Brain Tumor Classification via ensemble learning

Priyanshu Sukumar Ghosh (E23CSEU1107)

Batch 37

Abstract

Brain tumour classification from MRI scans is a critical task in medical diagnostics, often requiring high precision to assist in early detection and treatment planning. This project investigates a deep learning-based approach for automated classification of brain tumours into four categories: glioma, meningioma, pituitary, and no tumour. A custom Convolutional Neural Network (CNN) was initially developed and trained as a baseline model. To improve classification performance, transfer learning was employed using pre-trained VGG19 and InceptionV3 architectures. Each model was trained and evaluated individually on the same dataset with appropriate preprocessing and augmentation techniques.

To further enhance model robustness and accuracy, an ensemble learning strategy was adopted by aggregating predictions from the custom CNN, VGG19, and InceptionV3 models. The ensemble model demonstrated improved performance over individual models, suggesting that combining complementary architectures provides a more generalized and accurate prediction framework. Performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices.

The results validate the effectiveness of ensemble learning in medical image classification tasks and highlight the potential of deep learning models in supporting radiological diagnosis. This report presents the methodology, model architectures, training procedures, and comparative analysis of results in detail.

Table of Contents

8. **Discussion**
    8.1. Observations
    8.2. Limitations
    8.3. Potential Improvements and Future Work
9. **Conclusion**
    9.1. Summary of Contributions
    9.2. Key Takeaways and Impact
10. **References**

Introduction

Motivation

Brain tumours are among the most critical and life-threatening forms of cancer, impacting both cognitive and physical functions depending on their location and severity. Early and accurate detection of brain tumours is crucial to improve treatment outcomes and patient survival rates. Traditional diagnosis involves manual analysis of MRI scans by radiologists, which is not only time-consuming but also prone to inter-observer variability. In resource-limited settings, the lack of specialized radiologists further exacerbates the challenges of timely diagnosis. Thus, there is a growing need for intelligent, automated systems to assist in early tumour classification and decision-making.

Challenges in Manual Diagnosis

Despite the advancements in MRI imaging technologies, the manual identification and classification of brain tumours remain challenging due to:

- Subtle differences between tumour types (e.g., glioma vs meningioma),
- High inter-patient variability,
- Overlapping intensity patterns between tumour and healthy tissues,
- Human fatigue and diagnostic subjectivity.

These factors can lead to misclassification or delayed diagnosis, affecting the prognosis and treatment plan.

Role of Machine Learning in Medical Imaging

Machine learning (ML), particularly deep learning techniques such as Convolutional Neural Networks (CNNs), has shown remarkable performance in computer vision tasks. In medical imaging, ML models have been successfully applied to automate the detection and classification of various conditions, including lung cancer, diabetic retinopathy, and more recently, brain tumors. These models learn complex patterns directly from image data, often surpassing traditional hand-crafted feature-based techniques. Transfer learning from pre-trained networks (e.g., VGG19, InceptionV3) further enhances performance by leveraging knowledge from large-scale image datasets.

Project Objectives

This project aims to design and evaluate a robust brain tumour classification pipeline using multiple deep learning models:

- Develop a custom CNN tailored for MRI classification.
- Fine-tune state-of-the-art pre-trained models (VGG19 and InceptionV3) using transfer learning.
- Combine individual models through ensemble learning to enhance overall performance and reliability.
- Deploy the final model in a lightweight, interactive web interface using Gradio to simulate real-world applicability.

By focusing on classification rather than segmentation, this work offers a simpler and more accessible approach that can aid radiologists or serve as an educational and diagnostic support tool.

Related Work

Tumour Segmentation vs Classification

Brain tumour analysis in medical imaging predominantly falls into two categories: segmentation and classification. **Segmentation** focuses on identifying and delineating the tumour region within an MRI scan, often using techniques like U-Net or attention-based networks. This approach is highly beneficial in pre-surgical planning. In contrast, **classification** aims to categorise the type of tumour present (e.g., glioma, meningioma, pituitary), which is more aligned with diagnostic support. This work prioritises classification to provide a quick and automated way to assist clinicians in tumour type identification.

Literature Review

The field has seen numerous studies employing deep learning to tackle brain tumour detection. One such work, *"A review on brain tumor segmentation of MRI images"* by Wadhwa et al., offers a comprehensive overview of segmentation methodologies and highlights the effectiveness of different architectures across datasets. Although the paper primarily addresses segmentation, it lays foundational understanding of tumour types and medical relevance.
Another pivotal study, *"Brain Tumor Classification Using Convolutional Neural Networks"* by Seetha and Raja, explores the use of CNNs in classifying MRI images. The study illustrates how deep networks can learn intricate patterns in brain scans and underlines the importance of architectural design and preprocessing in achieving high performance.

Distinctiveness of This Approach

Unlike segmentation-based methods, this project focuses entirely on **tumour classification**, aiming for a lightweight, deployable solution that can be integrated into real-world diagnostic tools. It experiments with multiple architectures (a custom CNN, VGG19, InceptionV3) and employs ensemble learning to enhance robustness and accuracy. The integration with a Gradio-powered web interface also adds a practical dimension, making it more than just a theoretical contribution.

