

Brain Tumor Multi-Classification using Convolutional Neural Network

Sandhya Rao P¹, Shwetha Goldie U², Viswapujitha S³, Prof.V.Rajalakshmi⁴

^{1,2,3}B.E. students, Department of CSE, Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu, India.

⁴Assistant Professor of Sri Venkateswara College of Engineering, Dept. of Computer Science and Engineering, Sriperumbudur, Tamil Nadu, India

Abstract - Brain tumor classification is a pivotal task in detecting and classifying tumors to make a decision regarding treatment according to their classes. Among the available imaging modalities, Magnetic Resonance Imaging is preferred for having superior image quality and does not ionize radiation. Convolutional Neural Network, a class of Deep Learning, is more commonly applied to analyzing visual imagery attributing to its automated feature engineering. In this paper, a Deep Learning model based on CNN is proposed to differentiate between the three glioma grades (Grade II, Grade III, and Grade IV) using the publicly available REMBRANDT dataset. This dataset includes 21223 images for 73 patients.

Key Words: Machine learning, Convolutional neural networks, MRI, Brain tumor classification.

1. INTRODUCTION

A brain tumor is a mass or growth of abnormal cells in your brain and is one of the most dangerous causes of death among people with a survival rate of only 39%. If cancer spreads, effective treatment becomes more difficult, a person's chances of surviving are much lower. Brain tumor classification is a crucial task to evaluate the tumor at the early stage and make a treatment decision according to their classes. Magnetic Resonance Imaging (MRI) is one of the most used and popular for brain tumor diagnosis. The MR images are pre-processed and a deep learning model based on convolution neural networks is proposed to classify the different brain tumor glioma grades.

1.1 Brain tumors

Brain tumor is the abnormal growth of brain cells. The linked mortality and morbidity and the significant proportion of affected individuals has a major relevance on the death-adjusted life years contrasted to other malignancies although comparatively rare. The cases of brain tumor in India ranges from 5 to 10 per 100,000 populations with an increasing trend and accounts for 2% of malignancies [1]. The degrees of malignancy ranges from benign to aggressive. Each type of tumor has its own biology, treatment, and prognosis with different risk factors. Even "benign" tumors can be lethal due to their site in the brain, their ability to infiltrate locally, and their propensity to transform to malignancy.

This makes the classification of brain tumors a difficult science and creates problems in describing the epidemiology of these conditions [2]. The major ratio of adult tumors are supratentorial and gliomas contribute a majority of 86% namely astrocytomas, glioblastomas, oligodendroblastomas, and unspecified gliomas. Doctors use many tests to diagnose a brain tumor ranging from biopsy to imaging tests which unfortunately takes a long time where time is of essence.

1.2 Magnetic Resonance Imaging

Magnetic resonance imaging is used in radiology as a medical imaging technique. It forms images of the anatomy and physiological processes of the body and provides information that is different from other imaging modalities. Its major technological advantage is that it uses the physical and biochemical attributes of the tissues to identify and differentiate between them. MR image acquisition does not use ionizing radiation and hence does not have adverse health effects [3]. It requires little patient preparation and is non-invasive, patient acceptability is high.

1.3 Machine Learning

Machine learning (ML) is used to perform a specific task using patterns without human interaction by the study of statistical models and algorithms. ML has been widely used in the medical field under Artificial Intelligence. Most of the algorithms used are broadly classified into supervised and unsupervised and further into classification, regression or reinforcement. The proposed model is a supervised classification model that trains to find a mapping function of given input variable to either one of three output labels to predict new subject labels.

1.4 Convolutional Neural Network

Deep Learning is a subset of machine learning where artificial neural networks, algorithms inspired by the human brain, one of the most popular types of deep neural networks is known as convolutional neural networks (CNN). The key factor that makes this CNN architecture well adaptive to 2D data processing such as images is that it convolves learned features with input data, and uses 2D convolutional layers. CNN learn to detect different features of an image using tens or hundreds of hidden layers which is its biggest advantage. Every

hidden layer increases the complexity of the learning image features [4].

