Development of Artificial Intelligence Algorithm for Computer Aided Diagnosis of Brain Tumour (CADbrat) Using Tensor Flow, TFlearn Library and Magnetic Resonance Images

Raymond Confidence1, Dada O. Michael1, Musa Emmanuel1, Oni O. Michael1, Awojoyogbe O. Bamidele1,2

1. Department of Physics, Federal University of Technology, PMB 65, Minna, Niger State, Nigeria.

© AJMP is the official journal of the Federation of African Medical Physics Organizations (FAMPO). This is registered under Nigerian company number (CAC/IT/No 54182).
See http://fampo-africa.org/
ISSN 2043-5977

Received: 2021.01.22; Accepted: 2021.04.04; Published: 2021.05.15

Abstract

Early diagnosis of a brain tumour is very important. It improves patient’s quality of life and offers diverse options for treatment and reduces the financial burden of tumour treatment. In recent years, computational resources have improved, and they are capable of handling large amount of data which are generated during clinical diagnosis. Furthermore, the recent advances in machine learning have made large data analysis in medicine more tractable and interesting. Automated deep learning-based computer-aided diagnosis (CADx) can be deployed as a crucial and reliable tool used for early detection. In this study, a CADx algorithm with high accuracy and sensitivity has been developed to provide assistance to health care givers in the timely diagnosis of brain tumours. The artificial intelligence algorithm made use of Tensor Flow, TFlearn Library and Magnetic Resonance Images as such model optimization was achieved by cascading the convolutional layers with a down-sampling layer (Max pooling) and higher filter number at the inception convolution layer to extract high level features and provide translational invariance down the network. With this model, an average F1-score of 99.49% was obtained despite training the network on a CPU. This amounts to a slightly improved performance of a similar model which was, however, trained on NVIDIA Tesla K40 server (2880 CUDA cores and 12GB memory). To demonstrate the application of the developed convolutional neural network (CNN) algorithms in clinical scenarios, a graphic user interface (GUI) using Flask application programming interface (API) was developed for easy brain tumour diagnosis with MRI scans.

Keywords: Computer aided Diagnosis of brain tumour (CADbrat), Tensor Flow, TFlearn Library, CNN networks, Deep learning, Artificial Intelligence and Magnetic Resonance Images

Introduction

Brain tumour or spinal cord tumour are heterogenous masses of abnormal tissues with several morphology and behavioural signatures growing from the brain and its surrounding structures [1-2]. The American cancer society’s forecast that an approximate of 23,890 malignant tumours of the brain or spinal cord (13,590 in males and 10,300 in females) will be detected in 2020, with 18,020 expected deaths [3, 4], thereby becoming a major cause of mortality among children and adults in the world [5]. In addition, a depressing six-month to one-year survival rates, despite treatment efforts, have been recorded [6]. Radiologists spot the tumour by analysing medical brain images, the nature of the tumour is measured based on the location of tumours in the brain and their estimated coverage area [8, 12]. Nevertheless, the detection of a brain tumour is a tedious, time consuming and reoccurring responsibility, involving the need to accurately interpret anomalous brain images.
Brain tumour detection, most times, requires at least two experienced specialists to analyse and attest the report on imaging enquiries. Additionally, the shift from two-dimensional (2D) to 3D imaging has rapidly increased the number of images the radiologist has to investigate coupled with increased signal-to-noise ratio [14]. Invasive biopsy and surgery are explored when any complication is encountered [10]. Consequently, such a responsibility is very expensive [15] and is prone to misdiagnosis and errors. Magnetic resonance imaging (MRI) is the most common and effective brain imaging modality for appropriate diagnosis of brain tumour [10] due to its high sensitivity, and contrast resolution [5, 16]. MRI also provides anatomical details with high visual contrast while providing functional as well as insights into metabolic path ways with the use of a strong magnetic field, computer and radio frequency excitation pulses to produce a detailed scan of all internal body structures [17].

