Application of Neural Networks: Tumour

Matthew Fleischman (FLSMAT002)
Ariel Levy (LVYARI002)

October 2024

1 Introduction

A brain tumour is an abnormal growth of cells within the brain or its surrounding tissues. It is one of the most life-threatening forms of cancer, therefore having effective detection systems in place is of vital importance [1]. Brain tumours are often diagnosed via analysis of brain scans such as MRI, CT, and PET scans. However, human examination of scans is time-consuming and susceptible to human error.

1.1 Aims

The research aims to design a multi-class classification neural network to identify various forms of tumours from brain scan images. By automating the classification process, the model has the potential to serve as a supportive tool for medical practitioners, enhancing diagnostic accuracy and efficiency.

1.2 Dataset

The dataset consists of 3264 brain scan .jpg files [2], taken from various cross-sectional views. Each image is labelled according to the presence and type of tumour, facilitating the supervised classification task for the model.

1.3 Objectives and Metrics

The primary objectives of this investigation are:

- To develop a neural network model capable of classifying brain scan images into four categories, as described in section 2.1, achieving an accuracy that far surpasses random chance (25%).
- To evaluate the model's performance using metrics such as classification accuracy, F1 score, degree of overfitting, and training time.
- To analyse the model's strengths and weaknesses based on said metrics and discuss potential improvements.

1.4 Ethical Considerations

While AI offers several benefits in brain tumour classification, several ethical concerns must be considered.

- An overdependence on the AI model may develop, resulting in medical practitioners being unable to make brain tumour diagnoses by themselves.
- Data privacy is crucial to patient confidentiality laws [3]. Medical images contain sensitive information that must be protected.
- An experienced neurologist or radiologist may take into account nuanced details specific to the patient or case that go beyond the brain scan images themselves. These critical factors may be overlooked by the AI model.
- Biases in the training dataset could create disparities in healthcare outcomes for different patient groups. The potential for unequal performance of the AI model must be considered before use.

2 Data Preparation

2.1 Input-Output Definition

The input data consists of brain scan images. These images show different cross-sections of the brain and vary in resolution and colour depth. To standardise the data, the images are preprocessed by being resized to a fixed resolution (64x64) and converted into a single-channel colour space (greyscale). This ensures consistency across the dataset and reduces the model's computational complexity. In addition, greyscale pixel values were normalised from their typical range of 0 to 255 to a scale of 0 to 1.

The output is a multi-class classification label, predicting the presence and type of brain tumour in the scan. There are four possible output classes:

1. Meningioma tumour
2. Glioma tumour
3. Pituitary tumour
4. No tumour

2.2 Data Splitting

The 3264 labelled brain scans are split into three datasets: training, validation, and testing. The training set (68%), which constitutes most of the data, was used to train the model and optimise its parameters. The validation set (17%) was used during model development to tune hyperparameters and avoid overfitting by assessing the model's performance on unseen data. Finally, the testing set (15%) was reserved for the final evaluation of the model to gauge its real-world performance.

3 Baseline

3.1 Model

The baseline model is a simple convolutional neural network (CNN) that consists of just six layers [4]:

- **Input** – A 64x64x1 layer, representing the pixels of the preprocessed images.
- **Convolutional** – A layer with filters (e.g. 32 filters of size 3x3), followed by an activation function (e.g. ReLU), which extracts features from the input.
- **MaxPooling** – A downsampling layer which reduces the spatial dimensions of the feature maps (e.g. a window of size 2x2), retaining only the most important information.
- **Flatten** — A layer that converts the 2D feature maps into a 1D vector to prepare for the fully connected layers.
- **Dense** — A fully connected layer with a certain number of neurons (e.g. 64 neurons), activated by ReLU.
- **Output** — A dense layer with 4 neurons (one for each class), using a softmax activation function.

It utilises the same optimiser and hyperparameters established for the final model, as outlined in section 4.

3.2 Performance

The baseline model demonstrates strong performance with training, validation, and test accuracies of 96.8%, 84.1%, and 85.5%, respectively. However, the higher training accuracy compared to the validation and test accuracies suggests some degree of overfitting, which could be addressed in a more robust model.

4 Model Development

4.1 Architecture

The final CNN model builds on the baseline structure with several key enhancements: three sets of convolutional/max-pooling layers are introduced to progressively extract more complex features, and dropout layers are incorporated after all convolutional and dense layers to mitigate overfitting and improve generalisation. The structure of the final CNN is illustrated in Figure 1.

