Detection of Brain Tumours using Image Processing and Deep Features

Ali Mirza and Savas Konur

Department of Computer Science, University of Bradford,
Bradford, BD7 1DP, UK

Abstract. This paper presents the gaps/limitations of some existing approaches to classify brain tumour MRI images using AI methods and tools. A deep-learning (DL) methodology is proposed and then implemented for multiclass classification. The proposed deep-learning methodology encompasses deep features from the VGG-16 pre-trained model and GLCM features along with image processing and dimensionality reduction. The DL model achieved an accuracy score of 98.20% and a Matthews correlation coefficient score of 0.9736 with a dataset of 54,000 images.

Keywords: Brain tumours · Multiclass classification · Deep features · GLCM · Deep learning · Transfer learning

1 Introduction

Brain tumours occur due to the uncontrolled growth of abnormal cells in the brain. The unchecked growth of cells leads to either cancerous (malignant) or non-cancerous (benign) tissue forming into primary or secondary brain tumours. Tumours which begin from the brain tissue are primary brain tumours, whereas those that occur from cancer spreading to the brain from another area of the body (metastasis) are known as secondary brain tumours [8].

Despite the extensive research into DL approaches for brain tumour detection, challenges and limitations in the techniques used still remain. Specifically, existing methodologies for brain tumour classification lack the combination of deep features extracted from a pre-trained model and features from grey-level co-ocurrence matrices [1]. Another limitation of the current techniques used is the limited size of the datasets being used to train neural network classifiers as most datasets that are used tend to have fewer than 30,000 images [1]. Furthermore, existing state of the art solutions do not implement the global average pooling layer [5]. Research seems to agree that in terms of statistical feature extraction of images, second-order features alone deliver better results than first-order features as well as a combination of first and second-order features [11]. [9] presented a comprehensive comparison of the effects of first and second-order features on brain tumour classification accuracy of different forms of the SVM classifier. For instance, [1] employed a grey level run length matrix (GLRLM) and grey level cooccurrence matrix (GLCM) for second-order texture features such as coarseness, energy and contrast etc. [5] found that combining first-order, second-order and, wavelet transform feature sets returned better results for each classifier except Naïve Bayes, as opposed to only using second-order texture features. In particular, [5] highlighted an increase of 6.78% for the multi-layer perceptron classifier when the 38 combined features are used for each MRI modality like flair, T1 and T2. However, [5] did not apply their combined features methodology for a DL approach. A recurring gap in all the literature discussed above is the lack of a combination of deep and hand-crafted features alongside image-processing steps with a DL model that utilises transfer learning. [11] also emphasised this specific gap in their paper.

This paper addresses the aforementioned gaps and limitations by proposing a methodology which utilises a combination of image processing techniques and a combined set of features consisting of GLCM features and deep features from the application of the global average pooling operation on the output of the fourth max-pooling layer of the VGG16 model. Our studies have shown that the model trained on a combination of GLCM and deep features outperforms both traditional ML algorithms and neural network classifiers which were trained on only GLCM features or on only deep features.

2 Methodology

The proposed methodology is presented in Fig. 1. Brain MRI images were resized to the dimensions required for VGG-16 and augmentation was applied to reduce imbalance. In the transfer-learning-based stage, the output from the fourth max pooling layer of the VGG-16 CNN architecture was used as the set of deep features.

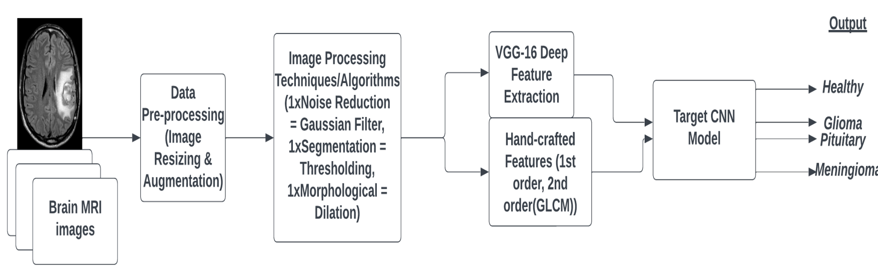

Fig. 1. Proposed Methodology

2.1 Tools and Packages

The Python programming language was used alongside several of its libraries and packages. All aspects related to the model were implemented on the computing platform/IDE known as PyCharm however, the kernel used for model development was a remote kernel provided through a private cluster by Paperspace. The private cluster which was used for development consisted of an Nvidia RTX A4000 GPU, 90GB of ram and a 16-core CPU.