MRI imaging modalities are used to probe the tumour location and extent, but typically, it is hard to determine the type of tumour with visual observation alone [18]. To overcome most of the challenges in brain tumour diagnosis such as time consumption, misdiagnosis, and the need for an experience radiologist, qualitative and objective computer-aided detection and diagnosis CAD programs have been developed. The heavy workloads in the neuroradiological field have led to the urgent need for new advances of CAD to detect different classes of brain tumours [14].

CAD systems include manifold components such as the application of intelligent retrieval systems including AI, machine vision, and medical image refinement for the sole purpose of diagnosis [19]. An accurate CAD can serve as an alternative option to the radiologists as well as aid physicians and radiologists in the classification of an MRI brain image [20]. A CAD module might demand segmentation, detection and classification all put together in a single systematic set-up to spontaneously help medical specialists achieve precise diagnosis [21]. CAD systems estimate tumour types and grades by means of probability models derived from Artificial intelligence (AI) algorithms.

Various studies on computer-aided diagnostics [9, 18, 25-29] have shown improved performance through the recent evolution of artificial intelligence algorithms, precisely deep learning concepts. The growth of low-grade gliomas has been successfully detected in a CAD system that permits substantial increase in tumour volumes form reference standard before detection [7]. This design could become the clinical gold stand for the future. An automated CAD system that employed K-means clustering and support vector machine (SVM) was proposed to suppress false positive and false negative values in delineating and classification of brain tumours respectively, the brain tumour classification was based on radiomic features (Gray level co-occurrence Matrices (GLCM)) extracted from the MRI scans [10]. Similarly, an automatic CAD scheme with a combination of automatic segmentation, via a versatile three-dimensional CNN and radiomics feature extraction using SVM, has been designed to enhance glioma detection capabilities. This method is unique because wide ranges of shape, first-order and texture feature were extracted from an MRI data [11]. An artificial intelligence framework with a fused prowess of a neural network and fuzzy logic principles called Adaptive Neuro Fuzzy Inference System (ANFIS) was deployed to unsupervisedly classify GLCM and Grid feature extracted from morphologically segmented MRI scans with images from expert radiologists [12]. The malignancy of diffused gliomas by extracting textural, moment distribution, global grey scale distribution features from MRI images was investigated. The global, local and the jointure of this feature furnished a CAD system with a promising diagnostic accuracy for clinical setting [22]. Several set-ups for classification and feature extraction were experimented in a quest to design an efficient CAD to differentiate between benign and malignant tumours using MRI images. The best performing combination of extractors and classifiers were GLCM characteristics and pattern net respectively [23]. GLCM has become a household feature extractor due its robustness. This matrix method was also combined with a Recurrent Neural Network (RNN) in a CAD system designed for pituitary, glioma and meningioma tumour classification. Also, pituitary, glioma and meningioma tumours have been classified using a novel NN algorithm in a CAD system where two-dimensional Discrete Wavelet Transform (DWT) and Gabor filter algorithms were employed in the conversion of MRI region of interest (ROI) segments to enable wholistic extraction of spatial information [30]. Amongst many more, a five-layer SVG CNN have been designed to classify wavelet information extracted using DWT from T1-weighted MRI images with the whole process fused into a single frame CAD. A summary of the reviewed CAD systems is provided in Table 1. Although many computer-aided detections (CAD) methods have been explored in brain tumour investigation in MR images [10, 23], the research is void of any MRI CAD for classification of: Astrocytomas, Oligodendroglomas, and Glioblastoma multiforme (GBM) tumours from a pool of abnormal MRI brain scans.