Dataset

Dataset Source

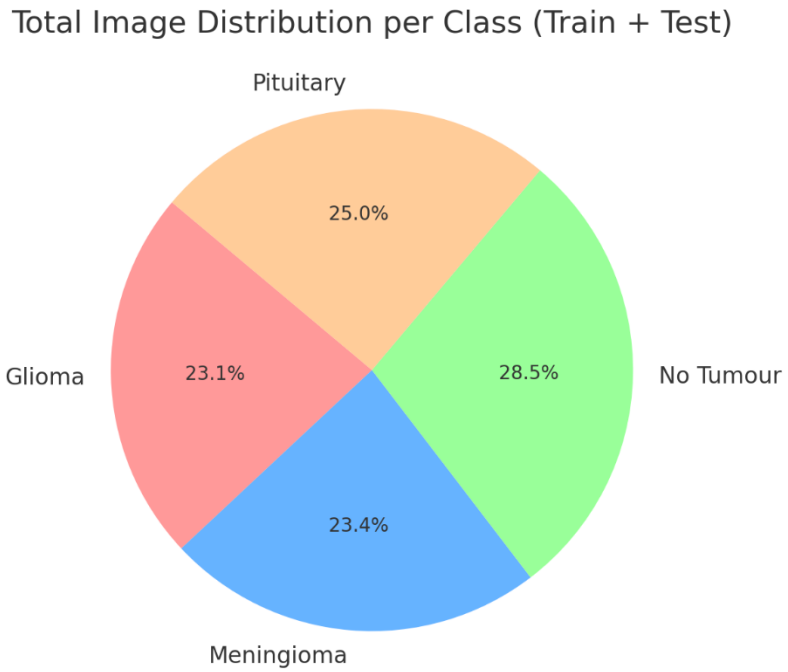
The dataset used in this project is the **Brain Tumour MRI Dataset** available on Kaggle. It contains **T1-weighted contrast-enhanced MRI scans** classified into four categories: **glioma tumours**, **meningioma tumours**, **pituitary tumours**, and **no tumour**. The dataset is organised into separate training and testing folders, each containing subfolders for the respective classes.

This dataset has become a common benchmark for evaluating tumour classification models due to its well-curated structure, decent class balance, and relatively clean MRI images.

Class Distribution

The dataset includes the following number of images per class:

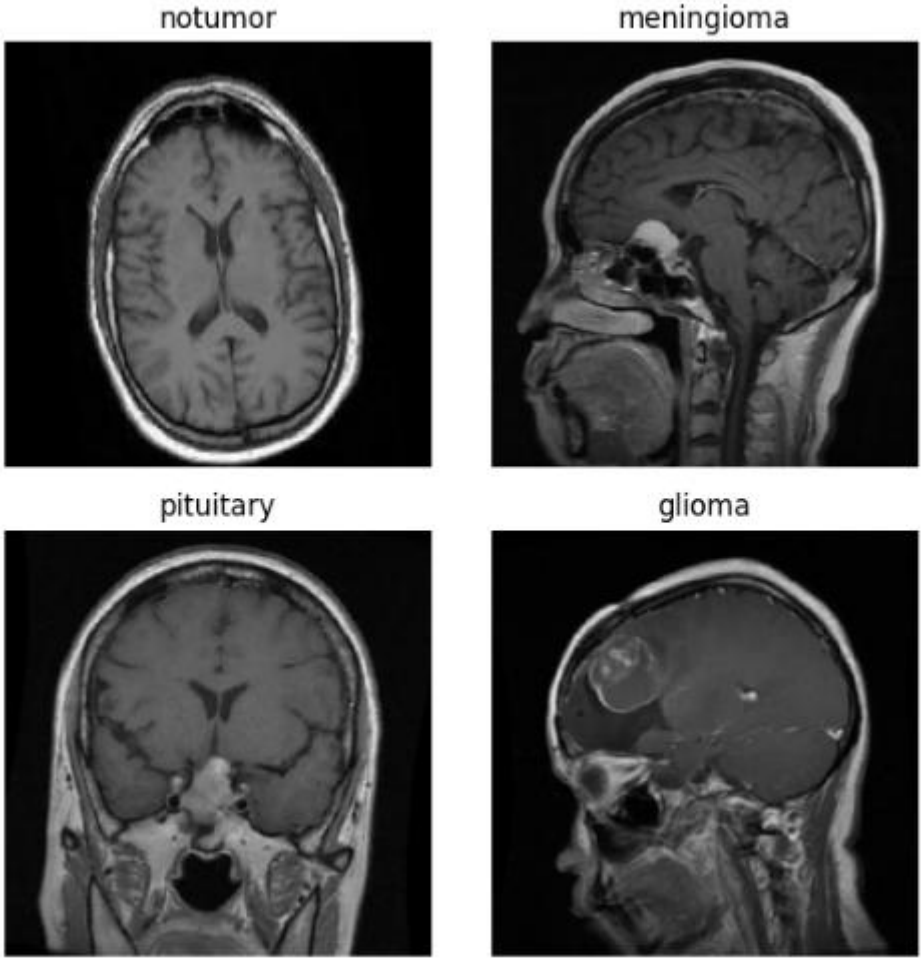
Class	Training Images	Testing Images
Glioma Tumour	1321	300
Meningioma Tumour	1339	306
No Tumour	1595	405
Pituitary Tumour	1457	300



While the dataset is generally balanced among tumour types, the **'no tumour'** category is slightly overrepresented in the training and testing set. However, the overall distribution is still sufficient for training and evaluation of deep learning models with appropriate augmentation.

Sample Visualisations

Below are sample images from each class in the dataset:



These visual examples highlight the complexity of visually distinguishing between different tumour types, reinforcing the importance of automated classification tools. Differences in tumour shape, size, and location can be subtle, making it challenging for traditional rule-based systems.

Methodology

Preprocessing Techniques

The dataset comprises T1-weighted brain MRI scans grouped into four categories: glioma, meningioma, pituitary tumour, and no tumour. Each image was resized to **224×224 pixels** and converted to **three-channel RGB** format to meet the input requirements of pre-trained models such as VGG19 and InceptionV3. Pixel values were scaled to the [0,1][0, 1][0,1] range via division by 255, facilitating more stable training.

The ImageDataGenerator class from Keras was used to streamline image preprocessing and augmentation. The training generator applied the following augmentations to improve generalisability:

- **Rotation**: up to 15 degrees
- **Zoom**: up to 10%
- **Horizontal flipping**

The test generator only performed rescaling, preserving the original test distribution for unbiased evaluation. The batch size was set to 16 for both generators. The code used for this setup is as follows:

```
correct_image_shape = (224, 224, 3)

train_image_gen = ImageDataGenerator(
    rescale=1./255,
    rotation_range=15,
    zoom_range=0.1,
    horizontal_flip=True
).flow_from_directory(
    train_dataset_location,
    target_size=correct_image_shape[:2],
    batch_size=16,
    class_mode='categorical'
)

test_image_gen = ImageDataGenerator(
    rescale=1./255
).flow_from_directory(
    test_dataset_location,
    target_size=correct_image_shape[:2],
    batch_size=16,
    class_mode='categorical'
)
```

These generators automatically label the data based on directory structure, yielding one-hot encoded labels suitable for categorical cross-entropy loss.

