Brain Tumour Classification With Explainable AI

University of Piraeus - NCSR Demokritos
Artificial Intelligence Application

First Author
Orfanos Stamatios
mtn2211
stamatisorfanos99@gmail.com

Abstract

Brain tumour recognition is a critical task in medical diagnosis and treatment planning. With the advancements in artificial intelligence (AI), there has been a growing interest in utilising Machine-Learning and Deep-Learning techniques to automate tumour detection and classification. However, the lack of interpretability in traditional AI models hinders their adoption in clinical practice. In this paper, we want to create a Deep Learning model for brain tumour recognition using explainable AI. Our method combines state-of-the-art deep learning architectures with interpretable feature visualisation techniques to improve both the accuracy and interpretability of the tumour classification process. We evaluate our approach on a data-set of brain MRI scans and demonstrate its effectiveness in achieving accurate tumour recognition while providing clinicians with valuable insights into the decision-making process.

1 Introduction

Brain tumours pose a significant health challenge worldwide, with early and accurate detection playing a crucial role in patient prognosis and treatment planning. Medical imaging techniques, such as magnetic resonance imaging (MRI), have proven to be valuable tools for tumour visualisation and characterisation. However, manual interpretation of MRI scans by radiologists is time-consuming and subject to human error. Therefore, there is a pressing need to develop automated systems that can assist radiologists in the accurate detection and classification of brain tumours.

In recent years, deep learning models have shown remarkable success in various medical image analysis tasks. Convolutional neural networks (CNNs) have achieved state-of-the-art performance in tumour recognition from MRI scans. However, one of the major challenges associated with deep learning models is their lack of interpretability. The black-box nature of these models limits their adoption in clinical practice, as clinicians require explanations for the decisions made by the AI system.

To address this limitation, we propose an approach that combines deep learning models with explainable AI techniques for brain tumour recognition. Our goal is to develop a system that not only achieves high accuracy in tumour classification but also provides interpretable explanations for its decisions. By integrating interpretable feature visualisation methods, such as gradient-weighted class activation mapping (Grad-CAM) and saliency maps, we aim to generate visual explanations that highlight the regions of the MRI scans that contribute most to the tumour classification.

In this paper, we describe our methodology for brain tumour recognition using explainable AI. We present an overview of the data-set used for training and evaluation, discuss the architecture of our deep learning model, and explain the interpretability techniques incorporated into the system. We then present experimental results that demonstrate the effectiveness of our approach in terms of both accuracy and interpretability.

2 Data

In this project we use the MRI Brain Tumour Data-set in order to create the classifier. This data-set includes four classes of different types of brain tumour being:

1. Glioma: Gliomas are tumors that develop from glial cells, which are supportive cells in the brain. They are the most common type of primary brain tumour. Gliomas can occur at any age and can be benign (non-cancerous) or malignant (cancerous).

2. Meningioma: Meningiomas are tumours that arise from the meninges, which are the protective membranes surrounding the brain and spinal cord. They are usually benign and grow slowly. Meningiomas are more common in women than in men and often occur in adults.

3. **No tumour**: In this case we get MRI's that have no tumour, which means that the brain of the patient is healthy.

4. Pituitary: Pituitary tumours develop in the pituitary gland, which is a small gland located at the base of the brain. Most pituitary tumours are non-cancerous, but they can still cause health problems by interfering with the normal hormone production of the pituitary gland.

Imbalanced data, where one class dominates over others, can lead to biased model predictions, poor generalisation, and to check if this problem is present in this data-set we have to get the cardinality of images for each one of the four categories.

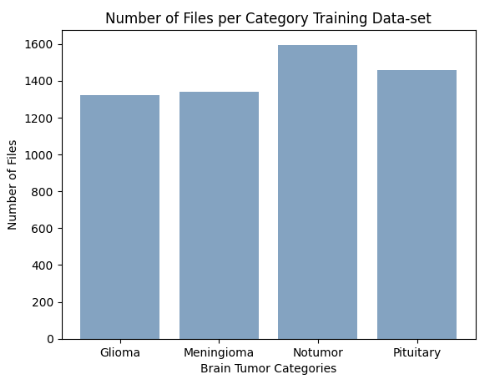

Figure 1: Training Data Distribution

In the training data-set we can find that we have a difference between the no-tumour class and the rest of the data classes. The difference in this case is around 250 samples a notable difference, that we have to keep in mind. In the same manner we are going to check for the distribution of data for the testing data-set in order to ensure the quality of the evaluation of the model:

