# Brain Tumour MRI Image Classification

Dharun Shoban A/L Ganesh
*Faculty of Computing and Informatics*
*Multimedia University*
Cyberjaya
1211300737@student.mmu.edu.my

Kolbe Alex A/L Soosai Nathan
*Faculty of Computing and Informatics*
*Multimedia University*
Cyberjaya
1211302223@student.mmu.edu.my

Danial bin Muhammad Firdaus Anson
*Faculty of Computing and Informatics*
*Multimedia University*
Cyberjaya
1201101577@student.mmu.edu.my

Low Zhao Pin
*Faculty of Computing and Informatics*
*Multimedia University*
Cyberjaya
1201100696@student.mmu.edu.my

***Abstract*— The experiment conducted in this study focuses on the classification of brain tumours in MRI images using Convolutional Neural Networks (CNNs). Preprocessing methods, including noise reduction, histogram equalization, thresholding, and image conversion with normalization, are applied to enhance image quality for effective model training. Four CNN models are implemented: a custom 3-layer CNN, a custom 9-layer CNN, and pre-trained models VGG16 and VGG19. Evaluation metrics such as accuracy, precision, recall, and F-score are utilized to assess model performance. Results indicate that while all models perform well in identifying tumour-free images, the VGG16 model achieves the highest overall accuracy of 89%, followed closely by the VGG19 model at 88%. The custom CNN models achieve accuracies of 87% and 85%, respectively. Additionally, further analysis reveals that the 9-layer CNN model demonstrates exceptional proficiency in identifying Glioma Tumours with an accuracy of 97%, while the 3-layer CNN exhibits notable accuracy across all classes, particularly excelling in identifying Pituitary Tumours with an accuracy of 92%. Notably, VGG16 excels in recognizing No Tumour cases with an accuracy of 96%, underscoring its efficiency in distinguishing normal brain images. These findings highlight the model-specific strengths in accurately classifying distinct tumour types, offering valuable insights for researchers and practitioners aiming to optimize model selection based on specific classification requirements in brain tumour MRI image analysis.**

**Keywords— CNN, Brain Tumour, Classification, Image Preprocessing, MRI Image, Neural Networks, VGG16, VGG19.**

## I. INTRODUCTION

85–90% of primary tumours of the central nervous system are brain tumours, which are a serious and sometimes deadly medical illness. With a comparatively low five-year survival rate of roughly 34% for males and 36% for women, approximately 11,700 new cases are diagnosed annually. The complexity and diversity of tumour characteristics present a problem in the classification of brain tumours, making manual diagnosis by radiologists a tedious and error-prone operation. The necessity for advanced diagnostic equipment is highlighted by the fact that prompt and effective treatment depends on an accurate and quick diagnosis.

The objective of this research is to help brain tumour patients' early diagnosis and effective treatment planning. By employing modern computer vision technology, our goal is to create and improve classification algorithms that can effectively identify between different kinds of brain tumours. Because of how difficult it is to handle and interpret MRI image data, this project is not only technically complex but also extremely significant, with the potential to evolve the medical diagnostics industry.

Four different Convolutional Neural Network (CNN) algorithms are implemented and compared in our research: 3-layer custom model, 9-layer custom model, VGG16 and VGG19. The first two models are developed manually, whereas the other two models are pre-trained models. The choice of these models aims to maximize classification accuracy by utilizing the advantages of each model. The study will assess both algorithms' performance thoroughly with a particular focus on improving brain tumour classification's accuracy and dependability. This will help medical professionals make better judgments and eventually lead to better patient outcomes.

This research paper is divided into five separate chapters, starting from the Introduction which introduces the motivation, problem, and overview of the research. Next, the relevant literature of the project is discussed in Related Works. Subsequently, the project framework is detailed in the Approach chapter. Subsequently, the type of experiments done, datasets utilized, and measurements used for evaluation are explained in the Experiment chapter. Finally, the paper concludes with the Conclusion chapter.

## II. LITERATURE REVIEW

In recent years, deep learning has significantly advanced the field of medical imaging, particularly in brain tumour classification. This section reviews contributions in this domain, highlighting the datasets, preprocessing methods used, CNN models implemented, and findings from each study.