Model Architectures

Three distinct CNN architectures were utilised for classification: a custom-built CNN, VGG19, and InceptionV3. The latter two models were two transfer learning models.

Custom CNN

The custom CNN architecture was designed from scratch to provide a lightweight yet expressive model for tumour classification. It begins with three convolutional blocks:

- **Conv2D (32 filters, 3×3)** → Batch Normalisation → Max Pooling
- **Conv2D (32 filters, 3×3)** → Batch Normalisation → Max Pooling
- **Conv2D (64 filters, 3×3)** → Batch Normalisation → Max Pooling

These extract spatial features while progressively reducing dimensionality. The output is then flattened and passed through two fully connected layers:

- **Dense (128 units, ReLU)** → Batch Normalisation → Dropout (30%)
- **Dense (32 units, ReLU)** → Batch Normalisation → Dropout (30%)

Finally, a softmax layer with 4 units produces class probabilities. This model was trained from scratch using the preprocessed grayscale MRI images converted to RGB.

VGG19 (Transfer Learning)

For the VGG19-based model, a pre-trained VGG19 network (trained on ImageNet) was used as a feature extractor with its convolutional base frozen:

- **Pre-trained VGG19 (no top layers)** → Flatten

The extracted features were then fed into a classification head:

- **Dense (128 units, ReLU)** → Batch Normalisation → Dropout (30%)
- **Dense (32 units, ReLU)** → Batch Normalisation → Dropout (30%)
- **Dense (4 units, softmax)**

This approach benefits from transfer learning, as VGG19's early layers provide rich general features applicable to medical imaging.

InceptionV3 (Transfer Learning)

The third model utilised a pre-trained InceptionV3 network from TensorFlow Hub, with the base frozen to prevent overfitting:

- **InceptionV3 KerasLayer (trainable=False)** → Batch Normalisation

The classification head is identical to the one used in the VGG19 model:

- **Dense (128 units, ReLU)** → Batch Normalisation → Dropout (30%)
- **Dense (32 units, ReLU)** → Batch Normalisation → Dropout (30%)
- **Dense (4 units, softmax)**

InceptionV3's architectural advantage lies in its ability to capture multi-scale features through factorised convolutions, making it highly effective in handling the complex patterns in MRI scans.

Training Configuration

Input Shape and Optimiser

All images were resized to a uniform input shape of **(224 × 224 × 3)**. Although the original MRI images were in grayscale, they were converted to RGB to be compatible with pre-trained models. The **Adam optimiser** was used across all models, configured with a low learning rate of **1e-5** to ensure stable and gradual convergence, especially during fine-tuning of the transfer learning models.

Loss Function and Callbacks

The models were compiled using the **categorical crossentropy** loss function, which is suitable for multi-class classification tasks. Accuracy was used as the primary evaluation metric during training.

To prevent overfitting and to monitor training progress efficiently, the following callbacks were employed:

- **EarlyStopping** on training loss (monitor='loss') with a patience of 2 epochs.
- **EarlyStopping** on validation loss (monitor='val_loss') with restore_best_weights=True, ensuring the best-performing model on the validation set is retained.
- **ModelCheckpoint** on validation accuracy (monitor='val_accuracy', mode='max'), saving only the best model to disk.

A sample training setup for the VGG19 model is shown below:

```
model.compile(
    loss=CategoricalCrossentropy(),
    optimizer=Adam(learning_rate=1e-5),
    metrics=['accuracy']
)

history = model.fit(
    train_image_gen,
    validation_data=test_image_gen,
    epochs=30,
    callbacks=[early_stop_loss, early_stop_val_loss, checkpoint_cb]
)
```

After training, the model weights and training history were saved to disk for further analysis and reproducibility. The training history was saved as a JSON file, allowing for visualisation of accuracy and loss trends.

Ensemble Strategy

In order to harness the strengths of multiple models and improve generalisability, an ensemble approach was implemented by combining the predictions of the three individually trained models: the Custom CNN, VGG19, and InceptionV3.

Ensemble Technique Used

A **late fusion strategy** was employed, where each of the base models was used to independently generate output probability distributions over the four tumour classes. These predictions were then averaged using a tf.keras.layers.Average() layer. This approach mitigates the weaknesses of any single model and reduces prediction variance.

After averaging, the ensemble passed through a custom classification head comprising:

- **Batch Normalisation** to stabilise and accelerate training,
- **Dense(128) → Dropout(0.3)** for learning higher-level representations,
- **Dense(32) → Dropout(0.3)** for further refinement, and
- A final **Dense(4)** layer with softmax activation for multi-class output.

This ensemble was wrapped into a new Keras Model instance and trained on the original dataset using the same loss, optimizer, and callback configuration as individual models.

Justification and Benefits

The ensemble method capitalises on the unique learning characteristics of each base model:

- The **Custom CNN** is lightweight and effective for capturing basic patterns.
- **VGG19**, being deep and hierarchical, extracts abstract features from transfer learning.
- **InceptionV3**, known for its multi-scale convolutional architecture, complements the others by capturing varied spatial hierarchies.

Combining these diverse feature extractors led to enhanced performance, as reflected in improved validation accuracy and generalisation on the test set.