The aim of this study is to i) show that a cascaded convolution-pooling CNN system demonstrates better
performance, by optimizing an existing CNN algorithm for brain tumour classification of MRI images and ii) develop an efficient, highly accurate CAD tool for the classification of Astrocytomas, Oligodendrogliomas, Glioblastoma multiforme (GBM), and unhealthy tumours. The completely automated design with high efficiency will classify brain MRI images into these categories without biopsy and surgery [24], thus, helping doctors in brain tumour diagnosis.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fathallah-Shaykh et al. [7]</td>
<td>Segmentation, Volume Estimation and Online abrupt Change point analysis</td>
<td>Change in tumour volume median of 57% in comparison to a median of 174% change of volume required to diagnose tumour growth (p &lt; 0.001).</td>
</tr>
<tr>
<td>Samanta &amp; Khan [10]</td>
<td>SVM</td>
<td>Accuracy of 99.28%</td>
</tr>
<tr>
<td>Chen et al. [11]</td>
<td>SVM, 3D-CNN</td>
<td>Accuracy of 91.27%, Precision of 91.27%, Sensitivity of 91.27% and F-1 score of 90.64%.</td>
</tr>
<tr>
<td>Kathirvel &amp; Batri [12]</td>
<td>ANFIS, GLCM</td>
<td>Sensitivity of 98.86%, Specificity of 98.95%, and Accuracy of 99.01%.</td>
</tr>
<tr>
<td>Ghahfarrokh &amp; Khodadadi [22]</td>
<td>DWT, GLCM FD, LE, ApEn, SVM, KNN</td>
<td>Accuracy of 98.9%</td>
</tr>
<tr>
<td>Roy et al. [23]</td>
<td></td>
<td>Accuracy of 88%</td>
</tr>
<tr>
<td>Ismael [30]</td>
<td>GLCM, RNN</td>
<td>Accuracy of 98%</td>
</tr>
<tr>
<td>Kathirvel [31]</td>
<td>DWT Gabor filter</td>
<td>Accuracy of 98.8%</td>
</tr>
<tr>
<td>Clark et al. [32]</td>
<td>DWT, SVG</td>
<td>Accuracy of 99.93%</td>
</tr>
</tbody>
</table>

Cross-Validation (CV)

Cross validation is mainly deployed for hyper parameters tuning and to better evaluate the performance of a model [38]. A CV technique is basically used in evaluating the capability of a classifier for unseen data. A defined number of data K is established (K-folds); the total tagged data set is stochastically grouped into K groups of the same size, for all groups [39]. For each group, the CNN is trained on the leftover K-1 groups and is validated on data from that group. The final accuracy is the mean of all K accuracies. In this study, k = 6, meaning 16% of the whole dataset is used as test set while 84% is used as a training set for each training split. The resulting model is validated on the remaining part of the data [40].

Convolution Neural Network (CNN)

Convolution neural networks as the name implies consists of artificial “neurons” with quantitative weights which transfer information amongst each other, such that when this trained network is given an image or data, it will perform with acceptable accuracy. Neural network has several layers of feature-detecting “neurons”. The layers are designed to extract unique features from an input, while subsequent layers detect features from features extracted from the previous layers [41]. Usually, CNN uses 5 to 25 different layers of features extraction.

In CNNs, a crucial factor needed in translating an input, which influences the output, is the weight. Weights (w) are numerical arguments that determine the impact of neurons on each other. For example, if a neuron has the inputs x1, x2, and x3, then the synaptic weights to employ on the inputs are w1, w2, and w3. The output is given as:

Materials and Methods

Patient Information

DICOM format dataset for 100 patients was collected from Repository of Molecular Brain Neoplasia Data (REMBRANDT) of the Cancer imaging archive (TCIA) [33], a source that houses MRI scans of The Cancer Genome Atlas (TCGA) patients for image processing and analysis. This dataset contains 65,427 MRI images, which include 21,307 Astrocytoma, 17,983 Glioblastoma, 12,460 Oligodendroglioma and 13,677 unidentified tumours, sorted based on patient ID. 556 normal (healthy) T1, T2 and/or proton density (PD) MRI images without tumour were collected from the Brain Images of Normal Subjects (BRAINS) Image Bank repository of University of Edinburgh [34] and 30,688 longitudinal volumetric T1 MRI images of The Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD). The control group of this dataset was the target since it also contained an Alzheimer’s Disease (AD) group [35]. Clinical metadata was used in manual classification of the images from REMBRANDT and their formats were translated to JPEG, likewise those of the MIRIAD dataset. These images were combined and arranged based on their labels and serialized using the Python pickle module into a single file for easier processing [36].
where \( i \) ranges from 1 to the total number of inputs available. It is important to note that (1) is a matrix which leads to the weighted sum. Bias \((b)\) behaves like the intercept in a linear regression. It serves as a constituent parameter needed to manipulate the output according to the weighted sum of the inputs to the neuron. This process is demonstrated by (2) and (3):

\[
Output = \text{sum(weight * inputs)} + \text{bias}
\]

\[
y = \sum x_i w_i + b
\]

A function known as activation function can be applied on the output of (3) such that the input of the next layer becomes the output of the neurons in the initial layer (as shown in Figure 1). The general structure of CNN consists of switching convolutional layers with down sampling layers such as pooling layers, accompanied by fully connected layers.