2.2 Dataset and Image Processing

The images in the dataset were collected from multiple sources like the Kaggle website and the REMBRANDT dataset [10,7,3]. It was imperative that any duplicates were removed by assigning images their unique hash codes and if images were found to have identical hash codes, the duplicate(s) were removed. To increase the final size of the dataset to 54,000 and eliminate any data imbalance image augmentation was applied. The decrease in data imbalance was vital as most classification algorithms assume that the data they are trained on is balanced. Therefore, the weights assigned to each of the samples and classes are equal but if the data is imbalanced, the classifier will be more biased towards the majority class. Furthermore, the issues surrounding data imbalance could be propagated in the implementation as it involves multiclass classification and could lead to a tumour type belonging to a minority class not being detected. Noise in images can lead to classification algorithms deducing patterns in the noisy pixels and eventually start making predictions based on the existing patterns in noisy data. We made the choice to apply the Gaussian spatial filter mathematical model (1) to reduce noise in the dataset.

$$G(i,j,k) = \frac{1}{(\sqrt{2\pi})^3 \sigma i \sigma j \sigma k} \rceil^{-(\frac{i^2}{2\sigma_i^2} + \frac{j^2}{2\sigma_j^2} + \frac{k^2}{2\sigma_k^2})} \tag{1}$$

The Gaussian filter was used as opposed to other filters because it can preserve image structure post-processing as opposed to the median filter which leads to a decrease in image structure. Threshold-based image segmentation which comprises the partitioning of pixels according to their intensity when compared to a set threshold value was applied. The specific threshold method used was THRESH_BINARY_INV which converts each image into a binary image by replacing pixel values that are less than the threshold value of 70 to 125 i.e., grey. Thresholding allows for the unnecessary areas of the cerebral MRI image to be removed and only the regions of interest to remain. The morphological operations, erosion and dilation were employed to reduce any imperfections that remain following image segmentation. Erosion removes pixels from the boundaries of objects post-image segmentation by iterating over the image and setting each pixel value to the minimum value of its respective neighbours. As a consequence, erosion allowed us to eliminate any unusual areas that were protruding from the tumour area of the MRI image. Dilation sets each pixel of an image to the maximum value of its neighbouring pixels. Dilation allowed for the removal of small gaps which existed in the region of interest (ROI) and also increased the size of the segmented object.

2.3 Feature Extraction Methods and PCA

Deep features were extracted using the output of the fourth max pooling layer of the pre-trained VGG-16 model. A global average pooling layer was appended to the end of the VGG-16 model to flatten the output of the fourth max pooling layer and to generate one feature map for each filter/type of feature. The decision to append a global average pooling layer to the VGG-16 model was due to its ability to categorise the hundreds of features by finding associations within the thousands of feature maps that are generated from the max pooling layer. The deep features generated using the global average pooling layer succeeding the max pooling layer allowed for the decrease of the complexity of the neural network classifier and as a consequence, avoid overfitting. Energy, correlation, dissimilarity, homogeneity and contrast were extracted for 5 different distances and angles using GLCM. The dimensions of the concatenated features were reduced through implementing PCA. Prior to dimensionality reduction, the combined features were scaled to a range individually using a min-max scaler. All components of the scaled features were plotted against the percentage variance to determine and select the optimum number of components without a significant loss of variance. The application of PCA to reduce the dimensions of the combined feature set reduced training time and increased the ability of the model to generalize to unseen instances due to a reduction in noise/outliers. 700 components were selected with a maximum variance loss of only 0.27%.

2.4 Neural Network Classifier

The neural network classifier encompasses only two layers that are fully connected and a dropout layer between the two dense layers. The first of the two dense layers has only 60 neurons and implements the ReLU activation function to remove all negative values from the concatenated feature set post-dimensionality reduction with PCA. The decision to only add 60 neurons to the first dense layer was due to a potential risk of the neural network overfitting the training feature set. Furthermore, 60 neurons were selected due to 60 being the optimum level of high training and high validation accuracies, suggesting no occurrence of overfitting. The optimum number of neurons was determined through hyperparameter tuning using a grid-search of parameters including optimizers, neurons and batch sizes. The ReLU activation function was implemented due to its simple computation of a gradient when compared to more complex activation functions such as the hyperbolic tangent function. The lower complexity of ReLU allowed for faster learning speeds and therefore, a reduction of computation time. The experiments of [2] supported the findings from our experiments involving hyperparameter tuning. The final dense layer includes 4 output neurons i.e., one neuron for each class that was predicted. The softmax function outputs the probability of each predicted image being in each of the 4 output classes. Finally, the dropout layer was inserted into the neural network architecture to reduce overfitting through setting some inputs to 0 at random with a frequency rate specified as 0.4 i.e., 40% of the inputs to the layer are set to 0 at random.

3 Results and Discussion

Classifiers	Accuracy with Image Processing	Accuracy without Image Processing
SVM	84.14%	76.68%
RF	88.55%	91.26%
KNN	83.12%	81.81%
Proposed DL model	89.34%	98.20%

Table 1. Accuracy scores of 3 traditional classifiers and proposed deep-learning classifier.