Datasets utilized across these papers consists of a wide range of MRI images, from specific tumour types like glioma, meningioma, and pituitary tumours in the study by [3], to a broad collection of 3000 brain MRI images sourced from Kaggle by [6], and a combined dataset approach by [4] to enhance tumour type variability. [5] focused on a binary classification of tumour versus non-tumour images, providing a robust dataset for model training and testing.

Data preprocessing is important for enhancing model performance, with methods ranging from pixel normalization,

120-2-1322-3232-1/20/$24.00 ©2024 IEEE

image partitioning, and size normalization to advanced techniques like intensity normalization, contrast enhancement, and data augmentation through rotation, flipping, and Gaussian blur. Researchers have used many types of preprocessing methods in their projects[3] utilized pixel normalization technique using MinMax scaling. [4] utilized intensity normalization, contrast enhancement, and data augmentation techniques to preprocess the image data.

The CNN models used show a variety of choices. [3] utilized the R-CNN with VGG16 architecture, while [6] experimented with various models including CNN, MobileNet V2, Inception V3, and VGG19. [4] introduced a deep CNN with ten convolutional layers, emphasizing the depth of the network for feature extraction. [5] employed a 9-layer CNN model, optimizing for high accuracy through a comprehensive layer and filter strategy.

Findings of these papers highlight the efficiency of CNN models in medical image analysis. The precision, accuracy, and overall performance metrics like sensitivity, specificity, and F1-scores reported in these studies underline the potential of deep learning in revolutionizing the early detection and classification of brain tumours. The accuracy of the VGG16 model utilized by [3] achieved an average precision of 75.18% for glioma, 89.45% for meningioma, and 68.18% for pituitary tumours, with a mean average precision of 77.60% across all classes. Researchers in [4] achieved an accuracy of 94.74% with their ten-layer CNN model. It demonstrated high performance in classifying brain tumours with fewer misclassifications, particularly for the meningioma class. [5] achieved accuracy rates of 93% for the custom CNN, 92% for MobileNet V2, 91% for Inception V3, and 88% for VGG19 models. The highest accuracy reported by [5] at 99.74% sets a new benchmark for future research in this domain.

## III. APPROACH

In this section, the project framework, methodologies, algorithms, and techniques employed for the classification of brain tumour MRI images are outlined and described. The primary objective is to explain the steps taken to achieve accurate and reliable results in the classification task.

### *A. Data Preprocessing*

The data preprocessing pipeline is an important phase of the project, aimed at enhancing the quality and relevance of the input data for neural network training. For each tumour type in the dataset, a series of image processing techniques are applied to the MRI scans. This ensures that the neural networks are provided with clean and normalized images, which enables effective learning and classification.

#### *3.1 Noise Reduction*

In the beginning, the MRI images are read in grayscale format, and subjected to noise reduction using Gaussian Blur. This step helps to smoothen the images and eliminate unwanted objects, which contributes to a more reliable feature extraction process. Gaussian Blur is a spatial domain filtering technique widely used in image processing. It operates by convolving the image with a Gaussian kernel, effectively reducing high-frequency noise, and enhancing important structural features. The Gaussian kernel is a two-dimensional function resembling the shape of a bell curve, characterized by its center and standard deviation:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where in (1), $(x, y)$ represents the spatial coordinates in the kernel, and σ is the standard deviation which controls the spread or width of the Gaussian distribution.

#### *3.2 Histogram Equalization*

Then, the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is applied to enhance image contrast, which is important for highlighting subtle variations in the intensity levels of the brain tumour structures. CLAHE is an adaptive histogram equalization method that addresses the limitation of traditional histogram equalization by preventing the overamplification of noise. The method divides the image into small, overlapping regions called tiles and computes the histogram equalization independently for each tile. This allows CLAHE to enhance local contrast without introducing artifacts. The cumulative distribution function (CDF) of each tile is then limited to a specified contrast threshold, preventing excessive amplification of intensities. The formula for CLAHE is given by:

$$I(x,y) = Clip\left(\frac{CDF_{tile}(I(x,y)) - CDFmin}{1 - CDFmin} \times L, 0, L-1\right) \quad (2)$$

where in (2), I $(x, y)$ represents the pixel intensity in the original image, CDFtile is the cumulative distribution function of the pixel intensities within the tile, CDFmin is a user-defined lower limit for the CDF, and L is the maximum intensity level.

