track their learning journey over time.

The VGG16 model's training history graph shows a steady increase in both training and validation accuracy, with little fluctuation in the curves. This smooth trajectory indicates that the model is effectively generalizing to the data. In parallel, the training and validation loss curves decrease gradually, reinforcing the model’s strong generalization capability. In contrast, other deep learning models applied in this study present a more erratic pattern on their history curves, characterized by wild swings from one extreme to another as show in Figure 4.2. These fluctuations suggest that these models had difficulties achieving consistent performance and did not reach the same level of generalization as did VGG16.

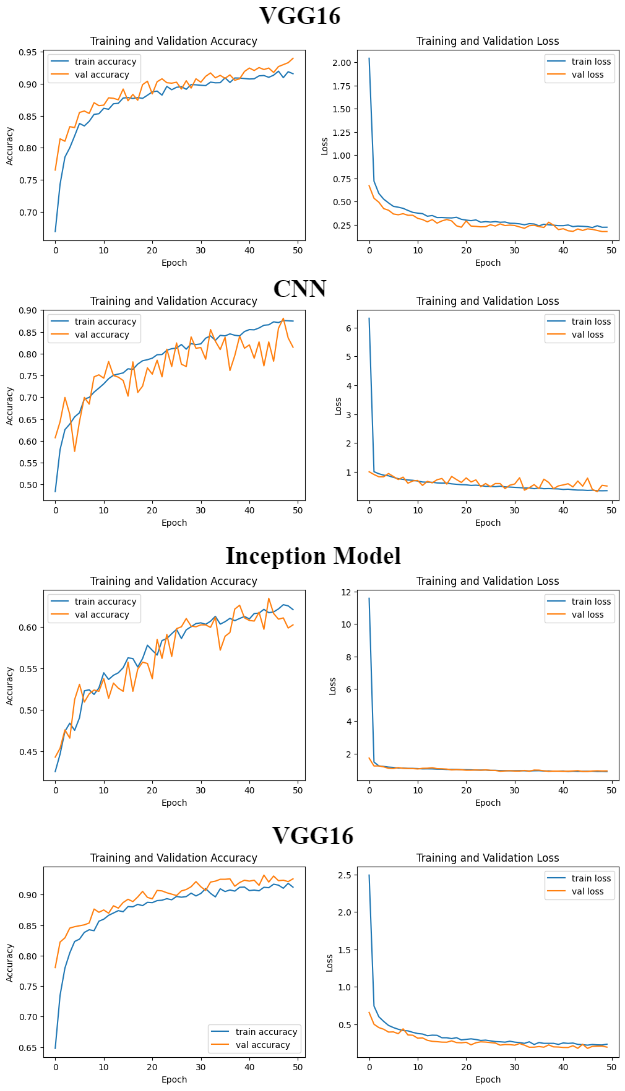


Figure . History graph of Deep learning models

The confusion matrix is a valuable tool for assessing the performance of deep learning models by providing detailed information on model predictions for various classes. This includes true positives, true negatives, false positives and false negatives for each class hence giving an overall summary of the model’s accuracy. For example, it can be observed that VGG19 model has more true positives than its false negatives which indicates high predictive abilities. On the other hand, Inception model has higher rates of false positive and negative especially in difficult classes which indicates poor performance and misclassification problems. In conclusion therefore, confusion matrix helps to evaluate the strengths and weaknesses in brain tumor classification for different models.