**Convolutional Layer**

The convolutional layer is a core component of a CNN. This layer deduces multiple channels “feature response maps” in the form of activation values from a defined spatial area [44]. Precisely, in this layer, several filters (kernels), which indicate the receptive field of a neuron, are used by multiplying the whole image and the subsequently acquired feature maps, creating several feature representations [45]. The arguments of the layers are the amount of feature representations, dimension of the representations and filter magnitudes. Each layer \((L)\) has feature representation maps \((M)\), Each map in Layer Ln is linked to each other maps in layer Ln-1. Weights are jointly used by neurons of a particular map but they do possess varying input fields [46].

**Sub-sampling Layers**

The pooling, drop-out and flatten layers in the convolutional neural networks are known as the sub-sampling layer because they dramatically down sample the size of the input data and resolutions of the features [48].

The pooling layers are broadly examined among the three layers. In this study, we focused on max pooling layer which is widely used and applied in our work. The layer segments the image to rectangular subsections (pixels) and it only maps out the highest valued sub-section. This layer forcefully down samples the spatial attributes of the input. It has the advantage of decreasing the number of training parameters or weights by about 75%, thus, reducing the computation burden [48]. Also, it minimizes over fitting of the model [49]. This layer takes for an instance a kernel (usually of 2x2 magnitude) and a stride of the equal distance (as shown in Figure 2). It then utilizes on the input data and output the max number in each sub-section that the kernel operates on.

**Activation or Non-linearity Layer**

The ability of the CNN to fit a given function (predominantly non-convex) is largely dependent on the non-linear activation functions. Activation functions take a vector and execute a given constant point-wise procedure on it. An example of an activation function employed in this study is Rectified Linear Unit (RELU):
Network was proposed by [36], we - {5×5), (3×3), which is ba - function

\[ y = \max(0, x) \] (4)

where \( x \) is the input to the neural network, and the threshold for the activation is simply at zero. For a CNN having \( l+1 \) layer, the Softmax layer is employed to transform the real valued features produced by the last Neural Network layer into many chances. Therefore, the output has a total value of 1, according to

\[
\sigma(x)_j = \frac{e^{x_j}}{\sum_{i=1}^{K} e^{x_i}}; \quad j = 1, 2, \ldots, K
\] (5)

where \( x \) is an absolute vector with true values \( x_j, j = 1, \ldots, n \), produced at the last layer of the NN and \( n \) is the magnitude of the vector [52].

**Fully Connected Layers (FC)**

These layers are usually the last layers of a CNN and have the largest number of parameters in the network. FC layers are employed to make predictions. They are most times implemented with SoftMax layer, thereby preceding the outcome of the Neural Network. A one-dimensional feature vector is generated from a given 2-dimensional input for next feature map.

**Optimizer**

This parameter often determines the learning and convergence of the developed model. Each optimizer has its peculiar modifiable parameters, but the learning rate is always inclusive. This parameter defines the extent to which the weights have been updated after each epoch. When Learning Rate is high, the change in the weight will be higher than for a small learning rate, after each epoch. These stochastic gradient descent-based optimizers are deployed with the aim of reducing an objective loss, which is basically the computed deviation between forecasted and reference data values. The optimization process encompasses the derivation of the specific parameters of the network which produces the most excellent results in the defined tasks such as classification [53].