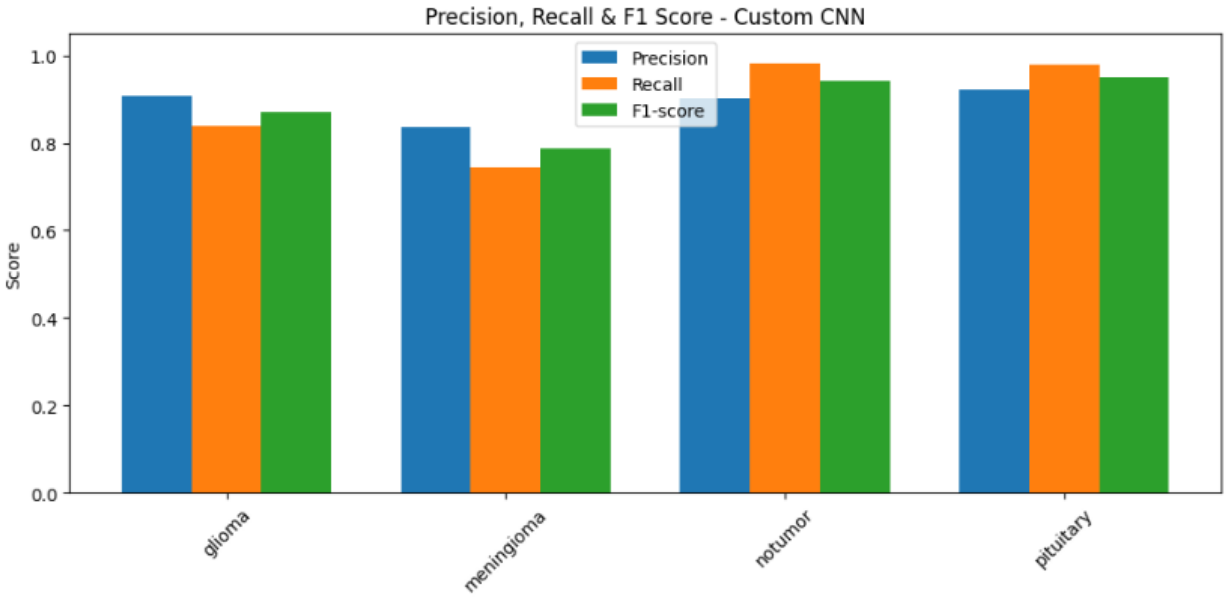
Results and Evaluation

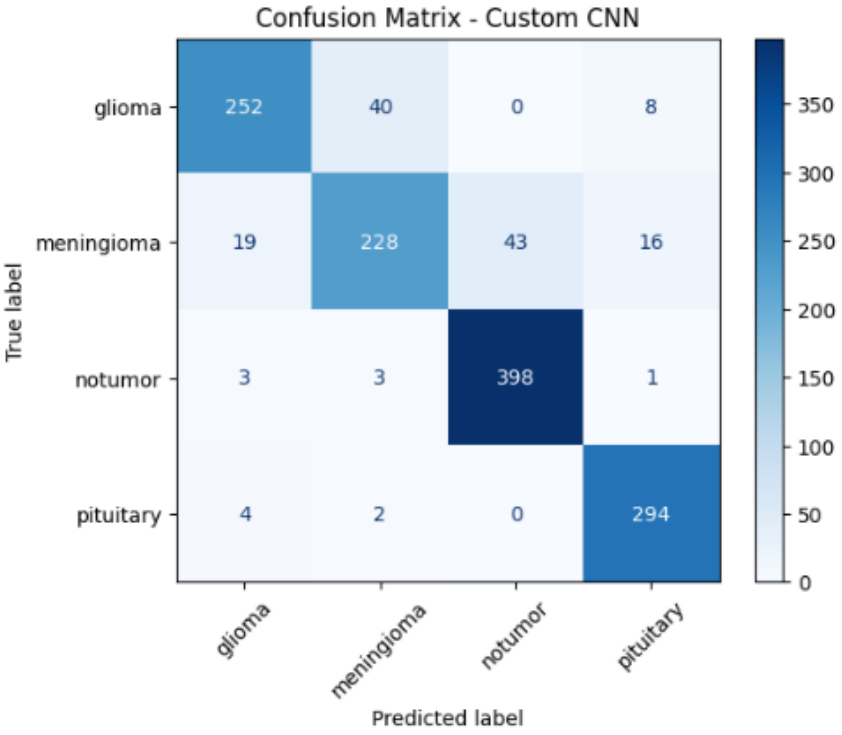
Performance Metrics (Accuracy, Precision, Recall, F1-Score)

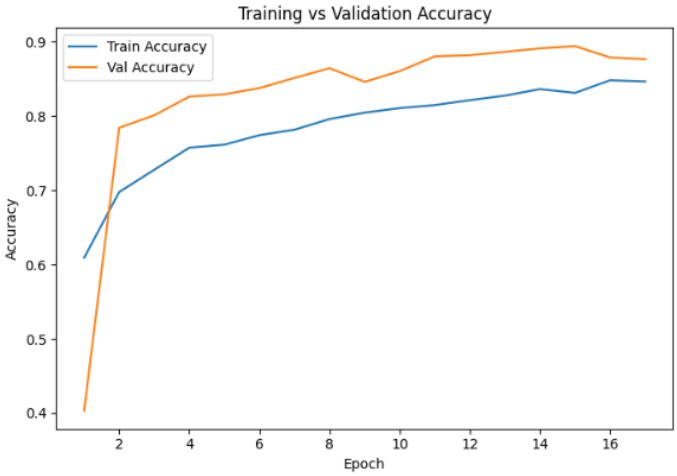
To evaluate the effectiveness of the individual and ensemble models, we employed the following metrics:

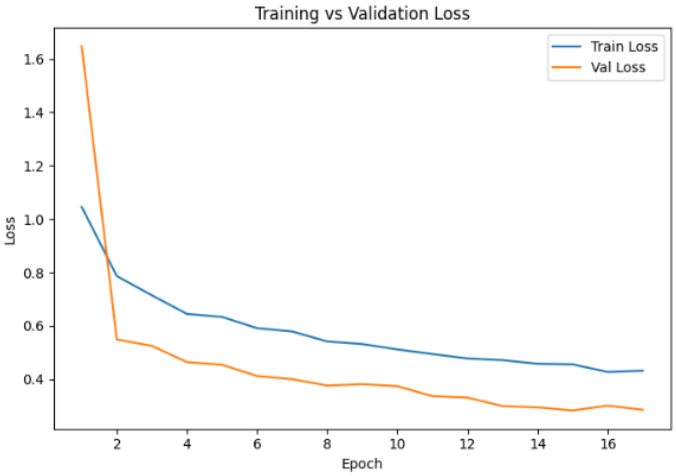
- **Precision**: Indicates how many predicted tumour classes were actually correct.
- **Recall**: Reflects the model's ability to detect all actual tumour instances.
- **F1-Score**: Harmonic mean of precision and recall; balances false positives and false negatives.
- **Confusion Matrix**: Visualises the model's classification performance by showing the correct vs. incorrect predictions across all tumour categories. It helps identify specific classes where the model might be underperforming.
- **Training vs Validation Curves**: Plots of accuracy and loss across epochs for both training and validation datasets. These are used to monitor model convergence, detect overfitting or underfitting, and guide early stopping decisions.