While other learning algorithms or models can also be used for image classification, CNN has emerged as the best model of choice for various reasons. These include the varied usage of the convolution operator in image processing. The benefits resulted by standard neural network training and the convolution operation to efficiently classify images is done implicitly by CNN. Further, being a neural network, CNN is also scalable for large datasets, which is often the case when images are to be classified.

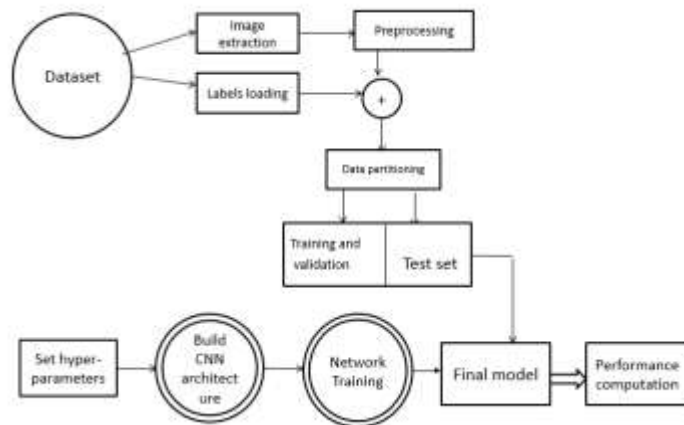


Fig-1: Block diagram of the proposed model.

2. METHOD

The block diagram of the proposed method is shown at FIGURE1 in which the system initiates by extracting and loading the dicom images along with its labels from the raw files of the REMBRANDT dataset. After preprocessing and shuffling technique the dataset is split into training and validation sets followed by which the proposed method is introduced along with setting the hyper parameters and an optimization algorithm. Lastly, the performance computations are shown.

2.1 Dataset

The dataset used in this method is obtained from The Cancer Imaging Archive (TCIA) public access repository [5]. MRI multi-sequence images from 130 patients with varied ages, grades, cases and races is contained in The Repository of Molecular Brain Neoplasia Data (REMBRANDT). These T1-weighted contrast-enhanced images contain different grades of glioma - Astrocytoma, Oligodendroglioma, Glioblastoma (Grade II,III,IV) as shown in FIGURE 2. Supplementary details about the description of the REMBRANDT dataset are shown in Table 1.

Table -1: Dataset description.

Tumor Category	Number of Patients	Number of tumor slices
Grade II	33	205
Grade III	19	129
Grade IV	21	182

2.2 Pre-processing stage

A preprocessing step is required before feeding the images into the proposed structure. The raw dataset contains varied images of pixel sizes - $N*N*1$ where $N=256, 288, 432, 512$. These images were downsampled to a uniform size of $256*256*1$ using bi-cubic interpolation for direct down sampling as it produces least noise relatively [6]. This down sampling is done to decrease dimensionality and hence computations resulting in direct calculations and helping the network to lower the time to show a better performance. To prevent focusing on tapered band of the whole dataset and to maintain the training on unsorted data the data is shuffled before splitting. Finally the data of 21223 images is divided into two sections: training and validation

2.3 Proposed CNN architecture

The proposed CNN architecture consists of 11 layers as shown in figure 2. It starts with the input layer that holds the images obtained from the pre-processing step passing through the convolution layers and their activation functions that are used in feature selection and down-sampling. A dropout layer is used to prevent overfitting followed by a fully connected layer and a softmax layer to predict the output. A classification layer is the final layer that gives the predicted class.