The ADAM-Optimizer is one of the most common adaptive step size methods [54]. Adaptive Moment (Adam) computes the individual adaptive learning rate for each parameter by calculating the first two moments of the gradients [55]. Loss functions measure the difference between expected chances and approximates thereof [56]. Given an input and a reference standard, they estimate the loss, that is, the discrepancy amongst output and reference standard, represented mathematically as \((y_{predicted} - y)\). It is a process of estimating the performance of the models on the given dataset.

The generalized Categorical cross entropy is the most used loss function for multi-class classification which is applicable in single label characterization. This happens when only one class is relevant for all data point [57]. Meaning, an example can belong to one class only. The loss for input vector \( y \) and the corresponding one-hot encoded target vector \( \hat{y} \) is:

\[
L(y, \hat{y}) = - \sum_{j=0}^{N} \sum_{i=0}^{N} (y_{ij} \log[\hat{y}_{ij}])
\] (6)

where \( \hat{y} \) is a one-hot encoded target vector \((y_{11}, y_{12}, \ldots, y_{10})\).

\[
y_{ij} = \begin{cases} 1 & \text{if ith element is in class j} \\ 0 & \text{otherwise} \end{cases}
\] (7)

\[
P_{ij} = f(y_i) = \text{Probability that ith element is in class j}
\]

Categorical cross entropy compares the probability dispersion of the estimated outcome with the reference probability dispersion such that the possibility of the reference class is fixed at one and zero for the other classes [57].

**Network Architecture**

The new network (which will be called “AWSNet”) is aimed at an automatic classification of the following brain tumours: Astrocytoma, Glioblastoma, Oligodendrogliaoma, healthy, and unidentified tumours. The earlier network was proposed by [36], we optimized the network by sandwiching Max pooling layer between each convolutional layer to improve the generalizability and improve the fitting function.

Small focal features were delineated from the MRI images by feeding the network with images of size 64x64 for faster computation.

a) The first group of the network consists of four convolutional layers with 32, 64, 128, and 128 number of filters and corresponding filter sizes of (5×5), (3×3), (3×3), and (2×2) respectively to extract numerical feature, a stride of 1, and Relu activation function to fire the neurons of each layer (as shown in Figure 3). The decrease in the filter size was to enable more local features to be extracted down the network.

b) Each of the convolutional layers is followed by a sub sampling layer that performs maximum pooling.
with a stride of 2 to provide dimensional invariance of the data transferred across each convolutional layer.

c) The second group of the network consists of 2 FC layers and a Dropout layer between them. The dropout layer is positioned between these layers to downsize the number of features extracted by 50%, hence, providing stability to the network and avoiding overfitting.

The extracted feature is finally fed into the last FC layer with 5 nodes corresponding to the 5 classes of images, and a Softmax activation function attached to it to predict outcomes.

**Network Implementation**

The selected frameworks for implementing this AWSNet are TFlearn [53], scikit-learn [54] and other Python libraries. TFLearn was employed in the coding of each layer of our model.

Training and testing of this model were implemented via 6-Fold Cross validation which separated the whole data into 6 groups such that each group of data was trained for 10 epochs totalling 60 epochs for the entire dataset. The number of epochs for training was chosen to overcome the overfitting challenge and reduce the time it takes to train the network within each epoch. ADAM optimizer was chosen because of its speed of convergence, simplifying learning rate [55], and its ability to discover a better minimum loss function [53]. A learning rate of 0.001 was applied to minimize training time and loss. The models were trained and tested on a computing device possessing an Intel Corei5 processor equipped with NVIDIA GEFORCE GT 650 with 1GB memory that operates at 2.6 Giga-Hertz with 6 GB of RAM and Windows 64-bit operating system.

**Network Training**

Training a machine learning algorithm is tantamount to minimizing two maps $y(x)$ and $\hat{y}(x)$, where the system is committed in discovering the minimal difference in distance between $y(x)$ and $\hat{y}(x)$ in each system of measurement. The fundamental basic of the training is given as follows:

$$\hat{y} = w^T x$$  \hspace{1cm} (8)

Here, $w$ is the weights to be optimized. They compute the mathematical mapping between the features $x$ and the output $\hat{y}$; determining the minimal distance between $y$ and $\hat{y}$ is referred to as mapping $y$ from $x$.