Custom CNN model –

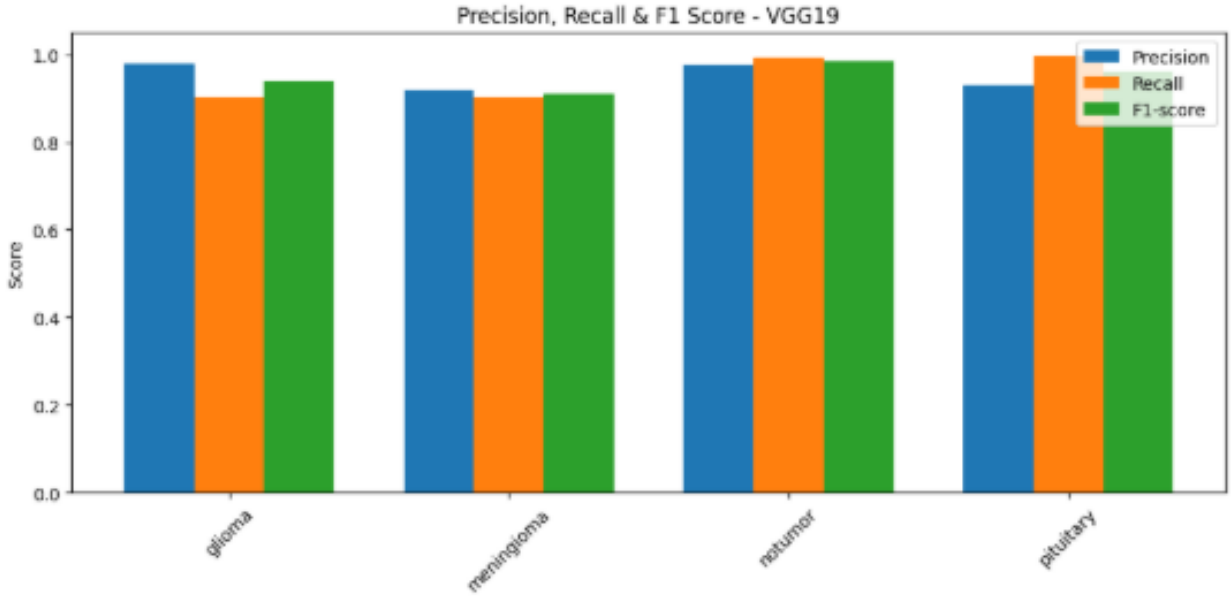


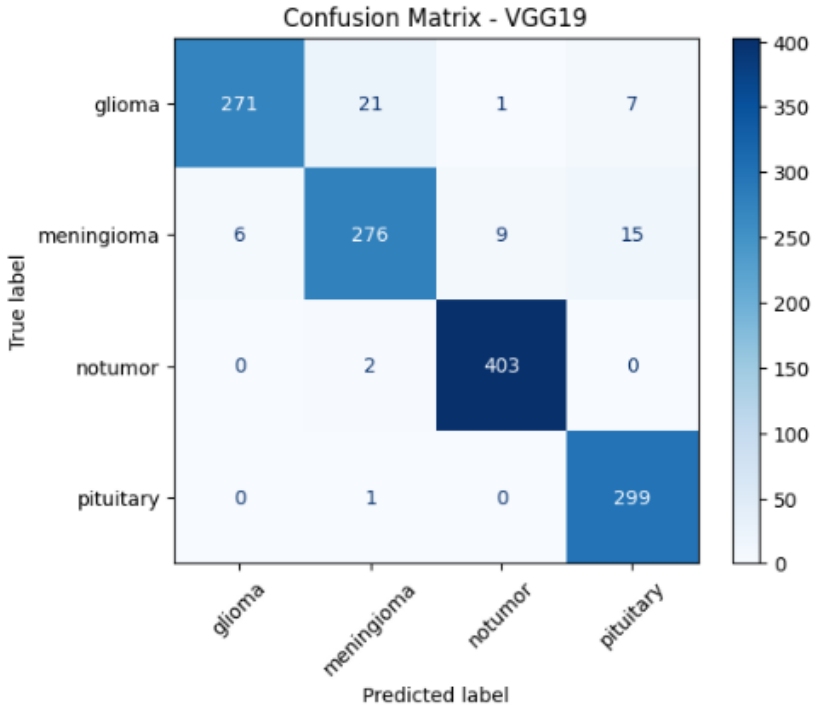

Confusion Matrix - Custom CNN
glioma
meningioma
notumor
pituitary
True label
252
40
0
8
19
228
43
16
3
3
398
1
4
2
0
294
350
300
250
200
150
100
50
0
glioma
meningioma
notumor
pituitary
Predicted label