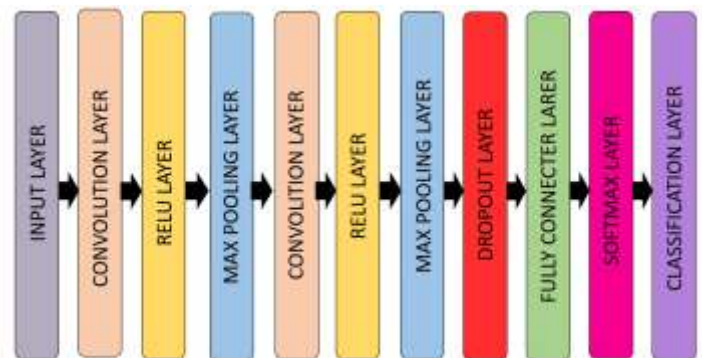


Fig-2 : The proposed CNN architecture.

First, Input layer should confirm the image data and its proposed size. In the proposed model, two convolutional layers are used. Each convolutional layer makes use of a set of K learnable filters or kernels.

These filters are used to detect the presence of specific features or patterns present in the original image which is the input. These kernels are usually expressed as a matrix of MxN. This filter is convolved at steps (S) across the width and height of the input file along with padding to retain edge information, and a dot product is computed to give an activation map. Varied filters to detect different features are convolved on the input file. A set of activation maps is given as output and is passed as input to the next layer in the CNN. FIGURE 3 shows an example of applying a kernel of size 2 × 2 (appears in blue) over a 3 × 3 image producing the same input dimensions of 3×3 after kernel sliding and dot product. The involved parameters we have used are; K=5,5, M×N=8×8, 16×16 and 2×2, S=[1, 1], P = 2,2 for the convolutional layers 1 and 2 respectively.

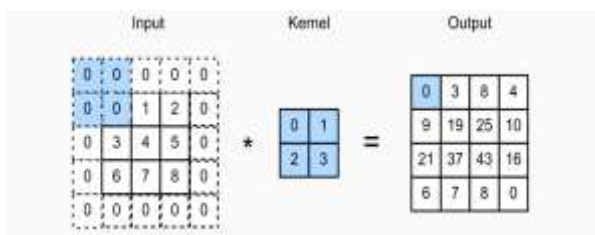


Fig -3: Example of convolutional layer.

A non-saturated activation function called ReLU follows every convolutional layer due to its better performance compared to other activation functions [7]. This rectified linear activation function will output either zero if negative or the input itself if positive. This piecewise-linear is easier to train and often achieves better performance. ReLU function is represented as

$$F(x) = \max(0, x) \quad (1)$$

The max pooling layer is that which calculates the maximum, or largest, value in each patch of each feature map. The output is either down sampled or is presented as pooled feature maps that emphasizes the most present feature in the patch, unlike the average presence of the feature in average pooling.

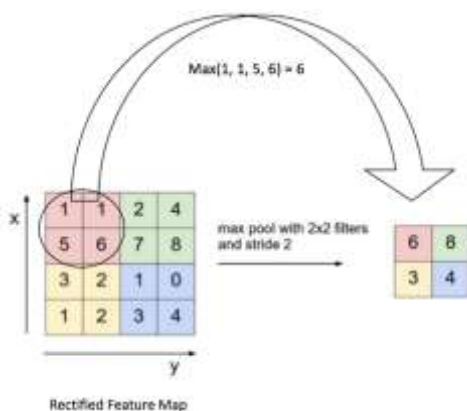


Fig -4: Example of a max-pooling layer (the maximum value out of a specific window is only considered).

The dropout layer is used for preventing overfitting in a neural network and is done by dropping out units both hidden and visible. Dropout refers to ignoring units or neurons during the training phase of which is chosen at random. In the proposed structure, we have found that 10% probability was most suited.

Finally, the advanced layers used are: Fully connected layer (FC), softmax layer and classification layer. The layers where all the inputs from one layer are connected to every activation unit of the next layer is called as a fully connected layer and is followed by the softmax layer also known as normalized exponential function.