**Performance Evaluation Metrics**

Performance evaluation methods used to evaluate the new model would be presented. The performance of the new method would be measured on the validation data in terms of specificity, sensitivity, classification accuracy, confusion matrix, F1 score. The basic terminology used in performance metrics are:

i. **True Positives (TP):** This involves a circumstance where the predicted class by the model was YES and it really was YES

ii. **True Negatives (TN):** This involves a circumstance where the predicted class by the model was NO and in reality, it was NO

iii. **False Positives (FP):** This involves a circumstance where the predicted class by the model was YES and actually, it was NO

iv. **False Negatives (FN):** This is an event where the predicted class by the model was NO and actually, it was YES

**Sensitivity or Recall**

This gives the ratio of cases accurately named positive from the entire positives.

$$\frac{TP}{TP + FN}$$  \hspace{1cm} (9)

The recall is a measure showing the proportion of patients that had a brain tumour as diagnosed by the algorithm as having a brain tumour.

**Specificity**

This gives the proportion of cases accurately named positive from all false negatives

$$\frac{TN}{TN + FN}$$  \hspace{1cm} (10)

Specificity estimates the ratio of persons that had a brain tumour predicted by the model as no-tumour.

**Precision**

Precision defines the ratio of cases accurately named as positives out of all predicted as positives.

$$\frac{TP}{TP + FP}$$  \hspace{1cm} (11)

Precision in a clinical context of brain tumour gives the fraction of people that were predicted to have a brain
Confusion Matrix or Table of Confusion
A table of confusion is a primary metric employed in determining the precision of a model; it gives us a matrix as output. The confusion matrix is not used as an evaluation metric oftentimes, but others or virtually all the evaluation tools are founded in this matrix and the magnitude of its quadrants. In the equations, positive and negative denote the two classes considered during binary classification.

Classification Accuracy
Classification Accuracy, oftentimes called accuracy, is the most common evaluation metric for classification problems. It gives the proportion of a few precise predictions to the sum of all cases, multiplied by 100 to turn it into a percentage. This metric performs better for dataset with the same amount of data for all cases.

\[
\text{Classification Accuracy} = \frac{TN + TP}{TN + FN + FP + TP}
\]  \hspace{1cm} (12)

F1 Score
F1 score gives outputs between the range of zero and one, defined as the weighted mean Precision and Recall. F1 is preferably employed than accuracy, precisely in situations where the dataset does not have the same amount of data for each class.

\[
F1 \text{ score} = \frac{2 \times (\text{precision} \times \text{sensitivity})}{\text{precision} + \text{sensitivity}}
\]  \hspace{1cm} (13)

![Figure 3. The proposed CNN architecture.](https://globalmedicalphysics.org/)
Results
The new model was used to perform brain tumour classification for Astrocytoma, Glioblastoma Multiforme, Oligodendroglioma, healthy tissue and unidentified tumour images. The F1 score for the new model trained with the dataset of the five classes are above 99% with mean F1-score of 99.48% (as shown in Table 2).

For healthy brain images, a high F1 score was obtained, and the table of confusion did not record wrong classifications results. This is probably caused by the lack of sufficient information about tumours in healthy brain MRI scans. The highest recorded accuracy was 99.95% received from Oligodendroglioma for the test dataset while a maximum accuracy of 99.92% was obtained from Astrocytoma for the whole dataset.

Accuracy for the test and entire dataset are shown in Tables 2 and 3 respectively. The model has shown that it is capable of performing classifications with accuracy of above 99% for Glioblastoma Multiforme, Oligodendroglioma and healthy tissue. However, classification performance for Astrocytoma was lower than 97%.