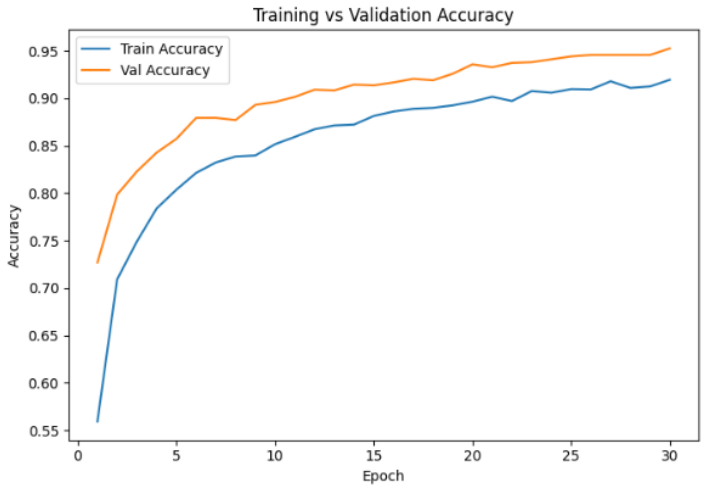

Training vs Validation Accuracy
Train Accuracy
Val Accuracy
Accuracy
Epoch

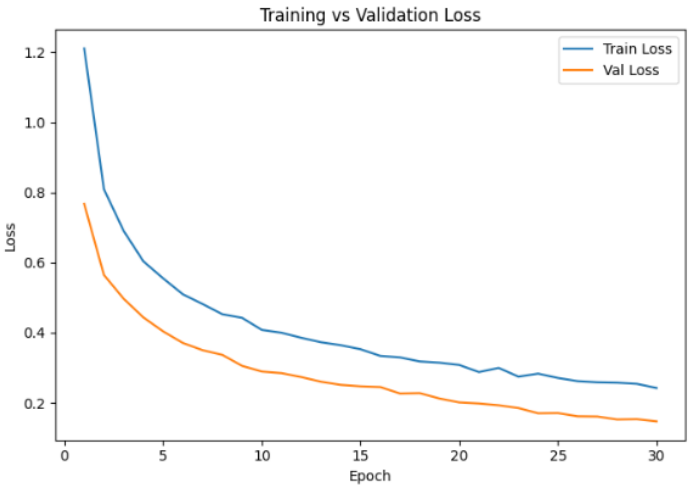

Training vs Validation Loss
Train Loss
Val Loss
Loss
Epoch

VGG19 model –

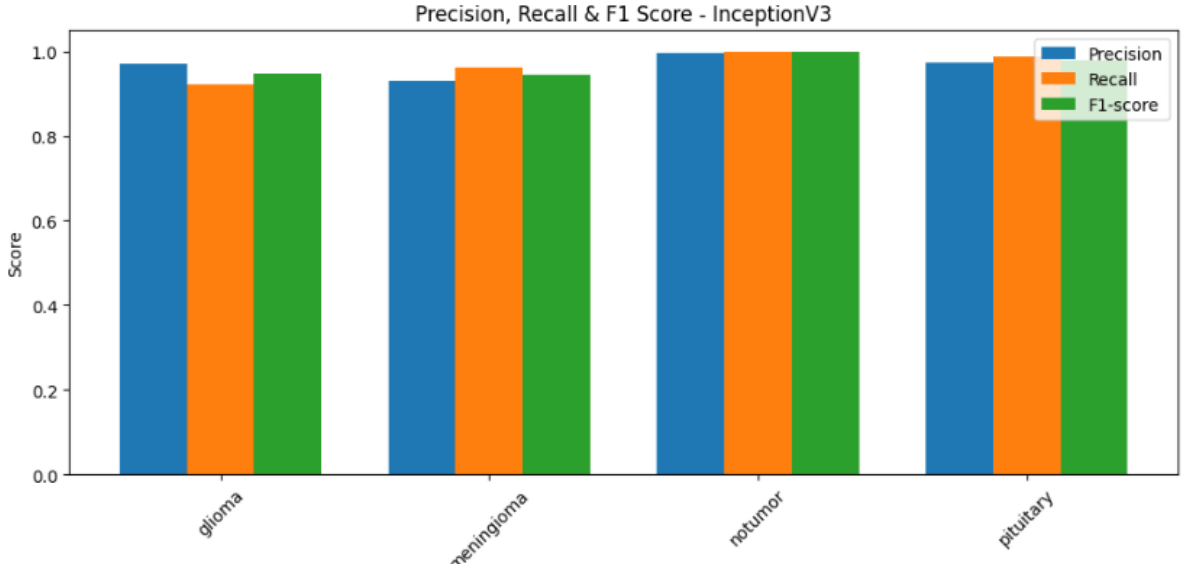

Precision, Recall & F1 Score - VGG19
Precision
Recall
F1-score
Score
glioma
meningioma
notumor
pituitary