#### *3.3 Thresholding*

Furthermore, thresholding is applied to convert the images into binary form to simplify the subsequent analysis. Thresholding is used to segment an image into distinct regions by setting a specific intensity value as a threshold. Pixels with intensities above this threshold are assigned one value, while those below are assigned another value, which converts the images into a binary format. Mathematically, the binary thresholding operation is expressed as:

$$T(x,y) = \begin{cases} 1, & if\ I(x,y) > \text{threshold} \\ 0, & otherwise \end{cases} \quad (3)$$

where in (3), *I (x, y)* represent the pixel intensity in the original image, *CDFtile* is the cumulative distribution function of the pixel intensities within the tile, *CDFmin* is a user-defined lower limit for the CDF, and *L* is the maximum intensity level.

### 3.4 Conversion and Normalization

Additionally, the preprocessed images undergo two separate transformations, which are conversion of images to RGB format and normalization. The conversion to RGB format is to ensure compatibility with the neural network architecture, which expects input images with three color channels. This process involves replicating the single-channel grayscale image to create three identical channels. Following this, the images are resized to a 150x150 pixel dimension. Finally, normalization is applied by scaling pixel values to a range between 0 and 1. This is done by dividing each pixel value by the maximum intensity (255) in the image. The normalization operation is expressed in (4) as:

$$I_{normalized}(x,y) = \frac{I_{resized}(x,y)}{\max intensity} \quad (4)$$

where in (4), $I_{normalized}(x,y)$ the normalized intensity value of the pixel at coordinates (x, y) in an image. $I_{resized}(x,y)$ is the intensity value of the corresponding pixel after resizing the image, and max intensity signifies the maximum intensity value present in the image.

### B. Custom 3 Layer CNN

The first model used for classification is a custom CNN model with 3 layers. It extracts hierarchical features from the input MRI images, followed by a fully connected layer for classification. The model begins with an input layer that expects input images with dimensions specified by the image size. This layer sets the dimensionality for the subsequent layers to interpret the input images. Then, three convolutional layers are added to the model, each followed by a Rectified Linear Unit (ReLU) activation function. These layers learn spatial hierarchies of features from the input images. The first convolutional layer has 32 filters of size 3x3, which captures the basic patterns. The following convolutional layers continue to learn more complex features by employing the same number of filters and kernel size. After each convolutional layer, a MaxPooling layer is added to down-sample the spatial dimensions, reducing the computational load, and enhancing the network's invariance. A 2x2 MaxPooling operation is applied after each convolutional layer, which reduces the size of feature maps. After that, a flatten layer is added to transform the 3D spatial data into a 1D vector. Two dense layers are then added to perform classification. The first dense layer has 64 neurons with ReLU activation, which enables feature extraction and abstraction. The final dense layer outputs probabilities for each class using the SoftMax activation function. The number of neurons in the last layer is set to 4, which represents each class in the dataset. The model is compiled using the Adam optimizer, sparse categorical cross entropy loss function, and accuracy as the evaluation metric. The sparse categorical cross entropy is suitable for multiclass classification tasks with integer-encoded class labels.

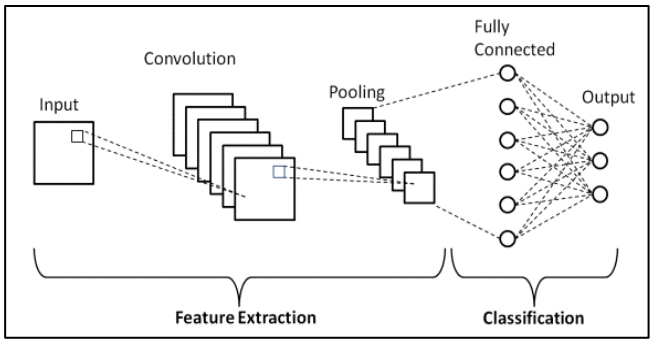



Fig. 1. A 3-Layer CNN Model Architecture Example [2].