<table>
<thead>
<tr>
<th>Tumour Type</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrocytoma</td>
<td>99.43%</td>
<td>97.78%</td>
</tr>
<tr>
<td>Glioblastoma Multiforme</td>
<td>99.32%</td>
<td>99.86%</td>
</tr>
<tr>
<td>Oligodendroglioma</td>
<td>99.17%</td>
<td>99.95%</td>
</tr>
<tr>
<td>Healthy tissue</td>
<td>99.99%</td>
<td>99.92%</td>
</tr>
<tr>
<td>Unidentified tumour</td>
<td>99.49%</td>
<td>99.93%</td>
</tr>
<tr>
<td>Average F1-score</td>
<td>99.48%</td>
<td>99.49%</td>
</tr>
</tbody>
</table>

Table 2. F1-score values obtained from AWSNet model and accuracies of the test set

<table>
<thead>
<tr>
<th>Tumour Type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrocytoma</td>
<td>99.92%</td>
</tr>
<tr>
<td>Glioblastoma Multiforme</td>
<td>99.51%</td>
</tr>
<tr>
<td>Oligodendroglioma</td>
<td>99.49%</td>
</tr>
<tr>
<td>Healthy tissue</td>
<td>99.49%</td>
</tr>
<tr>
<td>Unidentified tumour</td>
<td>99.49%</td>
</tr>
<tr>
<td>Average Accuracy</td>
<td>99.58%</td>
</tr>
</tbody>
</table>

Table 3. Accuracies for the entire dataset (a combination of the test and validation set)

CAD graphic user interface (GUI)
A simple graphic user interface (GUI) CAD system for brain tumour classification has been developed. The tool is capable of processing different MRI image formats (jpg, png, bmp and tiff) through the front end of the GUI as shown in Figure 4A. The CAD GUI provides a file-selection button to load the MRI image (illustrated in Figure 4B). A classify button classifies the input image as belonging to Astrocytoma, Glioblastoma, Oligodendroglioma, healthy tissues or unidentified tumours respectively. The web-app loaded image processing takes between 45 seconds to a few minutes (as shown in Figure 4C), and finally displays final prediction (as shown in Figure 4D). The CAD-GUI system is hosted locally through the use of Python FLASK API front-end, which is written using HTML, Mysql, Java, jQuery and CSS 5. The app was served locally using WSGIserver.

A confusion matrix that demonstrates the performance of our model on the whole datasets with recorded labels is shown in Table 4. Misclassifications were observed for several classes. Oligodendroglioma tumour recorded the highest number of misclassifications of 143 and healthy brain tumour class had the least number. Oligodendroglioma tumour was often misclassified as Astrocytoma and Glioblastoma Multiforme. Oligodendroglioma tumour class had the least MRI scans compared to other grades with named tumours and these scans most probably possess tumour at locations similar to those of other classes.

Table 4. F1-score values obtained from AWSNet model and accuracies of the test set

A confusion matrix that demonstrates the performance of our model on the whole datasets with recorded labels is shown in Table 4. Misclassifications were observed for several classes. Oligodendroglioma tumour recorded the highest number of misclassifications of 143 and healthy brain tumour class had the least number. Oligodendroglioma tumour was often misclassified as Astrocytoma and Glioblastoma Multiforme. Oligodendroglioma tumour class had the least MRI scans compared to other grades with named tumours and these scans most probably possess tumour at locations similar to those of other classes.

### Table 4. F1-score values obtained from AWSNet model and accuracies of the test set

<table>
<thead>
<tr>
<th>Tumour Type</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Astrocytoma</td>
<td>21173</td>
<td>77</td>
</tr>
<tr>
<td>B Glioblastoma Multiforme</td>
<td>33</td>
<td>17907</td>
</tr>
<tr>
<td>C Oligodendroglioma</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td>D Healthy tissue</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E Unidentified tumour</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

A = Astrocytoma, B = Glioblastoma Multiforme, C = Oligodendroglioma, D = Healthy tissue and E = Unidentified tumour.

A confusion matrix that demonstrates the performance of our model on the whole datasets with recorded labels is shown in Table 4. Misclassifications were observed for several classes. Oligodendroglioma tumour recorded the highest number of misclassifications of 143 and healthy brain tumour class had the least number. Oligodendroglioma tumour was often misclassified as Astrocytoma and Glioblastoma Multiforme. Oligodendroglioma tumour class had the least MRI scans compared to other grades with named tumours and these scans most probably possess tumour at locations similar to those of other classes.