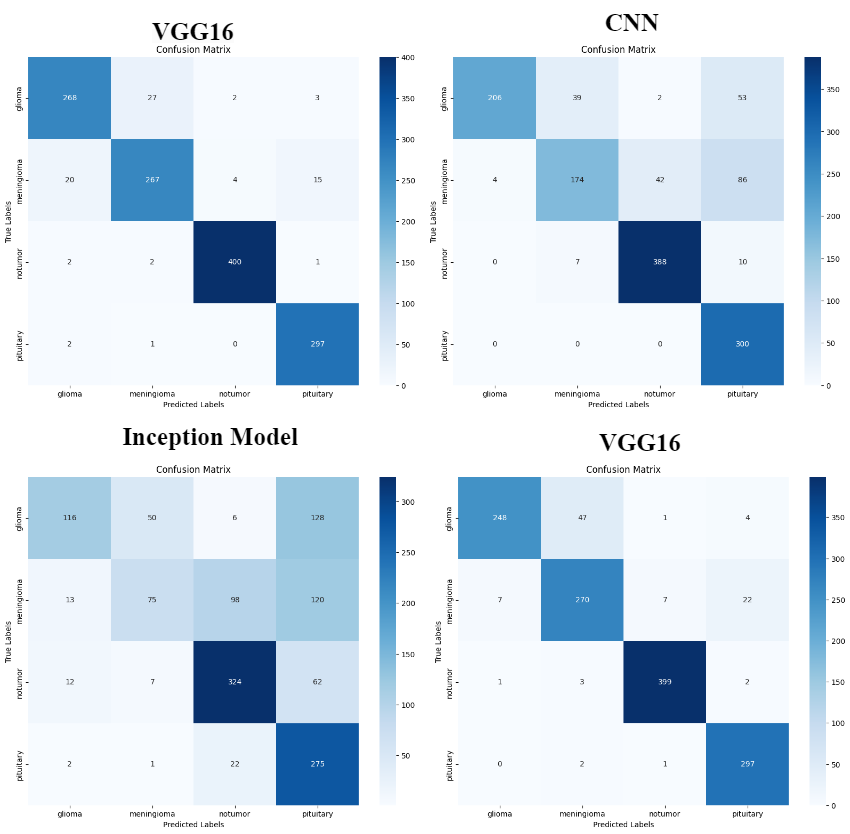


Figure . Confusion Matrix of Deep learning models

## Discussion and Future Work

The application of various deep learning models for the classification of brain tumours has led to different levels of effectiveness that were observed in this study. For all tumour classes, VGG19 was found as the most dependable model with respect to accuracy, precision, recall and F1-scores. This consistent performance demonstrates its stability and possibility for use in medicine. Another good alternative for the precise identification of tumours is VGG16 since it also performed quite well but left behind VGG19 rank-wise. On the other hand, some limitations were exhibited by custom CNN and Inception models particularly due to the Inception model which struggled with precise diagnosis rates at higher levels than expected values hence producing lots of false negatives (missed diagnoses). Therefore, although deep learning can be a powerful tool, it is evident that there exists much diversity in terms of performance among different architectures as well as training methods.

The future work should based on these models’ improvement especially through combining VGG19 and 16’s benefits while taking care of weaknesses noted in CNNs or Inception structures. Also, model generalization and robustness can also be improved through additional data expansion with greater diversity and quantity. In addition to this, research might look into how transfer learning and attention mechanism integration would improve classification accuracy plus reliability in clinical environment.

# References

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

Abiwinanda, N., Hanif, M., Hesaputra, S.T., Handayani, A., Mengko, T.R., 2019. Brain Tumor Classification Using Convolutional Neural Network. IFMBE Proc 68, 183–189. https://doi.org/10.1007/978-981-10-9035-6\_33

Ayadi, W., Elhamzi, W., Charfi, I., Atri, M., 2021. Deep CNN for Brain Tumor Classification. Neural Process Lett 53, 671–700. https://doi.org/10.1007/S11063-020-10398-2/TABLES/29

Saleh, A., Sukaik, R., Abu-Naser, S.S., 2020. Brain tumor classification using deep learning. Proceedings - 2020 International Conference on Assistive and Rehabilitation Technologies, iCareTech 2020 131–136. https://doi.org/10.1109/ICARETECH49914.2020.00032

Seetha, J., Selvakumar Raja, S., 2018. Brain Tumor Classification Using Convolutional Neural Networks. Biomedical & Pharmacology Journal 11, 1457–1461. https://doi.org/10.13005/bpj/1511