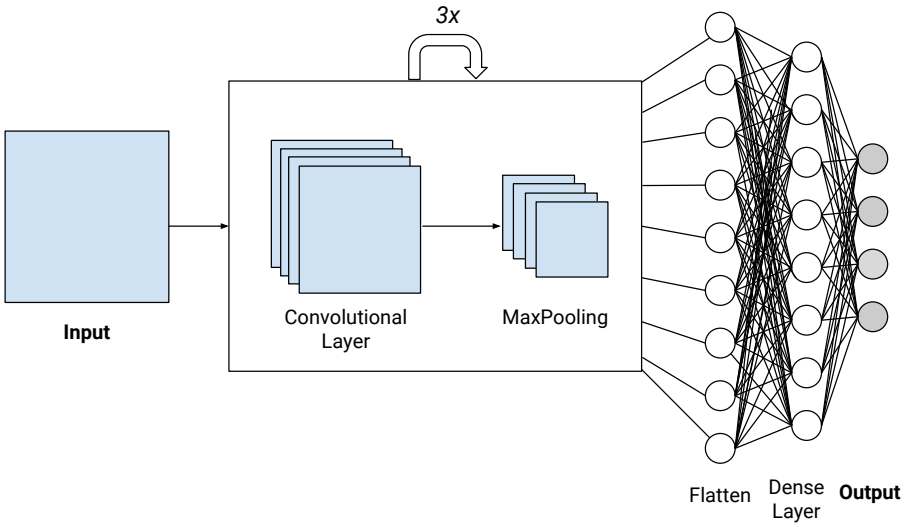

Figure 1: Architecture of CNN without Dropout.

Dropout is a regularisation technique that combats overfitting in a neural network layer by randomly deactivating a percentage of neurons, in a layer, during training. This process encourages the network to learn more robust features, and after training, its weights are scaled to account for the dropped neurons, ensuring stable performance [5].

4.2 Hyperparameter Tuning

Random search was employed to optimise the hyperparameters, with a focus on tuning the batch size and learning rate. Early stopping was used as the stopping criterion, monitoring changes in the validation loss to halt training when improvements became negligible, thus preventing overfitting. The hyperparameter values used in the final CNN are shown in Table 1.

Table 1: CNN Hyperparameters

Hyperparameter	**Value**
Learning rate	0.001
Batch size	64
Epochs	30

4.3 Training

The model was trained using the Adam optimiser, a widely used adaptive stochastic gradient descent algorithm. The loss function selected for this multi-class classification task was sparse categorical cross-entropy. Furthermore, accuracy – a suitable performance metric for classification tasks – was used to track the model's ability to correctly classify scans during training and validation.

5 Results and Analysis

The performance of the final CNN model for brain tumour classification was evaluated across training, validation, and testing datasets. This section presents the results for the considered metrics and compares the performance of the final CNN to the baseline model. The strengths and potential shortcomings of the trained model are discussed.

5.1 Performance of the Final CNN

The final CNN demonstrated high accuracy and F1 scores across the datasets, as shown in Table 2. The F1 score of 0.908 for the testing dataset implies that the model generalises well to new, unseen data, achieving a good balance between precision and recall.

Table 2: Accuracy and F1 Scores Achieved by the Final CNN

Dataset	**Accuracy (%)**	**F1 Score**
Training	98.02	0.980
Validation	89.37	0.893
Testing	90.82	0.908

In Figure 2 it can be seen that the model's accuracy improves consistently over the 30 epochs, with both the training and validation accuracy curves converging toward values above 88%. The high validation accuracy indicates that the model generalises well without signs of overfitting.

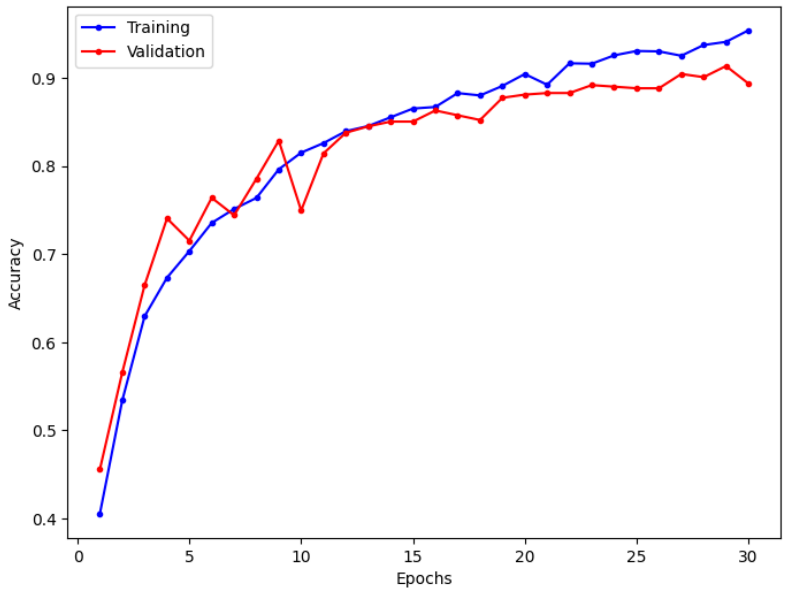

Figure 2: Classification Accuracy Achieved by the Final CNN for the Training and Validation Datasets over 30 Epochs.