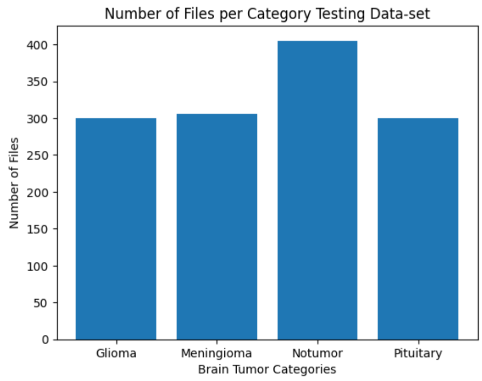

Figure 2: Test Data Distribution

In deep learning, image size plays a crucial role in model training. Varying image sizes can affect memory requirements, computational efficiency, and model performance, necessitating careful consideration and preprocessing to ensure compatibility and optimal results.

Starting with the training data we get the following image sizes:

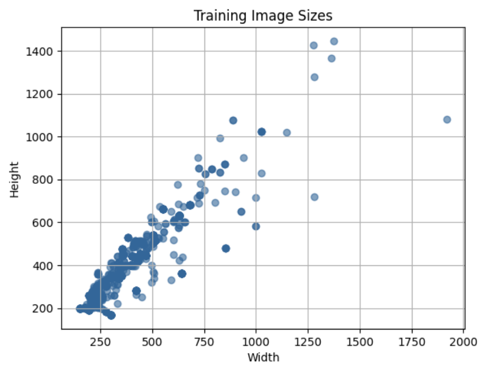

Figure 3: Training Data Sizes

Following the same strategy for the testing data we get the following results:

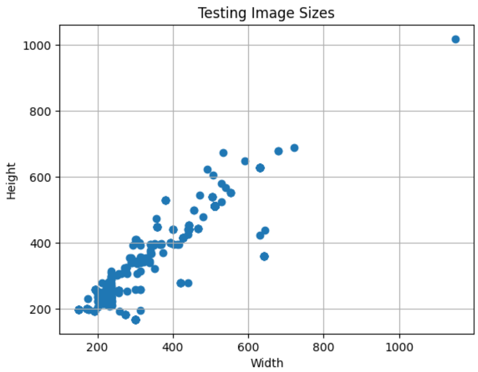

Figure 4: Training Data Sizes

3 Models

When comparing pre-trained models to CNN models for Explainable AI (XAI) purposes, several factors come into consideration. Pre-trained models usually are deep neural network architectures that has shown impressive performance in various computer vision tasks, thanks to its skip connections and residual blocks. It is often pre-trained on large-scale datasets like ImageNet, capturing diverse visual features.

Using pre-trained models for XAI provides the advantage of leveraging the learned representations and generalisation capabilities. With careful design choices, CNN models can be optimized for specific tasks like brain tumor classification, ensuring the inclusion of relevant image features and interpretability requirements.

The choice between pre-trained and custom CNN models for XAI depends on factors like task complexity, data-set size, interpretability needs, and available computational resources. While pre-trained models provide a strong starting point, custom CNN models can offer tailored interpretability and performance, allowing fine-grained control over the model's architecture and decision-making process.

3.1 CNN Model

A Convolutional Neural Network (CNN) is a type of deep learning model specifically designed for processing grid-like data such as images. It is inspired by the visual processing mechanism of the human brain. The CNN model consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters to the input image, capturing local patterns and features.

The pooling layers downsample the output of the convolutional layers, reducing the spatial dimensions and extracting the most important features. The fully connected layers further process the extracted features and perform the final classification. The CNN model learns the weights and biases of these layers through training on a large data-set, optimising them to make accurate predictions on unseen data. With their ability to automatically learn hierarchical representations from raw input data, CNNs have been highly successful in various computer vision tasks such as image classification, object detection, and image segmentation.

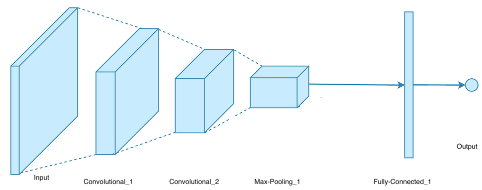

Figure 5: CNN Model

CNN's hierarchical structure learns patterns at different levels, aiding in tumor detection. Explainable AI (XAI) methods, like Grad-CAM, can be applied to CNNs to interpret the model's predictions by visualizing important regions in images, providing insights into the decision-making process and increasing trust in the model's outputs.

3.2 Resnet-50 Model

Comparing our model to the ResNet-50 model is of utmost importance, particularly when dealing with a medical data-set focused on brain tumours and striving for exceptional accuracy metrics. ResNet-50 is renowned for its effectiveness in image classification tasks, making it a widely accepted benchmark in the field. By comparing our model's performance against ResNet-50, we can assess its ability to accurately identify and classify brain tumours. This comparative analysis enables us to gauge the strengths and weaknesses of our model, identify areas for improvement, and fine-tune our approach accordingly.