	DL model with Image Processing	DL model without Image Processing
Accuracy:	89.34%	98.20%
F-1 score:	0.8901	0.9804
Matthews Correlation Coefficient:	0.8846	0.9736
Recall:	0.8566	0.9804
ROC/AUC Score:	0.8532	0.9986

Table 2. Results for each performance measure of proposed DL model.

Our results are presented in Table 1 and 2. The decrease of 2.71% (see Table 1) for RF could be due to the Gaussian filter as even though it reduces noise efficiently, it could be reducing fine details in the images due to the blurring effect. SVM is affected by noise due to target classes overlapping and the fact that noise can cause the SVM's separating margin to be a stochastic variable [6]. The proposed DL model, outperforms all three traditional classifiers in every performance metric which was measured (see Table 2) due to several reasons. For instance, the DL model was trained on a larger dataset than the traditional ML classifiers consisting of 54,000 cerebral MRI images. The larger dataset allowed for the neural network to not overfit the training combined feature set. Another potential reason for the discrepancy in performance between the proposed DL model and the traditional classifiers is that with the use of transfer learning in the DL model, i.e. deep feature extraction, allowed for more abstract and complex features like the specific tumour type to be detected and segmented in feature maps. After rigorous testing of the DL model with unseen images, a decrease in the accuracy of predictions was observed with certain images. We recognised the fact that cerebral MRI images of certain modalities were predicted with less accuracy than others. This difference in prediction accuracy between modalities could be due to modalities like T2 having fewer representations in the dataset due to a lack of T2 MRI images available online for use over other modalities.

4 Conclusions and Future Work

To conclude, the DL model detects the presence and form of a brain tumour in images accurately within seconds. The DL model also returns the percentage confidence of the prediction using the output from the softmax layer of the final layer of the neural network classifier. However, the model does not return the correct number of predictions between different modalities proportionately. This difference in performance between modalities shows that the lack of MRI images is still a very significant factor in the generalization abilities of brain tumour classification models. Future work could involve training the same model implemented using the proposed methodology on an even larger dataset which includes images from a range of different modalities to further increase the generalization ability of the DL model.

References

1. Abbas, H. K., Fatah, N. A., Mohamad, H. J. and Alzuky, A. A. Brain Tumor Classification Using Texture Feature Extraction. Bristol, 2021. Vol. 1892. IOP Publishing. https://go.exlibris.link/8CFVfcjQ.
2. Alhassan, A. M. and Zainon, W. M. N. W. (2021) Brain tumor classification in magnetic resonance image using hard swish-based RELU activation function-convolutional neural network. *Neural computing & applications* 33 (15), 9075-9087.
3. Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S. and Kanchan, S. (2020) Brain Tumor Classification (MRI).
4. Gelvez, E., Vera, M., Huérfano, Y., Valbuena, O., Salazar, W., Vera, M. I., Borrero, M., Barrera, D., Hernández, C., Molina, Á. V., Martínez, L. J., Sáenz, F., Vivas, M., Vanegas, J., Contreras, J., Restrepo, J., Salazar, J. and Contreras, Y. (2018) Smoothing filters in synthetic cerebral magnetic resonance images: A comparative study. *Revista latinoamericana de hipertensión* 13 (4), 335-338.
5. Latif, G., Iskandar, D. N. F. A., Alghazo, J. M. and Mohammad, N. (2019) Enhanced MR Image Classification Using Hybrid Statistical and Wavelets Features. *IEEE access* 7, 9634-9644.
6. Li, H.-X., Yang, J.-L., Zhang, G. and Fan, B. (2013) Probabilistic support vector machines for classification of noise affected data. *Information Sciences* 221, 60-71.
7. Nickparvar, M. (2021) Brain Tumor MRI Dataset.
8. *Primary and secondary brain tumours.* (2019) Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/brain-tumours/types/primary-secondary-tumours. Accessed 13/01/23.
9. Priya, K. M., Kavitha, S. and Bharathi, B. (2016) Brain tumor types and grades classification based on statistical feature set using support vector machine. *2016 10th International Conference on Intelligent Systems and Control (ISCO).* 7-8 Jan. 2016.
10. Scarpace, L., Flanders, A. E., Jain, R., Mikkelsen, T. and Andrews, D. W. (2019) *Data From REMBRANDT*. The Cancer Imaging Archive.
11. Shoaib, M. R., Elshamy, M. R., Taha, T. E., El-Fishawy, A. S. and Abd El-Samie, F. E. (2022) Efficient deep learning models for brain tumor detection with segmentation and data augmentation techniques. *Concurrency and computation* 34 (21), n/a-n/a.