Figure 3 shows the evolution of the loss function over the training process. The loss curves for both datasets exhibit an exponential decay, as expected. The loss curve of the validation dataset flattens after approximately 20 epochs, suggesting that the model has reached a point where further improvements are marginal.

5.2 Comparison to Baseline Model

The baseline model achieved a training accuracy of 96.83%, and a testing accuracy of 85.51%. Comparing the accuracy values from Table 2 to those of the baseline model, the final CNN showed an increase in training accuracy by 1.09% and an increase in testing accuracy by 5.31%. Therefore, the final CNN outperformed the baseline model, indicating better generalisation and robustness.

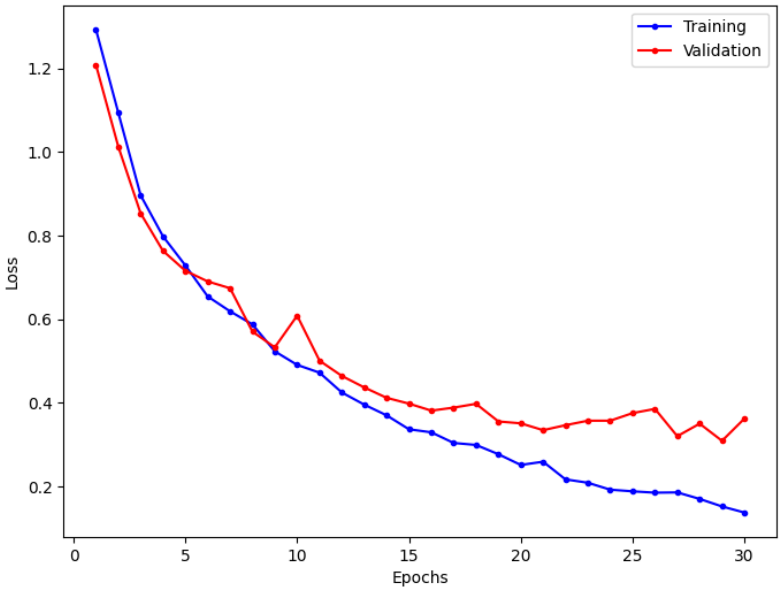

Figure 3: Loss Function Achieved by the Final CNN for the Training and Validation Datasets over 30 Epochs.

6 Conclusions

This research designed and developed a neural network for multi-class classification of brain tumours using MRI scans, aimed at improving diagnostic accuracy and reducing human error. The final CNN achieved higher accuracy and F1 scores, outperforming the baseline model by better generalising to unseen data. These improvements were driven by the addition of convolutional layers, dropout, and hyperparameter tuning, which effectively mitigated overfitting.

Expanding the classification to include more tumour types could further improve this neural network's utility. Future work should integrate the model with expert feedback to ensure nuances and patient-specific factors are not overlooked. Before real-world application, ethical considerations like dataset biases and over-reliance on AI must be addressed. A notable limitation of this model is the absence of uncertainty quantification in outputted classifications.

References

[1] Javaria Amin, Muhammad Sharif, Anandakumar Haldorai, Mussarat Yasmin, and Ramesh Sundar Nayak. Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex & intelligent systems*, 8(4):3161–3183, 2022.

[2] Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, and Swati Kanchan. Brain tumor classification (mri), 2020. URL https://www.kaggle.com/dsv/1183165.

[3] MPS. Aspects of confidentiality- High-profile patients — medicalprotection.org. https://www.medicalprotection.org/southafrica/casebook/casebook-may-2014/casebook-january-2014/aspects-of-confidentiality-high-profile-patients, 2014. [Accessed 09-10-2024].

[4] IBM. What are Convolutional Neural Networks? https://www.ibm.com/topics/convolutional-neural-networks, 2024. [Accessed 09-10-2024].

[5] Seb. Dropout regularization in neural networks: How it works and when to use it. *Programmathically*, 2021. URL https://programmathically.com/dropout-regularization-in-neural-networks-how-it-works-and-when-to-use-it/. [Accessed 14-10-2024].