ResNet-50 introduced the concept of skip connections and residual learning. These connections facilitate the propagation of gradients and help address the vanishing gradient problem, making it easier for the network to learn and optimise deeper layers. Residual learning allows the network to focus on learning the residual or the difference between the input and output of a layer, which aids in faster convergence and improved performance.

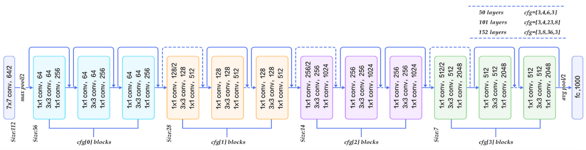

Figure 6: Resnet-50 Model Architecture

ResNet-50's depth and skip connections contribute to its excellent **generalisation** ability. It can learn high-level features that are transferable to various image recognition tasks, even when trained on different data-sets. This makes ResNet-50 a great choice for the classification task in this project.

4 Training

When training the two models, both the pre-trained ResNet-50 and the custom CNN model require similar data preparation steps. The data-set needs to be properly organised and split into training and test sets. Pre-processing steps such as resizing the images to a consistent size, applying normalisation, and transforming them into tensors are common for both models. The data normalisation techniques for the training and testing data are in both models are:

1. Resize image 224x224.

However, the major difference lies in the hyper-parameters. Hyperparameters like the number of layers, filter sizes, learning rate, batch size, and number of epochs can be adjusted based on the specific requirements of the custom model. Tuning these hyperparameters for the custom CNN model can play a vital role in achieving optimal performance. Knowing that information we had the following set-up for each model:

The **hyper-parameters** of the CNN model are the following:

1. Learning Rate: 0.001
2. Epochs: 25
3. Batch size: 128
4. Momentum = 0.9
5. Loss Function: Cross Entropy Loss
6. Optimiser Function: Stochastic gradient descent

The **hyper-parameters** of the Resnet-50 model are the following:

1. Learning Rate: 0.0001
2. Epochs: 5
3. Batch size: 32
4. Loss Function: Cross Entropy Loss
5. Optimiser Function: Stochastic gradient descent

The results of the experiments indicate remarkable performance for both models. The pre-trained ResNet-50 model showcases impressive accuracy, reaching close to 100% within just 5 epochs. Its ability to leverage pre-trained weights and learned representations contributes to its rapid convergence. On the other hand, the custom CNN model achieves a commendable accuracy of 90% after 25 epochs, demonstrating its effectiveness in the given task.

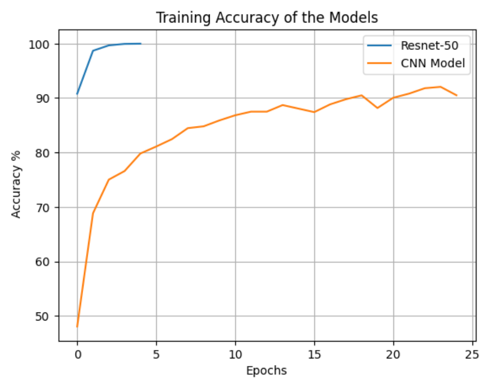

Figure 7: Models Training Accuracy

Despite taking slightly longer to converge, the CNN model's accuracy is still impressive, considering the complexity of the problem.

Notably, the training times of the two models differ by less than 100 seconds, highlighting their comparable efficiency. Overall, these results underscore the effectiveness of both models, with the ResNet-50 model achieving near-perfect accuracy and the CNN model performing remarkably well with 90% accuracy. Following the same principle we decided to also find the Loss both models have for the training data:

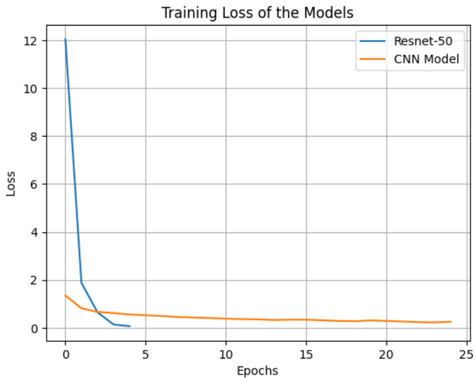

Figure 8: Models Training Loss

The confusion matrices obtained from the evaluation of a model on the test data, which the model has not seen during training, hold significant importance in assessing the model's performance. These matrices provide a comprehensive overview of the model's predictive capabilities across different classes. By examining the true positive, true negative, false positive, and false negative values, we gain insights into how the model is handling different classes and potential areas of improvement.