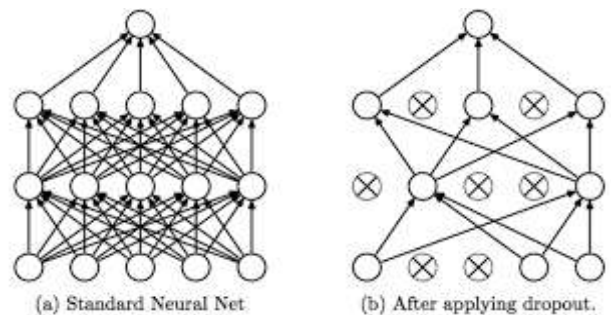


Fig -5: Example of a dropout layer.

Fully Connected layers in neural networks are the layers where all the inputs from one layer are connected to every activation unit of the next layer which is followed by the softmax layer which is also called the normalized exponential function.

The main aim of using a softmax layer is for the output probabilities range. The range is process to range from 0 to 1, where the sum of all the probabilities equals to one. In the proposed model, the softmax function is used for a multi-classification model that returns the probabilities of each class and the target class will have the high probability. The output of this layer can be calculated as:

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Finally, we use a classification layer that computes the cross entropy loss for multi-class classification problems with mutually exclusive classes and provides the final predicted class for each input image. Loss can be estimated from given equation, where p is the target labels vector, and q(x) is the output vector from the softmax layer.

$$H(p, q) = - \sum_x (p(x) * \log(q(x)))$$

2.4 Regularization and Optimization

Regularization is a method that tweaks the learning algorithm to make the model generalize better. This technique prevents learning a overly complex or flexible model to refrain from the risk of Overfitting.

Among the many techniques available for avoiding overfitting, data augmentation, dropout layers are used. To remove hidden units weights stochastically, dropout layers have been used. In L2 regularization, the values of weight matrices decrease because it assumes that a neural network with smaller weight matrices leads to simpler models and hence resulting in reduction of overfitting to an extent.

In L2 we have,

$$Cost\ function = Loss + \frac{\lambda}{2m} \cdot \sum ||w||^2$$

Here, lambda is the regularization parameter. It is the hyper parameter whose value is optimized for better results. Since it forces the weights to gradually decay towards zero, it is otherwise known as weight decay.

Optimization algorithms are also defined as error functions as it minimizes or maximizes an objective function. E(x) is a mathematical function that is dependent on the model's internal learnable parameters involved in computing the output values from the model's range of predictors. Adaptive Moment Estimation (Adam) finds the respective adaptive learning rates for every parameter. We have used this technique as it works favorably in practice and fares better in relation to other adaptive learning-method algorithms due to its fast convergence speed [8]. It rectifies all problems faced by other optimization techniques such as vanishing Learning rate, slow convergence or High variance in the parameter updates leading to fluctuating Loss function and the learning speed of the model is efficient and swift.

Table -2: Accuracy metrics for the proposed model.

Tumour Type	TP	TN	FP	FN	Precision	Sensitivity	Specificity	Accuracy
GRADE II	105	1004	21	93	0.834	0.530	0.915	90%
GRADE III	780	300	141	31	0.846	0.961	0.906	88%
GRADE IV	142	890	34	72	0.806	0.663	0.925	84%

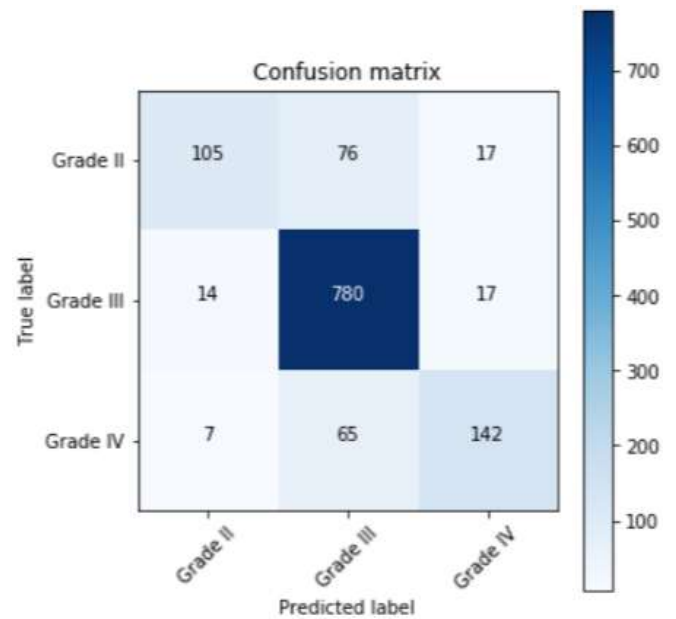


Fig -6: Confusion matrix for the proposed model.