### C. Custom 9 Layer CNN

The second model used for classification is a custom CNN model with 9 layers, designed to extract complex hierarchical features from the input images and enhance the classification abilities. The model is built upon the structure of the 3-layer model, integrating additional convolutional and pooling layers for increased complexity. The model initiation includes an input layer expecting images with dimensions specified by the image size, which allows the subsequent layers to interpret the input. The architecture then features four pairs of convolutional and max-pooling layers, each contributing to the extraction of spatial hierarchies of features. The convolutional layers utilize 32, 64, 128, and 256 filters successively, with a consistent kernel size of 3x3. Max-pooling operations follow each convolutional layer to down-sample spatial dimensions, reducing computational load and enhancing the invariance. Following these layers, a flattened layer transforms the 3D spatial data into a 1D vector, and two densely connected layers with 512 and 256 neurons respectively enable further feature extraction. The final dense layer outputs probabilities for each class using the SoftMax activation function, and the model is compiled using the Adam optimizer, sparse categorical cross entropy loss function, and accuracy as the evaluation metric.

### D. VGG16 CNN

The third model in our classification task is based on the VGG16 architecture, a well-known convolutional neural network model. The VGG16 model is pre-trained on large datasets, and its convolutional layers capture complex features from images, making it a powerful choice for image classification tasks. The VGG16 model is initialized with its input shape matching the specified image size, and the convolutional layers are frozen to retain the pre-trained weights. A new model is then constructed by adding a flattened layer to transform the 3D spatial data into a 1D vector. Two dense layers with 512 and 4 neurons respectively, are appended to perform classification, and the SoftMax activation function is applied to the final layer for probability output. The entire VGG16-based model is compiled using the Adam optimizer, sparse categorical cross entropy loss function, and accuracy as the evaluation metric. Subsequently, the model is trained on the provided training set with the specified number of epochs and batch size.

Evaluation on the test set follows the training phase, and the accuracy of the VGG16-based model is printed. The VGG16 model, with its ability to capture intricate image features, contributes to the overall objective of accurate brain tumour classification in MRI images.

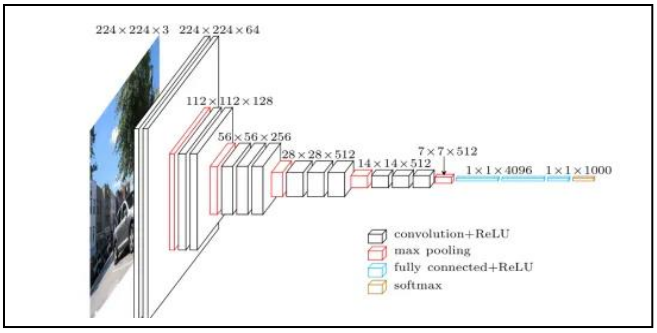



Fig. 2. A VGG16 CNN Model Architecture Example [1].

### E. VGG19 CNN

The fourth model used for our classification project is based on the VGG19 architecture, which is a deeper variant of VGG16 with 19 layers. Like the VGG16 model, VGG19 is pre-trained on large datasets and possesses an enhanced capacity for feature extraction. The VGG19 model is initialized with its input shape matching the specified image size, and the convolutional layers are frozen to retain the pre-trained weights. A new model is then constructed by adding a flattened layer to transform the 3D spatial data into a 1D vector. Two dense layers with 64 and 4 neurons, respectively, are appended to perform classification, and the SoftMax activation function is applied to the final layer for probability output. The entire VGG19-based model is compiled using the Adam optimizer, sparse categorical cross entropy loss function, and accuracy as the evaluation metric. Subsequently, the model is trained on the provided training set with the specified number of epochs and batch size. Evaluation on the test set follows the training phase, and the accuracy of the VGG19-based model is printed. The VGG19 model, with its increased depth, aims to capture even more complex features in brain tumour MRI images, contributing to the overall goal of accurate and reliable classification.

## IV. EXPERIMENT

In this section, the datasets used, type of experiments done, evaluation metrics utilized, and the results of the experiments are outlined and described. The primary objective is to explain results of the classification models and compare their performances. The results and comparisons are also visualized using plots and graphs, as well as utilizing confusion matrices to display each models' metric scores.

### A. Type of Experiments

In this research project, two distinct experiments were conducted to evaluate the performance of various classification models in brain tumour MRI image analysis. The first one is to determine which out of the four CNN models produces the overall best results in classifying brain tumour MRI images according to their respective tumour types. The second experiment done is to determine which model produces the best results for each distinct tumour category.