The evaluation of both models yielded highly impressive confusion matrices, showcasing their strong predictive capabilities. Notably, the ResNet-50 model's confusion matrix exhibited slightly superior performance compared to the custom CNN model. The custom CNN model's confusion matrix displayed commendable results, with accurate predictions and relatively fewer misclassifications.

The ResNet-50 model demonstrated exceptional accuracy in correctly classifying instances across various classes, with minimal instances of misclassifications. While the performance of both models was commendable, the ResNet-50 model's confusion matrix highlighted its ability to make highly accurate predictions and maintain a better balance across classes. These excellent confusion

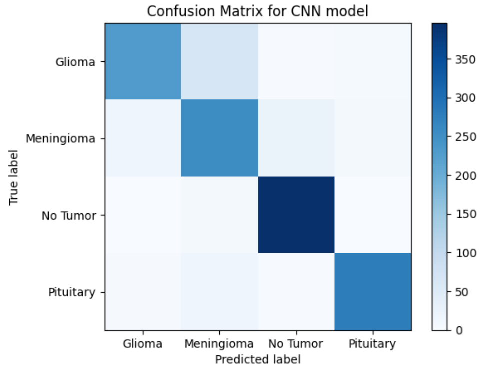

Figure 9: CNN Confusion Matrix

matrices validate the effectiveness of both models in the given task, with the ResNet-50 model showcasing a slightly better performance overall.

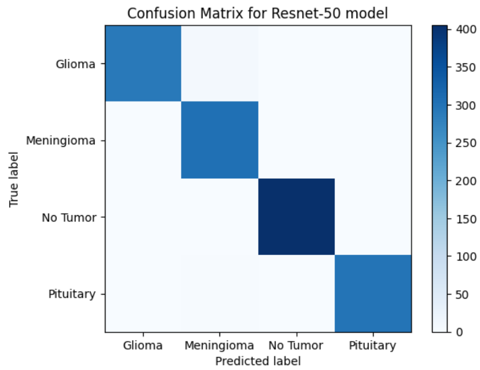

Figure 10: Resnet-50 Confusion Matrix

Explainable Artificial Intelligence (XAI) techniques aim to provide insights into the decision-making process of deep learning models. In this context, the custom CNN model excels with its slightly superior Grad-CAM implementation compared to the ResNet-50 model. The custom CNN's architecture allows for precise control over interpretability methods, resulting in a more fine-grained visualisation of important regions contributing to the model's predictions.

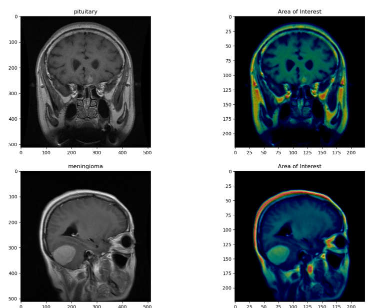

Figure 11: CNN - Grad-Cam Results

The Grad-CAM technique offers detailed and accurate visual explanations of the CNN model's attention, enabling users to gain a deeper understanding of the features and regions that drive its decisions. Although both models deliver impressive performance, the custom CNN's slightly better Grad-CAM implementation enhances its interpretability, making it an excellent choice for applications that require precise visual insights into the model's decision-making process.

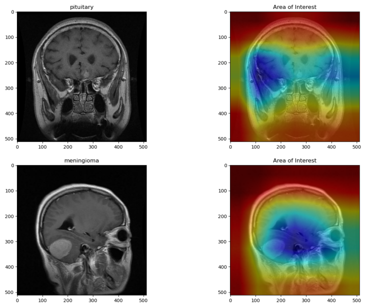

Figure 12: Resnet Grad-Cam Results

5 Conclusion

In conclusion, both the custom CNN and the pre-trained ResNet-50 models demonstrated exceptional performance in the given task. The ResNet-50 model showcased near-perfect accuracy within a short period, leveraging its pre-trained weights and learned representations. On the other hand, the custom CNN model achieved an impressive accuracy of 90% after a slightly longer training duration. Although the performance of both models was remarkable, the custom CNN model exhibited a slight advantage in terms of the Grad-CAM implementation. With its flexibility and tailored design, the custom CNN model allowed for a more fine-grained control over interpretability methods like Grad-CAM. This advantage enables better visualisation and understanding of the important regions contributing to the model's decision-making process. Hence, while both models delivered amazing performance, the custom CNN model offered a slightly superior Grad-CAM implementation, enhancing its interpretability and providing valuable insights into the decision-making process.

References

MRI Brain Tumour Data-set
Glioma
Meningioma
Pituitary
Resnet-50