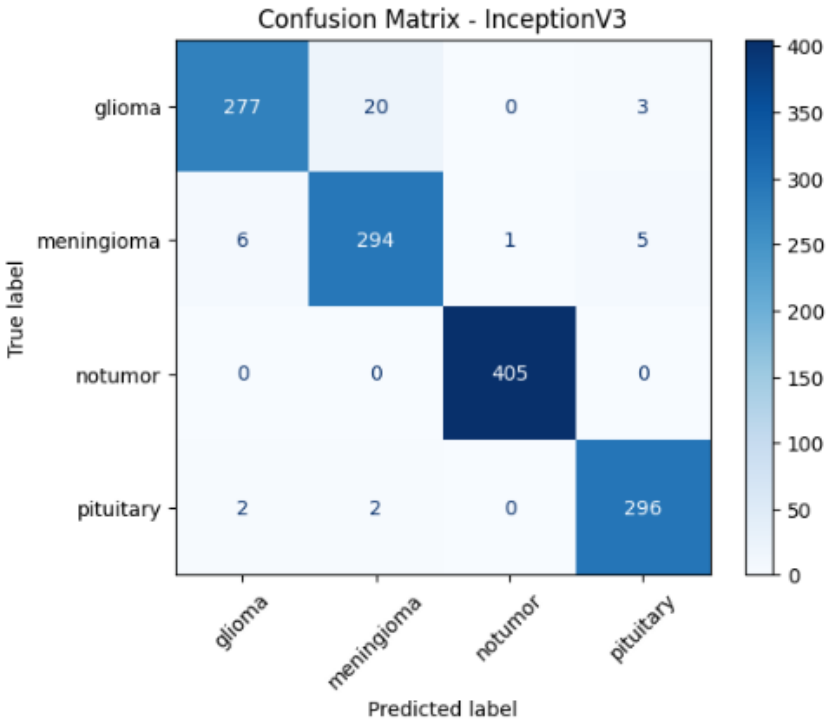


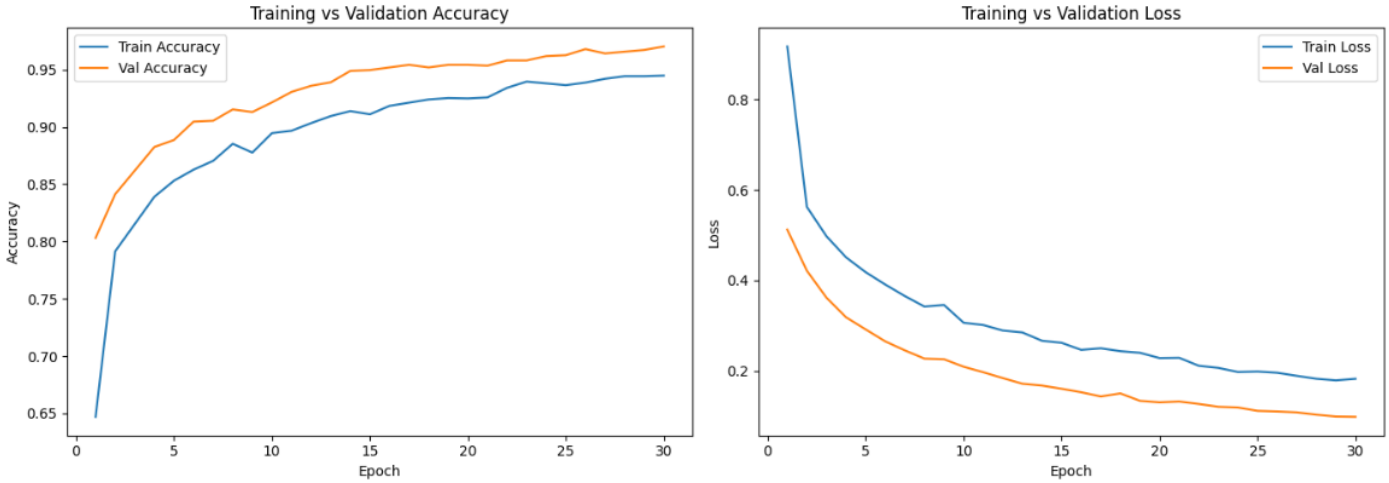




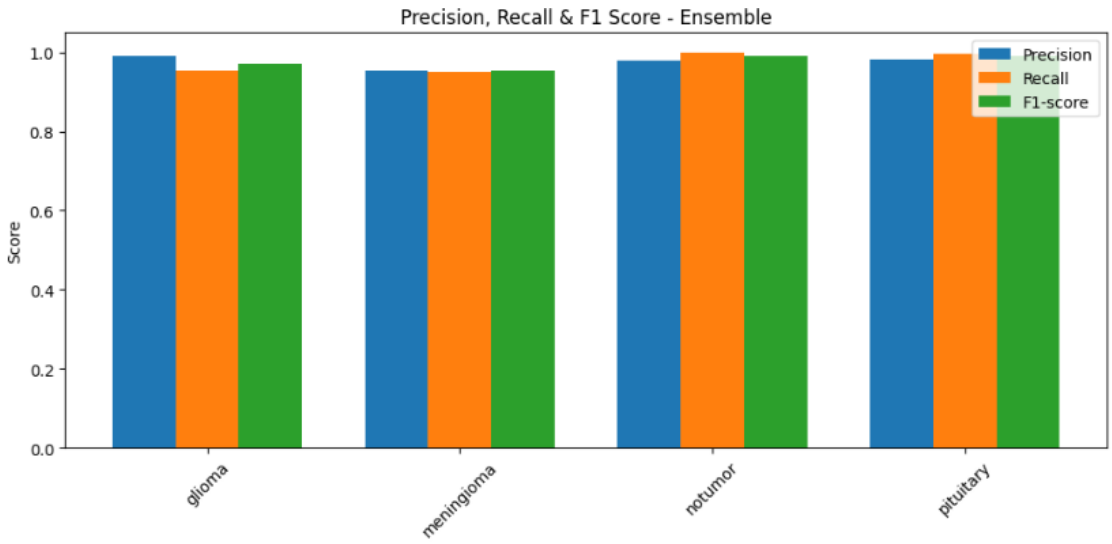
InceptionV3 model –

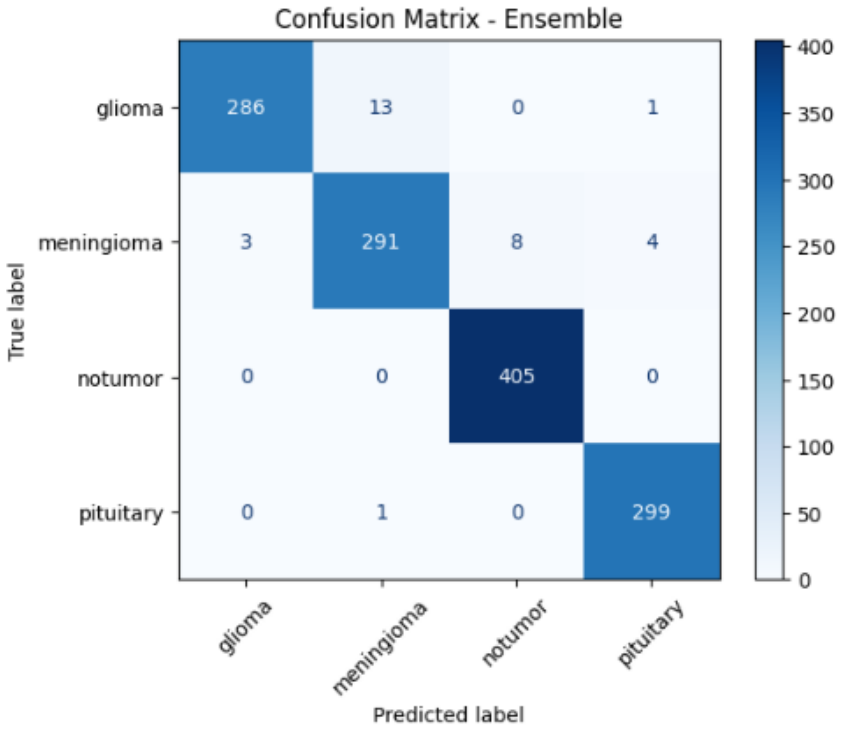


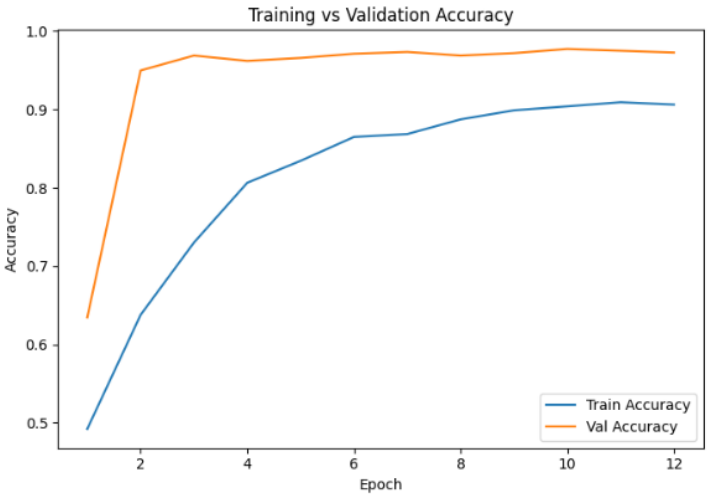


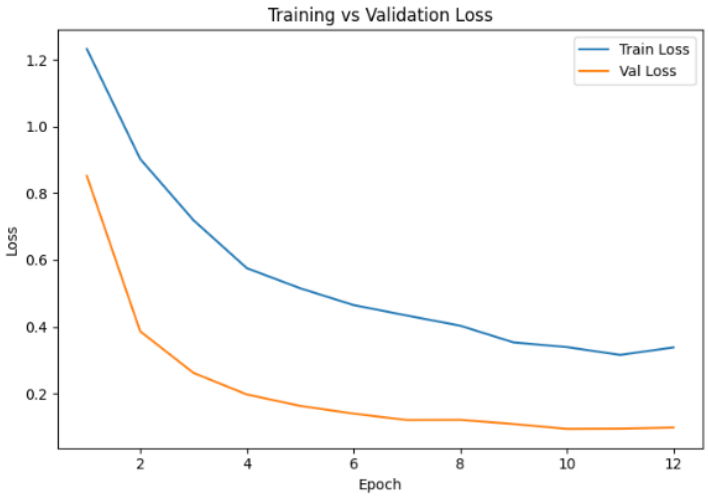


Ensemble Learning model –









Model Comparison and Insights

- **Custom CNN** was fast to train and performed reasonably well, showing that even lightweight architectures can be effective with preprocessing and augmentation.
- **VGG19** benefitted from deep feature extraction but had a larger parameter footprint.
- **InceptionV3** captured multi-scale features and performed slightly better than VGG19.
- **Ensemble** of all three offered a synergistic boost in performance, especially in reducing misclassifications among similar tumour types.

Key Insight: Combining diverse model architectures with different inductive biases increases robustness, especially when dealing with subtle inter-class MRI differences.

Deployment & Interface

Gradio-Based Web Application

To provide a user-friendly interface for interacting with the trained tumour classification model, we developed a web application using Gradio, an open-source Python library for creating intuitive ML model interfaces. Gradio simplifies the deployment process and allows for rapid prototyping without the need for complex frontend development.

The application enables users to upload a brain MRI image and receive a classification prediction in real-time, indicating whether the scan shows signs of a glioma, meningioma, pituitary tumour, or no tumour.

Workflow and User Experience

The webapp follows a streamlined process:

1. **Image Upload**: The user selects or drags an MRI image (in JPG/PNG format) into the upload section.
2. **Preprocessing**: The image is resized and normalised to match the model input specifications.
3. **Prediction**: The ensembled model processes the image and returns probabilities for each tumour class.
4. **Result Display**: The predicted class with the highest confidence is displayed.

Discussion

Observations

Throughout the development of the brain tumour classification models, several trends were observed. Firstly, transfer learning using InceptionV3 and VGG19 consistently outperformed the custom CNN in terms of validation accuracy, highlighting the strength of pretrained architectures for medical imaging tasks with limited data. The ensemble approach further improved classification robustness, reducing the variance seen in individual model predictions. Tumour types with more distinct visual patterns, such as pituitary tumours, achieved higher precision and recall compared to glioma, which displayed greater intra-class variability.

Limitations