3. CONFUSION MATRIX

FIGURE 6 shows the confusion matrix that summarizes the model's performance where the x-axis and y-axis are the true labels and predicted labels respectively. Table 2 shows the accuracy metrics as Precision, Sensitivity, Specificity and Accuracy extracted from the confusion matrix according to the following equation:

$$Precision = TP / (TP+FP)$$

$$Sensitivity = TP / (TP+FN)$$

$$Specificity = TN / (TN +FN)$$

$$Accuracy = TP+TN / (P+N)$$

where,

True Positive (TP) is the number of positive predicted cases and are actually positive.

True Negative (TN) is the number of negative predicted cases and are also actually negative.

False Negative (FN) is the number of negative predicted cases and are actually positive, also called (type two) error.

False Positive (FP) is the number of positive predicted cases and are actually negative, also called (type one) error.

We have achieved accuracy of 90.6% in classifying Grade II, 88.3% for Grade III and 84.3% for Grade IV.

4. TOOLS AND TIME CONSUMPTION

The proposed deep neural network structure is trained on *Inteli7-8550U CPU (1.8GHz), NVIDIA GeForce MX150 (4GB) GPU, 8GB RAM, Pytorch and Python. The training time was 150 minutes for 21223 images. The average test execution time was 25.44 milliseconds per image.

4. CONCLUSION

In this work, we have proposed a system for the classification of brain tumor MR images into different grades (Grade II, Grade III and Grade IV) using a custom deep neural network structure. The proposed network is constructed from 11 layers starting from the input layer having the preprocessed images going through the convolution layers and their activation functions (2 convolution, 2 ReLU and 2 Maxpooling layers). Also, to prevent overfitting a dropout layer is used followed by a fully connected layer and a softmax layer to predict the output. Lastly, a classification layer produces the predicted class. The proposed architecture has achieved the accuracy of 87.8% concerning the REMBRANDT dataset used in this paper.

5. REFERENCES

- [1] Dasgupta A, Gupta T, Jalali R, "Indian data on central nervous tumors: A summary of published work," *South Asian J Cancer*, 2016;5(3):147–153.
- [2] McKinney PA, "Brain tumours: incidence, survival, and aetiology," 2004 *Journal of Neurology, Neurosurgery & Psychiatry*, 75: ii12-ii17.
- [3] N. Kumar, "Health effects of non-ionized electromagnetic radiation", 2008 *International Conference on Recent Advances in Microwave Theory and Applications*, Jaipur, 2008, pp. 276-2761.3.
- [4] Y. Bengio, G. Hinton, Y. LeCun, "Deep learning,'" *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [5] K. Clark, J. Freymann, J. Kirby, P. Koppel, D. Maffitt, S. Moore, S. Phillips, M. Pringle, F. Prior, K. Smith, L. Tarbox, and B. Vendt, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [6] Debin Zhao, Jian Zhang, Ruiqin Xiong, Wen Gao, and Yongbing Zhang, "Interpolation-Dependent Image Downsampling", *IEEE Transactions on Image Processing*, vol. 20, no. 11, November 2011.
- [7] G. Shen and Y. Yuan, "On Theoretical Analysis of Single Hidden Layer Feedforward Neural Networks with Relu Activations," 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 2019, pp. 706-709.
- [8] Y. Lei, Y. Sun, J. Xing, C. Xue, and J. Zhang, "Research on Edge Detection Operator of a Convolutional Neural Network," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 49-53.