### B. Datasets Used

In our investigation, we will utilise the "Brain Tumour Classification (MRI)" dataset, obtained from Kaggle. This dataset is useful in developing an algorithm for classifying brain tumours through MRI scans, summarising the complexity and variance found in real-world diagnostic imaging. The main dataset consists of two sub datasets: Training and Testing sets. Both the sub datasets include a diverse set of images classified into four categories: glioma tumour, meningioma tumour, no tumour, and pituitary tumour. The training set contains a total of 2870 images, whereas the testing set contains a total of 394 images. The overall dataset contains a total of 3264 images. This extensive collection of images enables us to conduct an efficient training and validation process for developing an accurate brain tumour classification model.

TABLE I. NUMBER OF IMAGES IN EACH CLASS

| Classes | Training Set | Testing Set |
|---|---|---|
| Glioma Tumour | 826 | 100 |
| Meningioma Tumour | 822 | 115 |
| No Tumour | 395 | 105 |
| Pituitary Tumour | 827 | 74 |
| Set Total | 2870 | 394 |
| Dataset Total | 3264 | |

The first images of each tumour type present in the training set are displayed for reference:

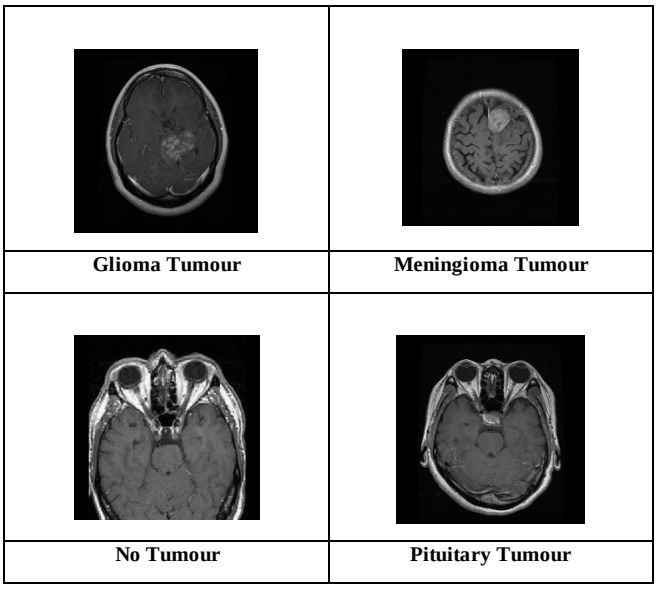



Fig. 3. First Image of Each Tumour Type in Training Set.

To visualize the number of images in each class of both the training set and the testing set, bar charts are generated with each bar representing a certain class contained in the dataset:

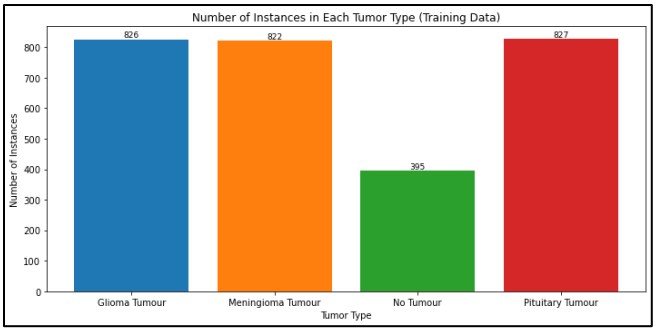



Fig. 4. Number of Images of Each Class in Training Set.

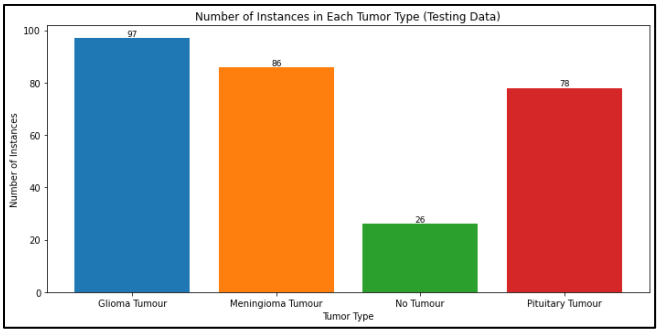



Fig. 5. Number of Images of Each Class in Training Set.

## C. Evaluation Metrics

A confusion matrix of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) can be visualized from the classifications done. The confusion matrix provides a breakdown of predictions for each class, allowing us to derive several important evaluation metrics. Based on the current research, we have selected four types of evaluation metrics that can be computed from the confusion matrix: accuracy, precision, recall, and F-score. The equations used to calculate these evaluation metrics are as follows:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F-Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (7)$$

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (8)$$

These formulas are metrics commonly used to evaluate the performance of classification models. Precision (5) measures the accuracy of positive predictions by calculating the ratio of true positives (correctly predicted positives) to the sum of true positives and false positives (incorrectly predicted as positive). Recall (6) measures the model's ability to identify all relevant instances by dividing true positives by the sum of true positives and false negatives (instances missed by the model). The F-Score (7) combines precision and recall into a single metric, providing a balanced assessment of a model's performance. Finally, Accuracy (8) quantifies the overall correctness of predictions by considering true positives, true negatives (correctly predicted negatives), false positives, and false negatives. These metrics are crucial for assessing the effectiveness and reliability of machine learning models, particularly in tasks where the balance between false positives and false negatives is crucial.

## D. Experimental Results

The results of each classification model are visualized and compared using confusion matrices, metric score tables, and graphs.

### 4.1 Overall Model Performance Comparison

TABLE II. EVALUATION METRICS TABLE FOR EACH MODEL

| Model | Accuracy | Precision | Recall | F-Score |
|---|---|---|---|---|
| 3 Layer CNN | 0.87 | 0.88 | 0.87 | 0.87 |
| 9 Layer CNN | 0.85 | 0.86 | 0.85 | 0.85 |
| VGG16 | 0.89 | 0.88 | 0.89 | 0.88 |
| VGG19 | 0.88 | 0.88 | 0.88 | 0.88 |

Table II presents a complete evaluation of the four CNN models used in the analysis of brain tumour MRI images. The table outlines key performance metrics, including Accuracy, Precision, Recall and F-Score, for each model. The 3-layer CNN demonstrated an accuracy of 0.87, with similar precision, recall, and F-Score values. The 9-layer CNN exhibited slightly lower metrics at 0.85, while VGG16 and VGG19 demonstrated higher accuracy at 0.89 and 0.88, respectively. Although all the models displayed similar performances, the VGG16 model exhibited the highest overall performance compared to the other models. This may be due to the deeper architecture and the increased complexity of the VGG16 model, which allows it to capture complex features and patterns in the images more effectively. The higher accuracy of 0.89, along with balanced precision, recall, and F-Score values, indicates the model's robust ability to correctly classify brain tumours across various types.

*4.2 Tumour-Specific Model Performance Comparison*

TABLE III. MODEL ACCURACY FOR EACH TUMOUR TYPE

| Model | Glioma Tumour | Meningioma Tumour | No Tumour | Pituitary Tumour |
|---|---|---|---|---|
| 3 Layer CNN | 0.88 | 0.85 | 0.81 | 0.92 |
| 9 Layer CNN | 0.97 | 0.73 | 0.81 | 0.85 |
| VGG16 | 0.87 | 0.84 | 0.96 | 0.94 |
| VGG19 | 0.92 | 0.80 | 0.88 | 0.91 |

Table III shows the analysis of the accuracy achieved by each classification model in classifying brain tumour MRI images based on specific tumour types. The table consists of all the four tumour categories. For the Glioma Tumour classification, the 9-layer CNN model demonstrated the highest accuracy at 0.97, showcasing its exceptional proficiency in identifying this tumour type. The 3-layer CNN exhibited notable accuracy across all classes, with particularly high performance for Pituitary Tumour at 0.92. Interestingly, VGG16 excelled in recognizing No Tumour cases with an accuracy of 0.96, highlighting its efficiency in distinguishing normal brain images. VGG19, on the other hand, demonstrated strong overall performance for each tumour type, but does not achieve the highest accuracy for any of the classes. These results emphasize the model-specific strengths in accurately classifying distinct tumour types, providing valuable insights for researchers and practitioners seeking to optimize model selection based on the specific classification requirements of Glioma, Meningioma, No Tumour, and Pituitary Tumour cases in brain tumour MRI image analysis.

*4.2 Confusion Matrix of Models*

TABLE IV. CONFUSION MATRIX FOR 3 LAYER CNN MODEL

| | Glioma Tumour | Meningioma Tumour | No Tumour | Pituitary Tumour |
|---|---|---|---|---|
| **Glioma Tumour** | 85 | 9 | 0 | 3 |
| **Meningioma Tumour** | 6 | 73 | 2 | 5 |
| **No Tumour** | 0 | 5 | 21 | 0 |
| **Pituitary Tumour** | 2 | 4 | 0 | 72 |

TABLE V. CONFUSION MATRIX FOR 9 LAYER CNN MODEL

| | Glioma Tumour | Meningioma Tumour | No Tumour | Pituitary Tumour |
|---|---|---|---|---|
| **Glioma Tumour** | 94 | 2 | 0 | 1 |
| **Meningioma Tumour** | 16 | 63 | 3 | 4 |
| **No Tumour** | 3 | 2 | 21 | 0 |
| **Pituitary Tumour** | 8 | 3 | 1 | 66 |

TABLE VI. CONFUSION MATRIX FOR VGG16 CNN MODEL

| | Glioma Tumour | Meningioma Tumour | No Tumour | Pituitary Tumour |
|---|---|---|---|---|
| **Glioma Tumour** | 84 | 10 | 0 | 3 |
| **Meningioma Tumour** | 8 | 72 | 1 | 5 |
| **No Tumour** | 0 | 1 | 25 | 0 |
| **Pituitary Tumour** | 2 | 2 | 1 | 73 |

TABLE VII. CONFUSION MATRIX FOR VGG19 CNN MODEL

| | Glioma Tumour | Meningioma Tumour | No Tumour | Pituitary Tumour |
|---|---|---|---|---|
| **Glioma Tumour** | 89 | 7 | 0 | 1 |
| **Meningioma Tumour** | 12 | 69 | 1 | 4 |
| **No Tumour** | 0 | 3 | 23 | 0 |
| **Pituitary Tumour** | 4 | 1 | 2 | 71 |

Tables IV to VII describe confusion matrices for the four classification models. Each matrix presents the performance of the respective CNN model in distinguishing among the four classes. The rows indicate the true classes, and the columns represent the predicted classes. The diagonal values (top left to bottom right) represent correctly classified cases, while off-diagonal values indicate misclassifications.

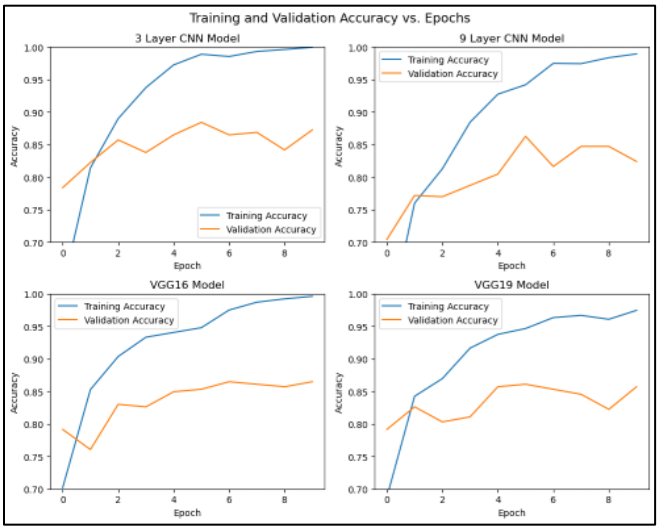
Table IV presents the confusion matrix for the 3-layer CNN model. It correctly identifies 85 Glioma Tumours, 73 Meningioma Tumours, 21 cases of No Tumour, and 72 Pituitary Tumours. Misclassifications are also noted, such as 9 Glioma cases being incorrectly predicted as Meningioma Tumours and 6 Meningioma cases as Glioma Tumours, among others.

Table V presents the results for the 9-layer CNN model, indicating a higher accuracy in identifying Glioma Tumours (94 correct predictions) but more misclassifications in Meningioma Tumours (16 predicted as Glioma Tumours). The model also correctly identified 21 No Tumour cases and 66 Pituitary Tumours, with several misclassifications between the categories.

Table VI presents the confusion matrix for the VGG16 CNN model. This model demonstrates a balanced performance across the board, with 84 correct predictions for Glioma Tumours, 72 for Meningioma Tumours, 25 for No Tumour (showing an improvement in identifying cases with no tumour), and 73 for Pituitary Tumours. Misclassifications are relatively low, indicating a robust model performance.

Table VII presents the confusion matrix for the VGG19 CNN model. It shows an improvement in correctly identifying Glioma Tumours (89 correct predictions) and a good performance on Meningioma Tumours (69 correct predictions), No Tumour (23 correct predictions), and Pituitary Tumours (71 correct predictions).

### *4.3 Training and Validation Accuracy Comparison*




Line graphs are used to compare the training and validation accuracy for all four CNN models over the number of training epochs (10). The training accuracy represents the model's performance on the dataset it learns from, while the validation accuracy measures how well the model generalizes to new, unseen data. For the 3-layer and 9-layer CNN models, both accuracies increase over time, indicating learning, but the validation accuracy for the 9-layer model shows more volatility, suggesting potential overfitting. The VGG16 and VGG19 models, which are more sophisticated, also show an increase in training accuracy, but their validation accuracies stabilize earlier and exhibit less volatility compared to the 9-layer model. Overall, while the more complex models of VGG16 and VGG19 might be expected to perform better, they don't show a significant advantage in validation accuracy over the simpler models in these results.

## V. CONCLUSION

Our study investigated the effectiveness of Convolutional Neural Networks (CNNs) in classifying brain tumours using MRI images. Four different CNN models were compared, two custom-built CNN models with 3 and 9 layers and another two pre-trained models, VGG16 and VGG19. All the models were evaluated on a single dataset of 3264 categorized MRI images, which include a variety of MRI images that categorizes into four classes: glioma, meningioma, pituitary tumours, and no tumour.

While all models performed well in identifying tumour-free images, the VGG16 model has successfully achieved 89% accuracy and recall, as well as an 88% precision and F-score from the tests. This likely comes from its deeper architecture and pre-trained weights, which allows it to leverage pre-existing knowledge for better adaptation and accuracy rates. While not surpassing VGG16, the custom-built CNN models still demonstrated impressive results, with the 9-layer model reaching 85% accuracy and the 3-layer model at 87%. This suggests that carefully designed custom CNNs can hold their own and be compared against pre-trained models, offering alternative approaches. The VGG19 pre-trained model however, scored a lower accuracy of 88% compared to the VGG16 model.

While this study offers promising results, it is important to acknowledge the limitations of this study. The main limitation faced is hyperparameter tuning. Various hyperparameter tuning options have been explored such as random search, Bayesian optimization, and Optuna optimization on the CNN models, but these efforts did not improve the model's accuracy at all. This may be due to the overall nature of the dataset, limited computational resources at hand, as well as time constraints. Further research with a wider range of hyperparameter combinations and optimization techniques is needed to fully explore this aspect. Additionally, the dataset size could potentially benefit from further expansion and inclusion of more diverse patient data to enhance the generalizability of the model's performance.

Overall, this research highlights the promising potential of custom and pre-trained CNNs for brain tumour classification in MRI images, but it also highlights the need for further exploration and refinement. By addressing the limitations identified and actively pursuing future research avenues, higher model accuracy and generalizability can be achieved, which will ultimately contribute to earlier diagnosis and more effective treatment plans for brain tumour patients.

## REFERENCES


[1] Rohit Thakur. (2019, August 6). Step by step VGG16 implementation in Keras for beginners. Medium; Towards Data Science. https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c

[2] Balaji, S. (2020, August 29). Binary Image classifier CNN using TensorFlow. Medium. https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697

[3] Bhanothu, Y., Kamalakannan, A., & Rajamanickam, G. (2020, April 23). Detection and Classification of Brain Tumour in MRI Images using Deep Convolutional Network | IEEE Conference Publication | IEEE Xplore.Ieeexplore.ieee.org.https://ieeexplore.ieee.org/abstract/document/9074375

[4] Ayadi, W., Elhamzi, W., Charfi, I., & Atri, M. (2021). Deep CNN for Brain Tumour Classification. Neural Processing Letters, 53(1), 671–700. https://doi.org/10.1007/s11063-020-10398-2

[5] Chattopadhyay, A., & Maitra, M. (2022). MRI-based brain tumour image detection using CNN based deep learning method. Neuroscience Informatics, 2(4), 100060. https://doi.org/10.1016/j.neuri.2022.100060

[6] Lamrani, D., Cherradi, B., El Gannour, O., Amine Bouqentar, M., & Bahatti, L. (2022). Brain Tumour Detection using MRI Images and Convolutional Neural Network - ProQuest. *Www.proquest.com*. https://www.proquest.com/openview/d39b1b3915b69b5fc14a94d8cb1ed0d0/1?pq-origsite=gscholar&cbl=5444811