### Table 4. F1-score values obtained from AWSNet model and accuracies of the test set

<table>
<thead>
<tr>
<th>Tumour Type</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Astrocytoma</td>
<td>21173</td>
<td>77</td>
</tr>
<tr>
<td>B Glioblastoma Multiforme</td>
<td>33</td>
<td>17907</td>
</tr>
<tr>
<td>C Oligodendroglioma</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td>D Healthy tissue</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E Unidentified tumour</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

A = Astrocytoma, B = Glioblastoma Multiforme, C = Oligodendroglioma, D = Healthy tissue and E = Unidentified tumour.

A confusion matrix that demonstrates the performance of our model on the whole datasets with recorded labels is shown in Table 4. Misclassifications were observed for several classes. Oligodendroglioma tumour recorded the highest number of misclassifications of 143 and healthy brain tumour class had the least number. Oligodendroglioma tumour was often misclassified as Astrocytoma and Glioblastoma Multiforme. Oligodendroglioma tumour class had the least MRI scans compared to other grades with named tumours and these scans most probably possess tumour at locations similar to those of other classes.

A confusion matrix that demonstrates the performance of our model on the whole datasets with recorded labels is shown in Table 4. Misclassifications were observed for several classes. Oligodendroglioma tumour recorded the highest number of misclassifications of 143 and healthy brain tumour class had the least number. Oligodendroglioma tumour was often misclassified as Astrocytoma and Glioblastoma Multiforme. Oligodendroglioma tumour class had the least MRI scans compared to other grades with named tumours and these scans most probably possess tumour at locations similar to those of other classes.
Discussion

The design and implementation of the computer-aided diagnosis (CAD) system made use of Softmax classifier to classify brain tumours such that features were automatically extracted from the images by convolutional layers, while feature selection was performed using down-sampling layers. The methods in this study employed higher translational invariance [58] by introducing a large number of non-linearity to the network making it train faster and robust to noise and distortion with the help of max pooling layers. This has been demonstrated by the observed less misclassification and the improved Average F1 score in comparison to the model performance by [36].

This study has demonstrated that a CAD system can significantly enhance the performance, and processing time of brain tumour diagnosis. It offers a means of brain tumour delineation with minimum or no biopsy.

The advantage of the CAD was evident in the high F1 score and accuracy observed. However, it is worthy of note that, because we could not obtain the same number of images for each tumour class, the measurement of sensitivity and precision in the classification of brain tumour was not considered in this study. Consequently, we had to focus on measuring the general stability of the model by using an F1-score metric.

The CAD-GUI worked at optimal level since it was specifically designed to classify just the classes of tumour selected for this study and it returned the diagnosis result in less than a minute.

Conclusions

We have developed a neural network algorithm for the classification of brain tumours from MRI images. The performance of the model was tested and compared with similar methods that are available in the literature, and was found to have better performance. Classification accuracy of 99.95% and F1-score of 99.48% were obtained from the model. Classification of brain tumours through the new model was found to be faster with improved accuracy when compared to the manual classification carried out by experts in the clinics. The use of T1-weighted and T2-weighted MR images lead to high stability, accuracy and output because of their better soft tissue contrast. The CAD-GUI developed from the AWSNet2 algorithm (with codes based on Python, Java and MySQL) has an interactive interface and can be invoked via HTML enabled-web link, thus, making it quite accessible to users with limited computer experience.

Abbreviations

CAD: Computer-aided diagnosis; GLCM: Gray level co-occurrence Matrices; CNN: Convolutional neural network; ANFIS: Adaptive Neuro Fuzzy Inference System; DWT: Discrete Wavelet Transform; GBM: Glioblastoma multiforme; MRI: Magneti resonance imaging; PD: Proton density.

Acknowledgement

The authors would like to express their gratitude to Dr. Udunna Anazodo of the Department of Medical Biophysics, Western University and Lawson Health Research Institute, London, Ontario, Canada for providing encouragement throughout the course of this study.
Author Contributions
All authors contributed equally to this study and gave their final approval.

Competing Interests
The authors have declared that no competing interest exists.

References