While the ensemble model achieved improved overall accuracy, several limitations were identified:

- **Data imbalance** across tumour classes may have biased learning, despite augmentation.
- **Layer compatibility** during ensembling posed constraints; merging outputs from heterogeneous models (e.g., Inception and custom CNN) required careful architecture alignment.
- **Grayscale-to-RGB conversion** was necessary for compatibility with pretrained networks but may have introduced redundancy or artefacts.
- **Dataset quality and generalisability**: The models were trained and tested on a specific dataset (Kaggle Brain Tumor MRI), which may not generalise to other clinical settings or MRI protocols.

Potential Improvements and Future Work

Future work could focus on:

- **End-to-end tumour segmentation and classification**, incorporating pixel-wise localisation to enhance clinical utility.

- **Fine-tuning pretrained layers** on a larger, domain-specific dataset to improve feature representation.
- **Incorporation of volumetric data (3D MRI slices)** to capture spatial context across multiple axial views.
- **Explainable AI techniques**, such as Grad-CAM, to visualise regions influencing predictions and increase model transparency for medical practitioners.

Conclusion

Summary of Contributions

This project presented a deep learning-based pipeline for classifying brain tumours into four categories—glioma, meningioma, pituitary tumour, and no tumour—using MRI scans. A custom CNN, VGG19, and InceptionV3 were trained, evaluated, and later combined via ensemble learning to improve prediction reliability. A Gradio-based web application was developed to provide a user-friendly interface for practical deployment.

Key Takeaways and Impact

The use of transfer learning significantly boosts classification performance in medical imaging domains with limited annotated data. Ensemble learning adds robustness and mitigates overfitting from any single model. Despite the challenges, the system demonstrates the potential of deep learning for aiding early diagnosis of brain tumours, paving the way for more advanced decision-support tools in healthcare.

References

- Wadhwa, A., Bhardwaj, A., & Verma, V. S. (2019). A review on brain tumor segmentation of MRI images. *Computerized Medical Imaging and Graphics*, 75, 1–13.
- Seetha, J., & Raja, S. S. (2018). *Brain Tumor Classification Using Convolutional Neural Networks*. *Biomedical & Pharmacology Journal*, 11(3), 1457–1461.
- TensorFlow Documentation. https://www.tensorflow.org/
- Gradio Documentation. https://www.gradio.app/docs/
- Kaggle Notebooks Documentation. https://www.kaggle.com/docs/notebooks
- Brain Tumor MRI Dataset (Kaggle). https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset