Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLASSIFICATION OF BRAIN TUMORS: USING DEEP TRANSFER LEARNING

NOR AZURA HUSIN[1], MOHAMED HUSAM[2], MASNIDA HUSSIN[2]

[1]Intelligent Computing Research Group, Faculty of Computer Science and Technology, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia

[2] Faculty of Computer Science and Technology, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia E-mail: [1]n_azura@upm.edu.my, [2]gs60282@student.upm.edu.my, [2]masnida@upm.edu.my

ABSTRACT

Brain tumor classification is important for diagnosing and treating cancers. Deep Learning has improved medical imaging with Artificial Intelligence (AI). Brain tumor's shape, size, and intensity make subclassification difficult. Medical imaging data is scarce. Any medical data involves privacy of the patients, hence unlike other image data, medical image data is not easily available. There are only few medical image data that is freely available for researchers. This project aims to develop a deep transfer learning model that can accurately classify brain cancers utilizing limited Medical Resonance Images (MRI) images. To achieve the goal, a modified GoogleNet model was used. Various learning algorithms were tested. The experiment also examined transfer learning and data augmentation. Finally, F1-average and confusion matrix were used to evaluate the model. Our model outperformed the state-of-the-art model in various research articles, according to performance matrices. Experimenters employed data augmentation and learning algorithms.

Keywords: *Deep Learning, Transfer Learning, Brain Tumors, Learning algorithms, Medical Imaging.*

1. INTRODUCTION

Deep learning has had a significant impact on the field of medical imaging. Medical imaging has changed disease diagnosis in numerous ways, including the ability of medical practitioners to accurately diagnose disorders (1). A diagnostic tool in medicine that does not necessitate any intrusive procedures is imaging. This method has become the primary source of the procedure because other diagnostic methods necessitate an even more invasive method of collection. Since medical image analysis is used by doctors and surgeons to plan more accurate and robust treatment for their patients, this is a common practice.

The identification and categorization of tumors is one of the most prevalent forms of diagnoses that may be obtained via the use of medical imaging. Depending on the circumstances, these tumors may be benign or malignant. Patients who do not receive effective treatment for malignant tumors face a high risk of death if the disease progresses. Tumors of the brain are among the deadliest and most difficult-to-treat cancers (2). Because of the brain's complexity and unique morphology, it can be extremely difficult to locate a tumor in the brain. Those that are malignant (cancerous) are referred to as such, whereas those that are benign (not cancerous) are referred to as benign. Additionally, there are numerous subtypes of malignant tumors, each with their own grade or stage of development. There have been numerous attempts by medical researchers to classify brain tumors; nevertheless, the WHO-approved classification standards are largely recognized throughout the medical industry (3).

The treatment for various forms of brain tumors varies, so it is critical to classify them. Glioma, meningioma, and pituitary tumors are the three most common forms of brain tumors. Because the physical characteristics of most tumors are indistinguishable, even radiologists(3) have difficulty classifying brain cancers.

Radiologists must analyze many MRI in order to classify a brain tumor. Radiologists must be extremely careful during the categorization procedure because a misclassification can lead to a disastrous outcome. However, as the number of imaging modalities grows, many images are generated in order to diagnose a single patient. Radiologists are also exhausted by having to look through all these images. Misclassification or human error may occur as a result of this. According to researchers in their paper (4), human error in medical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

image interpretation is responsible for 10-30% of all misdiagnoses.

To help radiologists and physicians in image classification Computer Aided Detection (CADs) systems are being used. CADs used machine learning and deep learning for the classification of medical images. The purpose of computer-aided design (CAD) systems is to improve radiologists' accuracy while simultaneously reducing the amount of time needed to examine images (5).

As opposed to Machine Learning, where features must be manually obtained, Deep Learning does it automatically (4). In addition, deep learning approaches have increased performance in computer-aided medical diagnoses in recent years (4)(6)(7). However, DL is not used in clinical diagnosis by radiologists and is limited to use for segmentation, rarely.

Though recent researched have shown a higher accuracy of classification of tumors using DL models, when unseen data is shown to the model, its accuracy significantly decreased. This may be a reason for the limited use of DL by radiologists.

DL need large amount of data, as it is a data driven model. Due to the lack of medical image data, to formulate a more efficient model is a challenge. Using transfer learning, the limitation of data is being addressed by researchers. Transfer learning (TL) allows the use of a CNN or DL model that has been trained using the ImageNet database for medical image classification. This strategy not only addresses the issue of limited data, but also accelerates the model training process.

Because protecting patients' privacy is crucial to collecting any medical data, medical image data, in contrast to other types of image data, is not readily available. Researchers only have access to a small fraction of the medical image data that is readily available.

MRI of Brain were utilized in the experiment detailed in the researchers' paper, where five distinct deep learning models, including AlexNet, GoogleNet, ResNet50, ResNet101, and SqueezeNet, were given these images. The researchers then implemented TL methods on the dataset. Even though the dataset utilized in this experiment was extremely small (8), all the models that were pre-trained using ImageNet achieved an accuracy of greater than 90 percent, with AlexNet attaining 99.04 percent. Additionally, for the classification of Alzheimer's researchers used deep transfer learning, with an accuracy of 99.22% (9). Extensive research is being undertaken on the application of transfer learning to the classification of medical images.

However, by employing TL to address the data limitation issue, the model becomes more biased. Therefore, the model is ineffective when presented with new dataset. Data augmentation is a method adopted by researchers to address the issue of bias. Using data augmentation, which does not necessitate the acquisition of additional data, it is possible to dramatically improve the diversity of the data used to train models. Data enhancement techniques, including as clipping, padding, and horizontal flipping, are commonly used in the training of huge neural networks (10).

Building a deep learning model that is more effective in the actual world has been hindered by the insufficiency of available medical image data. Existing state-of-the-art DL models have not been employed in clinical diagnosis by radiologists because they add large errors when introduced to unseen data (11). This is a result of overfitting. In addition, most models are trained so narrowly that they can successfully classify just tumors as malignant or benign. As a result of this binary categorization, even if the model performs well, radiologists must manually identify and classify brain tumors into their respective kinds. And the latter step is more important since it paves the path for doctors to design a medical treatment plan for patients. Insufficient training data for deep learning models is the root cause of overfitting. To solve data availability difficulties, researchers have employed a variety of strategies in TL, including the fine-tuning of a model that was pre-trained using generic images (8)(12).

Using Transfer Learning in DL models does not, however, address the issue of model overfitting. As these models are typically identified as being biased. Due to these constraints, radiologists are unable to fully integrate these bias models into clinical diagnosis. When a model is determined to be biased, generalization is hindered, and the model cannot be applied to a different issue. Even if the output task remains unchanged, these models will be unsuccessful if the input domain problem changes. Existing Deep Learning models are incapable of handling multiclass classification issues without degrading performance. Typically, models perform well for binary classification, however their performance drops for multiclassification situations.

With the increasing of image modalities, more images are being produced for a single patient diagnosis. Hence, radiologists spend more time on evaluating these images, resulting in fatigue and human error. That eventually may lead to a misdiagnosis. Since DL are becoming more efficient in classification of objects in images and have recently become more efficient than humans in classification and detection of objects in general images, DL models are being researched for medical image classification.

Furthermore, it is important to note that most of the research in this sector has been motivated by the need to increase the accuracy of existing models. Some research publications lack enough performance matrices, particularly those (9) that have showed more accuracy in their studies.

Consequently, the purpose of this study is to address the issue of the availability of medical image data via the techniques of data augmentation and transfer learning. Using a modified pre-trained GoogleNet model, research was undertaken to classify brain cancers into three categories: meningioma, glioma, and pituitary tumor.

The model images are contrast-enhanced T1-weighted MRI scans of a brain tumor. These images are loaded into a modified version of the GoogleNet. The findings demonstrate that the model's accuracy is great. The contributions of this research study can be described as follows: Using a deep transfer learning technique, a modified pretrained model with improved performance has been employed. The remaining sections are arranged as follows: In Section 2, we present a literature assessment of current brain tumor categorization techniques. Next, section 3 describes in full the technique employed for this research. In Section 4, experimental results are provided. Section 5 contains a discussion of the results, followed by a section 6 conclusion and future study.

2. LITREATURE REVIEW

2.1 Classification of Brain Tumors Using DL and ML

For the classification of brain tumors into three types, researchers in their research (12), used a modified GoogleNet. In the study they modified the pre-trained GoogleNet last layer and used it as a standalone classifier. Then the same model was again modified, and features extracted using TL technique, was used for the classification using SVM and KNN. This hybrid model performance was better compared to the stand-alone classifier. However, this model was only tested using the same dataset, and though the accuracy was higher, the performance matrices showed that model could be overfitting. Researchers (13) used GoogleNet with SVM trained on Medical Image Repository available from Harvard Medical School, to outperform the existing state-of art models. Furthermore, their model performed very well for the FIGSHARE dataset, that was not used in the training. Hence, the generalizability of this model can be seen. The accuracy of the model did drop from 100 percent to 95 percent when FIGSHARE dataset was used for testing. Since the model was trained with a very small dataset, and the accuracy score of 100 percent could have been because of the bias. This was evident when FIGSHARE dataset was used. To counter the imbalance of the dataset, no technique like data augmentation was employed. Though it also proved that using of DL with ML, where DL was used as a feature extractor and ML was used for classification, for classification can beneficial. Using of Pre-Trained Neural (PTN) network, like AlexNet, GoogleNet with SVM was used in the paper (14). For achieving higher accuracy with limited data, they have also used data augmentation and fine tuning of the deep learning models. However, as they used multiple deep learning models time complexity was high.

Another research paper (15) that used a majority voting-based algorithm for the classification of brain tumors, using ensemble learning. Five models of DL and ML were trained and using ensemble learning five DL and five ML model's performance were evaluated. Researchers used extensive data pre-processing; hence the training process will take a lot of time. Since usage of multiple DL models and ML models, the model will require a very high computational power, that would increase the cost. DL showed better performance compare to ML. Using of DL to extract features and ML for classification can be categorized as a hybrid model of DL. Though, it should be noted that these types of hybrid model associates with high computational complexity and time complexity.

2.2 Classification of Brain Tumors Using DL

To address the issue of manual feature engineering and manual analysis of MR images, researchers (16) in their study, used a two-stage training method to increase the accuracy and to automate the full process of diagnosing brain damage into five categories. First stage involved in training the last layer of the pretrained network while the first layers were frozen. Afterwards, in the second stage, all the layers were trained, using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

multiple learning rate. First layers were trained with lower learning rate while the last layers were trained with higher learning rate, to overcome the issue of vanishing gradient problem. To increase the dataset and to minimize the bias data augmentation was used. Researchers (16) used five common pretrained model; AlexNet, VGG16 and ResNet in their research to compare the performance. According to the results ResNet was more efficient. However, the validation loss graph indicates the model was overfit. The overfit occurrence hinders the model's usage in the real world.

Discrete Wavelet Transform (DWT) is a widely used feature extractor because of its ability to reduce the dimension by excluding the irrelevant features. DWT was used for feature extraction by researchers (17) for the classification of brain tumors. They used DWT and Genetic algorithm to select the most relevant features form the MR images, and the classification was done using Deep Neural Network (DNN). The research paper did not have performance matrices like confusion matric or validation loss graph to support their results. DWT technique was also used by the researchers (18) for feature extraction, segmentation was done using fuzzy c means. However, these approaches have not been used in more recent research papers. As the DL models can extract feature without any manual settings, the use of a different algorithm for feature extraction seems to be not useful.

While traditional machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbor (KNN) are used by some researchers for the classification process, the amount of time taken to manually feature engineering makes these models irrelevant in a clinical setting. Hence, more recent approach has been to use ML techniques with DL techniques. Although, currently DL models are being used for both feature extraction and classification due to the simplicity and minimization of time for all the pre-processing stages. A pre-trained GoogleNet was used for the classification of brain tumors in to normal and abnormal with a higher accuracy by the researchers (19). Using BRATS dataset, they used a Fully Convolutional Neural Network (FCNN) for the segmentation of the brain tumors. The process was done to find the Region of Interest (ROI), so that the accuracy of the model would be much higher. After the segmentation process the information was send to a GoogleNet for the classification. Accuracy of the model was shown to be 98 percent. Segmentation process is one of the techniques, in image classification. However, for the detection of brain tumor, or classification of brain tumor some of the researchers (20), in their research paper used a pre-trained AlexNet for the classification brain tumors. Researchers (20), argued that the segmentation process was more time consuming, and the current idea is to assist radiologists with a higher accuracy and faster model that would classify brain tumors. Their experiment showed that the pre-trained AlexNet took 2 minutes 37 seconds for the training of the model with an accuracy of 100 percent. However, the dataset used for training was very small. There is also a possibility that the model was biased or overfit as the data imbalance was shown in the Harvard medical image dataset.

2.3 Classification of Brain Tumors Using TL

Transfer Learning is most common approach, in deep learning, for any small dataset. Researchers (21), in their study, approached to classify glioblastoma into high grade, lower grade with a third class being normal, using a modified AlexNet. The las layer of AlexNet was replaced with two fully connected layers and a SoftMax layer for the classification. The model took 3 hours for training and accuracy was 91 percent. The experiment was done without using transfer learning. The accuracy and the training time indicate the importance of using transfer learning in the case of small dataset.

Transfer learning technique is one of the most common used technique in Deep Learning. It is used because it gives an advantage of training the model with less time and training a model with a small dataset. Researchers (22), in their study used ResNet-50, MobileNet V2 and Xception for classification of brain tumors. MobileNet V2 achieved an accuracy of 98%, highest compared to other models used in their research. The size of data set was increased using Data Augmentation technique like Rotation, height, width shift, brightness, etc. Same TL and Data Augmentation techniques can be found adopted in another research (23) by researchers. However, researchers (23) used four optimizers to compare their proposed model, ResNet50. This research was more focused on finding an optimal optimizer for the DL model. By fine tuning the optimizers they achieved an accuracy of 99% for multi-classification of three types of tumors. Although, they had a higher accuracy, with a small dataset from FIGSHARE, the class imbalance was observable as well as the absence of a validation a loss graph, means there is lack of performance matrices.

Many research papers employed a solution to discover an appropriate hyper parameter of the model to address the limitations of TL and Data Augmentation. To avoid over fitting a regularization and optimization technique was adopted in their research by researchers (24). Techniques that were used in their research are Data augmentation, drop out layers and an early stop method was utilized. Adam optimizer was used as an optimizer. A same approach was used by researchers (10) for binary classification of brain tumors. TL and Fine tuning for the multi-classification of brain tumors are shown in their studies (25), by researchers. They have used a pre trained VGG19 and fine-tuned the parameters. Instead of using layer-wise fine tuning, researchers used block wise fine tuning, by dividing VGG19 layers into six blocks. This increased the training time significantly.

Model overfitting issues are common in deep learning models. To overcome these researchers (26) used drop-out layers in their Deep Convolutional Generative Adversarial Network (DCGAN) model. In the research Deep convolutional network was trained as a discriminator in a GAN to distinguish between fake and real MR images of brain tumor. Drop-out rate was fixed and was used between the layers. Drop-out technique was also used by researchers (2,27,28) in their research to minimize the overfitting.

A novel architecture was used for multi-grade classification of brain tumors by researchers (29), in their study. In the three-stage classification process, 1) Segmentation was done to segment the brain tumors, 2) Data Augmentation was done on the segmented brain tumors to increase the number of data and 3) Deep feature extraction and classification. Extensive data augmentation was used in their research. However, still the class imbalance was observed. For multi-classification of brain tumors, a two-channel novel model was used in the research (2) by researchers. Researchers used a simple CNN architecture, to build their model for lower computational complexity. The training of the model is shown to take less time compared to the other well-known state-of-art models. However, since there was no comparison done with other pre-trained networks like GoogleNet, this claim cannot be proven. Cropped and uncropped with different size of brain MR images were compared for the efficient classification of brain tumors into three types by researchers (30) in their study. Researchers were using extensive data pre-processing techniques for achieving a better result. However, the time taken for data pre-processing will be high, and it would not be proved to be efficient in the real world. To overcome the issue of bias, data augmentation was used by researchers (27), in their paper. A pre-trained modified AlexNet was used to classify brain tumors, using FIGSHARE dataset. Model accuracy was 98 %. In the data pre-processing stage data augmentation was applied to training set to increase the data and address the issue of data imbalance. This also aided to prevent the model from being biased.

For fully automatic classification and ensemble learning technique with grid-search algorithm for fine tuning the hyper parameters were used in the multi- classification of brain tumors in the research paper (31), by researchers. They used a cascading technique, initially starting form binary classification of brain image. That is the detection of brain tumors, if a brain tumor is there it is then passed on to classify the type of brain tumors and finally grade or stage of tumor is classified. The fully automatic capability of this model may be applicable for the clinical diagnosis.

3. PROBLEM STATEMENT

The lack of medical image data hinders creating a real-world deep learning model. Researchers have employed Transfer Learning to overcome data availability challenges, including fine-tuning a model pre-trained with generic images(12,24). Existing state-of-the-art DL models contain considerable errors when introduced to unseen data (11) due to overfitting. Hence, it can be derived that lack of training data causes overfitting.

Most models are trained to classify just malignant or benign tumors. Even if the model is doing well in binary classification, radiologists must manually classify brain tumors to their types. The latter phase is more important since it allows clinicians to develop patient treatment plans. Most of the existing models perform well for binary classification, but not for multi classification (27,31).

In addition to overfitting, models are known to be biased due to insufficient training data (20,32). Biased models impede generalization, not only resulting in the model being regarded a narrow model but also reducing its clinical diagnostic use. Even though the output task remains the same, these models won't work if the input domain changes.

Most research in this sector has been motivated by a desire to increase model accuracy. Some study articles (9,12,14) with improved accuracy lack performance matrices such validation loss graph, f1-score, and confusion matrix.

Several studies have been done to address the above issues; however, the clinical usage of DL models have been limited due to the above-mentioned constraints. With all the developments

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and popularity on DL models, the question is, why DL models are fully not utilized in the clinical diagnosis? And can we minimize the computational complexity by using only simple DL model, that is more efficient?

The objective of this study focuses on the above research questions, by addressing to minimize the overfitting and bias issue. Minimization of the overfitting and bias issue will improve the efficiency of the model. Additionally, minimizing the error of model for multi-classification of tumors to meningioma, glioma and pituitary tumors.

A simple, efficient DL model will be more advantageous to the medical industry, consequently increasing the usage of DL models in clinical diagnosis and aiding radiologists.

To implement such model, with limited data can be a challenging task. However, in this study, we have used limited Data (from FIGSHARE) to train our model.

Efficiency of a model should be evaluated by proper performance matrices, and how the model performs on multi-classification scenarios. Hence, relevant performance matrices like, confusion matrices, training and validation loss graph and f1-score were used.

4. METHODOLOGY

This study will be undertaken in three phases. Initially, medical images of the brain will be collected from MRI modalities. These are weighted T1-CE images from FIGSHARE. During this initial phase, data will also be pre-processed. The second step will be to modify a Deep Learning Model that can determine the underlying representation of the presented data. Following training, the model will be evaluated and, based on its accuracy, it will be fine-tuned, and the training and testing process will be repeated until the model achieves a high accuracy score. And the final phase will involve testing this model using performance measures and comparing it to the baseline model for this study.

4.1 Data

T1-CE Weighted images of brain will be taken from FIGSHARE. FIGSHARE contains 3064 T1 contrast enhance image slices of a total of 233 patients. This MRI images will contain all three types of brain tumor, Glioma, Meningioma and Pituitary Tumor. All the data taken will be converted to JPG from MAT. As shown in the figure 1 after image conversion data set is split in to training, validation and test set. Dataset will be randomly distributed, in the ratio of 80% to training, 10% for validation and 10% for testing.

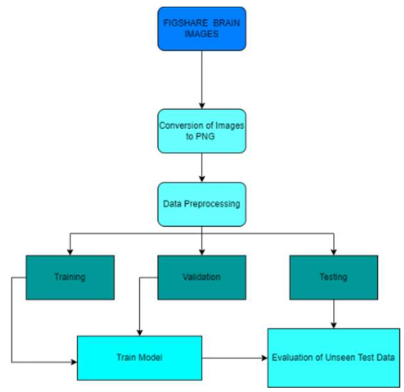

Figure 1: Proposed Data Preparation Method

For the imbalance of data and overfitting of the model, data augmentation was applied. This technique has been used by researchers to minimize the overfitting issues and model bias (22). Table 1 shows the data augmentation technique used in the research.

Table 1: Data Augmentation Technique

No	Data Augmentation Technique
1	Random Rotation
2	Random Vertical and Horizontal Flip
3	Shear
4	Sharpen
5	Edges Detection
6	Skew (Tilt)

4.2 Model Designing

By adopting the TL approach, GoogleNet and other successful models have been employed in medical imaging in the paper (10), GoogleNet was successfully employed to detect brain tumors following the application of TL. GoogleNet and other pre-trained models have been demonstrated to be successful in various researches (8) and (9).

Szegedy et al. created GoogleNet in 2014 and it is the first ILSVRC 2014 winner trained on the ILSVRC dataset (14). As shown in figure 2, the design includes nine inception modules, two convolutional layers, four max-pooling layers, one convolutional layer for dimension reduction, one average pooling layer, two normalization layers, one fully connected layer, and lastly a linear layer with SoftMax activation at the output. Furthermore, each

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

inception module has one max-pooling layer and six convolutional layers, four of which are employed for dimension reduction. Dropout regularization is employed in all completely connected layers, while the ReLU activation function is used in all fully connected layers. Furthermore, it outperforms AlexNet on the original ILSVRC dataset (12). GoogleNet is significantly slow compared to VGG16 or VGG19, though there is an inbuilt drop-out layer within the GoogleNet architecture. Hence, the model is more efficient compared to other models. This is because the gradient vanishing issue is minimized by the said layer. Additionally, it also helps in minimizing the overfitting issue.

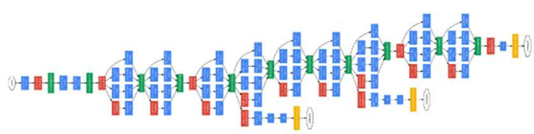

Figure 2: GoogleNet Architecture

However, the classification accuracy of GoogleNet is worse than that of VGG and ResNet (2)(12) with limited data. To overcome this problem, GoogleNet is utilized as a feature extractor, and classification is done only by adding a machine learning algorithm like SVM or KNN. Although this has a major impact on GoogleNet performance, the computational complexity means that training will take longer.

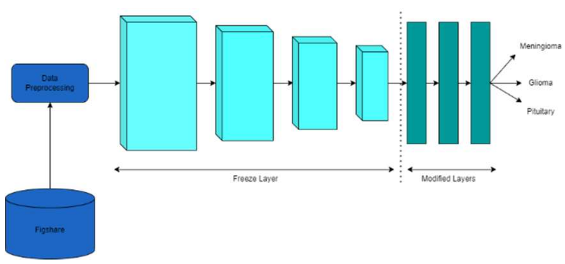

Figure 3: Proposed Model

Figure 3 illustrates the proposed model for this research and the model's workflow. The initial layers are frozen because they have been trained using the ImageNet Database. The model supplied has already been trained. All low-level features such as lines, edges, and contours are already captured in the initial layers. The modifications are done to the final GoogleNet layers. Using categorical cross-entropy, the final layers are adjusted to classify brain tumors into three sorts or categories. Utilized were hyper parameters such as Adam optimizer. Utilizing the Adam optimizer will reduce computational time complexity, and it is effective for noisy or sparse problems.

4.3 Measurement Metrices

The performance of the proposed model will be evaluated based on its Confusion Matrix, F1-Score, Accuracy, and Precision. In addition, the Validation Loss graph will be evaluated to assess model overfitting. The ground-truth and expected outcome will be evaluated using a confusion matrix. In the confusion matrix, the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) will be represented in a matrix format. Accuracy is the summation of a model's performance across all classes. It is advantageous when all classes are of similar importance. It is computed by dividing the number of accurate predictions by the total number of forecasts.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The Precision is determined as the ratio of the number of correctly categorized Positive samples to the total number of Positive samples. The precision of the model in identifying a sample as positive is measured. It indicates how accurate the model is in classifying Positive samples. The harmonic means of precision and recall values which are calculated gives the F1- Score. Precision and Recall is calculated as shown in the equation 2 and 3.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall}=\frac{TP}{TP + FN} \tag{3}$$

In the above equation for the precision, True Positive (TP) is divided by adding TP to False Positive (FP). And in recall the only thing that changes is the FP in the denominator. Instead of FP, False Negative (FN) is added. F1- Score is calculated as shown in equation 4.

$$\text{F1-Score} = 2 * \frac{\text{Precission} * \text{Recall}}{\text{Precission} + \text{Recall}} \tag{4}$$

5. RESULTS

The experiment was performed on Kaggle utilizing the following hardware: two virtual central processing units (vCPU) from Intel Xeon with a base

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

frequency of 2.4 GHz, 13 GB of RAM, and Nvidia Tesla 4 for the GPU (Graphical Processing Unit).

Experiment was conducted using a dataset that is available in FIGSHARE. Two experiments were undertaken to examine the impact of training data size on the model. In the first experiment, the data set was split so that 80 percent was used for training, 10 percent for validation, and 10 percent for testing accordingly. In the second experiment, the dataset was segmented as follows: 50 percent for training, 25 percent for validation, and 25 percent for testing. Evaluation of the suggested model was done using a variety of different optimizers for each one of the experiments.

For the classification of brain tumors into Meningioma, Pituitary and Glioma, the proposed modified GoogleNet was used. The model was tested using four learning algorithms, Adam (Adaptive Moment Estimation) and AdaGrad (Adaptive Gradient Descent). One cycle learning rate (33) was used during the training of model. After feature extraction the information is passed through a modified sequential layer with two linear layers, ReLU as an activation function. For each learning rate same data split was used. Epochs were set to 40 and batch size was set to 124.

Classification accuracy is an important measurement for any model. The overall accuracy of a model shows how well it is performing. However, only accuracy score is not enough for the evaluation of a model. As the dataset used in this experiment was an imbalance dataset. It should be noted that if the dataset is a balance dataset accuracy score can be enough to evaluate the model. Though, in real world to get a balance dataset could be difficult. Even more if the dataset is a medical dataset. Accuracy score for the model with different optimizers are depicted in the table 02.

Table 2: Optimizers and Their Accuracy

Optimizers	Accuracy
Adam	98.69 %
AdaGrad	95.75 %

For the imbalanced data accuracy as mentioned earlier does not give a very good information about the model. In Figure 4, 5 shows the F1-Score, Precision, Recall of the model with different optimizers.

	precision	recall	f1-score
meningioma	0.97	0.97	0.97
glioma	0.99	0.99	0.99
pituitary	1.00	0.99	0.99
accuracy			0.99
macro avg	0.99	0.98	0.99
weighted avg	0.99	0.99	0.99

Figure 4: Precision, Recall, F1- Score with Adam Optimizer

	precision	recall	f1-score
meningioma	0.92	0.92	0.92
glioma	0.96	0.98	0.97
pituitary	0.99	0.96	0.97
accuracy			0.96
macro avg	0.95	0.95	0.95
weighted avg	0.96	0.96	0.96

Figure 5: Precision, Recall, F1-Score with AdaGrad Optimizer

For the imbalanced data accuracy as mentioned earlier does not give a very good information about the model. In Figure 4, 5 shows the F1-Score, Precision, Recall of the model with three different optimizers.

The accuracy of the proposed model was higher when Adam optimizer was used during the training, compare to other learning algorithms. Figure 6 and Figure 7 shows the confusion matrix of the experiment carried out using Adam optimizer and AdaGrad respectively. Additionally, figure 8 and figure 9 shows the validation loss graph, that steeps down and converges within almost epoch 10 and epoch 5. Figure 10 and 11 show sample of the predicted images for Adam and AdaGrad optimizers, respectively.

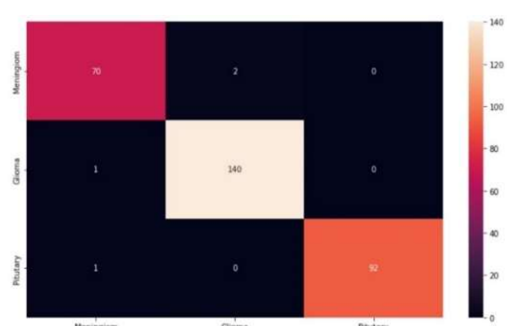

Figure 6: Confusion Matrix with Adam Optimizer

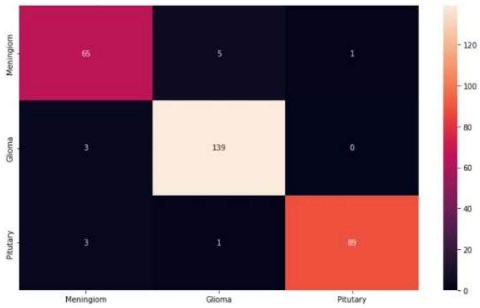

Figure 7: Confusion Matrix with AdaGrad

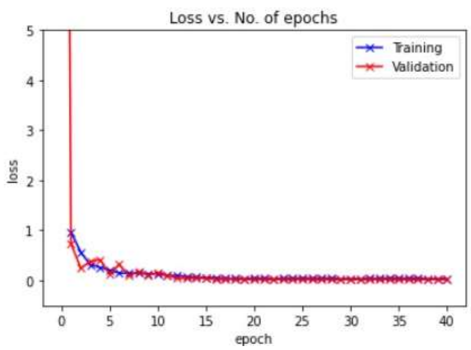

Figure 8: Training and Validation Loss Graph for Adam Optimizer

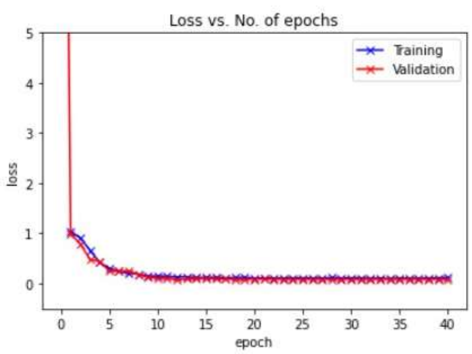

Figure 9: Training and Validation Loss Graph for AdaGrad Optimizer

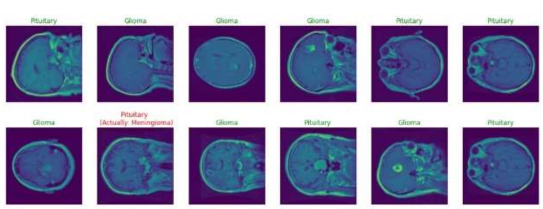

Figure 10: Sample of Predicted Images with Adam Optimizer

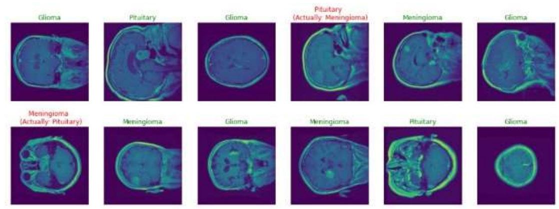

Figure 11: Sample of Predicted Images with AdaGrad Optimizer

Experiment was also done to evaluate the model without using data augmentation. Data augmentation is an important concept when there is a limited data availability. As mentioned earlier medical image data unlike most of other image data, is not easily available. Due the data privacy issues most of the medical images are not put into free repositories. Additionally, annotated medical data is immensely difficult to get. Since the Adam optimizer was identified as the best learning algorithm for this dataset, furthermore experiment was carried out without Data augmentation. The F1 average score of this experiment was 86% and accuracy score was 88% as shown in figure 12.

	precision	recall	f1-score
meningioma	0.74	0.82	0.78
glioma	0.92	0.93	0.92
pituitary	0.94	0.84	0.89
accuracy			0.88
macro avg	0.86	0.86	0.86
weighted avg	0.88	0.88	0.88

Figure 12: Precision, Recall and F1-Score without Data Augmentation

Figure 13: Confusion Matrix without Data Augmentation

To determine the impact of transfer learning, a second experiment was conducted excluding transfer learning. This raised the training duration from twenty minutes to two hours. The outcomes are depicted in figures 14, 15 and16. The

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

F1- Score of the aforementioned configuration (without TL) was 79%, showing a significant decline in the accuracy of the model.

	precision	recall	f1-score
meningioma	0.70	0.62	0.66
glioma	0.80	0.92	0.86
pituitary	0.92	0.78	0.84
accuracy			0.81
macro avg	0.81	0.77	0.79
weighted avg	0.81	0.81	0.81

Figure 14: Precision, Recall and F1-Score without Transfer Learning

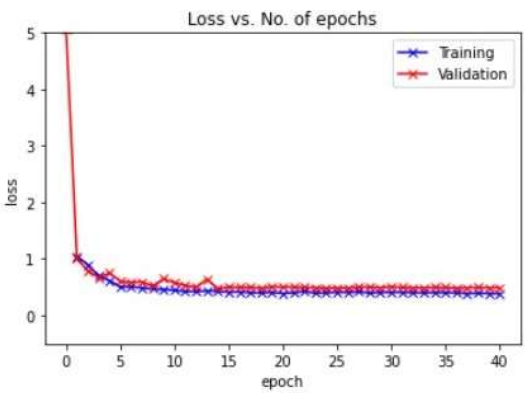

Figure 15: Training and Validation Loss Graph without Transfer Learning

Figure 16: Confusion Matrix without Transfer Learning

6. DISCUSSION

The aim of doing an experiment to analyses the learning algorithm or optimizers was to find the best optimizer for the proposed model. From the results perceived from Figure 4 and 5, F1-score for the two optimizers are 98.69% and 95.75%, for Adam and AdaGrad respectively. Figures 9 and 10 illustrate the training and validation loss graphs from this experiment. Both graphs show a steep slope, with the training and validation lines almost converging. Figure 7 indicates that Adam optimizer successfully identified 70 of 72 Meningiomas and 2 as Gliomas. 140 Gliomas were successfully identified, while 1 was Meningioma. 91 of 92 Pituitary tumors were accurately classified, while 1 was Meningioma. Figure 8 displays AdaGrad confusion matrix, which exhibits lower values for successfully identified examples. Both experiments used the same dataset and distribution (80 for training, 10 for validation, and 10 for testing) with 40 epochs and 124 batches.

From the carried out experiments it is inferred that the learning algorithms or optimizers and Data augmentation are important parameters in deep learning models. These two attributes can make a significant change to a model's behavior. With limited data a model can perform very well with a correct optimizer and with data augmentation using transfer learning. To compare out model to the state-of are models with other researchers a compression is shown in the table 03, that have used FIGSHARE dataset.

Table 3: Proposed Model Comparison with State-Of-Art Models.

Work	Method	Accuracy
(29)	Pre-Trained VGG19 CNN	94.58%
(25)	Block-Wise VGG19	94.82%
(12)	Deep CNN + SVM	97.1%
(3)	[ResNet18 + ShallowNet] + SVM	98.02 %
Proposed Model	Modified GoogleNet with Transfer Learning	98.69%

Comparing the results of figure 10, with a f1 score of 86% shows that the model does not perform well without data augmentation. Furthermore, this is evident from the confusion matrix shown in figure 20. Correctly classified Meningioma is 59, Glioma 131 and Pituitary tumor is 78. The highest accuracy was Glioma with 92% while other two types of tumor Meningioma and Pituitary had an accuracy of 81% and 83% respectively. With the Data augmentation the F1 average score was 98.33% and Accuracy was 98.69%. Most of the researchers, publish the model accuracy score, however, with imbalance datasets these score does not help in evaluating the model's efficiency. Hence, F1 average score is shown in this study. Comparing to the accuracy scores in table 03, the proposed models F1-average score is still higher.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Adam optimizer performance is evidently better than other optimizers used in the experiment. Adam optimizer is a technique for calculating adaptive learning rates for each parameter. It holds the decaying average of previous gradients, comparable to momentum, as well as the decaying average of previous squared gradients, akin to RMS-Prop and AdaDelta. As a result, it incorporates the benefits of both strategies. It also has an advantage over other optimizers, as it uses less memory and computationally efficient (12). The performance drop of the model with other learning algorithms like SGD, AdaGrad shows that the significant role it plays in the convergence.

To further analyze the effect on the model with limited data, dataset was split to 50% for training, 25% for validation and 25% for testing. The accuracy and the F1-average of the model drops 98.17% and 98% respectively as shown in figure 17. This drop is not significant, and hence, proves that with data augmentation and efficient optimizer, the model can perform efficiently.

	precision	recall	f1-score	support
meningioma	0.94	0.98	0.96	177
glioma	0.99	0.98	0.99	366
pituitary	1.00	0.99	0.99	222
accuracy			0.98	765
macro avg	0.98	0.98	0.98	765
weighted avg	0.98	0.98	0.98	765

Accuracy of the network on the 3369 test images: 98.17%

Figure 17: Precision, recall and F1-score after data split (0.5, 0.25, 0.25)

This study demonstrates that GoogleNet is an effective deep learning model. Moreover, its architecture incorporates dropout layers, giving it an advantage over the majority of models. Dropout is essential for generalizing and mitigating model overfitting problems. Adding drop-out technique or more layers to a modified layer will increase the model's complexity. It is also known through a survey of the relevant literature that the risk of vanishing gradient decent increases with the depth of the layers. The problem of vanishing gradient decent was handled in GoogleNet using auxiliary layers. Because GoogleNet was designed with drop-out layers and auxiliary layers for addressing overfit and bias issues, a more commonly used model, such as ResNet or VGG 19, was not used to solve the stated problem in this study.

7. CONCLUSION AND FUTURE WORKS

In this study, we proposed a modified version of the GoogleNet model and used data augmentation and transfer learning techniques. During the training, our model was shown to be more effective than a few other models that are state of the art. With the data that was available, our model was able to correctly classify three distinct forms of brain tumors.

Researchers, in their paper (12), have used the GoogleNet with a Support Vector Machine (SVM) to classify brain tumors. The use of SVM (a machine learning technique) and a Deep learning model like GoogleNet can lead to more processing time and computational complexity. In this research, we have proven that GoogleNet with a modified layer can perform better compared to the model used in the above-mentioned research paper.

Generalization of deep learning models are an ongoing research. The proposed model's generalizability can only be evaluated using a new MRI dataset. However, it is assumed that the proposed model will be generalizable compare to other models, as it uses drop-out.

In future works, research will be conducted on domain adaptation and multi-task learning adaption in a hybrid model. This will be done so that the limitations of the proposed model can be circumvented. In addition, Natural Language Processing (NLP) will be utilized to generate medical reports that are nearly on par with those produced by humans.

Currently, there are also studies being conducted on the topic of establishing an effective model that can produce medical reports better or equivalent to radiologists (34). Researchers are devoting an increasing amount of their time and energy to the study Explainable Artificial Intelligence. Because it is anticipated that this will revolutionize the way that most people think of DL, which is currently viewed as a black box. However, this is still not possible due to the limited computer power that is now available. Despite the fact, it is possible to say that Explainable Artificial Intelligence can be accomplished with the development of a more stable quantum computing system.

Even if Artificial Intelligence (AI) is becoming more advanced daily and more deep learning models are demonstrating that they are more effective in categorization problems, the use of these models in the actual world is still somewhat limited in the medical industry. As a consequence of this, it is of the utmost need to carry out qualitative study about the application of DL models by radiologists. In order to design a model for medical image diagnosis that is more effective, it will be vital to gain an understanding of the perspective of medical professionals.

ACKNOWLEDGEMENT

This work is supported by Ministry of Higher Education under Grant FRGS FRGS/1/2020/ICT02/UPM/02/2, project code 08-01-20-2315FR.

REFRENCES

[1] Panayides AS, Amini A, Filipovic ND, Sharma A, Tsaftaris SA, Young A, et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. IEEE J Biomed Health Inform. 2020 Jul;24(7):1837–57.

[2] Kesav N, Jibukumar MG. Efficient and low complex architecture for detection and classification of Brain Tumor using RCNN with Two Channel CNN. Journal of King Saud University - Computer and Information Sciences. 2021;(xxxx).

[3] Öksüz C, Urhan O, Güllü MK. Brain tumor classification using the fused features extracted from expanded tumor region. Biomed Signal Process Control. 2022;72(November 2021):103356.

[4] Dutta P, Upadhyay P, De M, Khalkar RG. Medical Image Analysis using Deep Convolutional Neural Networks: CNN Architectures and Transfer Learning. Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020. 2020;175–80.

[5] Amin J, Sharif M, Haldorai A, Yasmin M, Nayak RS. Brain tumor detection and classification using machine learning: a comprehensive survey. Complex & Intelligent Systems. 2021;

[6] Karimi D, Warfield SK, Gholipour A, Member S. Transfer Learning in Medical Image Segmentation: New Insights from Analysis of the Dynamics of Model Parameters and Learned Representations. 2020;116(April).

[7] Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. Comput Methods Programs Biomed. 2020;187:105242.

[8] Mehrotra R, Ansari MA, Agrawal R, Anand RS. A Transfer Learning approach for AI-based classification of brain tumors. Machine Learning with Applications. 2020;2(September):100003.

[9] Kamran I, Naz S, Razzak I, Imran M. Handwriting dynamics assessment using deep neural network for early identification of Parkinson ' s disease. 2021;117:234–44.

[10] Talo M, Baloglu UB, Yıldırım Ö, Rajendra Acharya U. Application of deep transfer learning for automated brain abnormality classification using MR images. Cogn Syst Res. 2019;54:176–88.

[11] Ahn E, Kumar A, Feng D, Fulham M, Kim J. Unsupervised deep transfer feature learning for medical image classification. ArXiv. 2019;(Isbi):2019–22.

[12] Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med [Internet]. 2019;111(March):103345. Available from: https://doi.org/10.1016/j.compbiomed.2019.103345

[13] Sekhar A, Biswas S, Hazra R, Sunaniya AK, Mukherjee A, Yang L. Brain Tumor Classification Using Fine-Tuned GoogLeNet Features and Machine Learning Algorithms: IoMT Enabled CAD System. IEEE J Biomed Health Inform. 2022;26(3):983–91.

[14] Rehman A, Naz S, Razzak MI, Akram F, Imran M. A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning. Circuits Syst Signal Process [Internet]. 2020;39(2):757–75. Available from: https://doi.org/10.1007/s00034-019-01246-3

[15] Tandel GS, Tiwari A, Kakde OG. Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. Comput Biol Med. 2021;135(February).

[16] Talo M, Yildirim O, Baloglu UB, Aydin G, Acharya UR. Convolutional neural networks for multi-class brain disease detection using MRI images. Computerized Medical Imaging and Graphics. 2019;78:101673.

[17] Kumar S, Dabas C, Godara S. Classification of Brain MRI Tumor Images: A Hybrid Approach. Procedia Comput Sci. 2017;122:510–7.

[18] Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM. Classification using deep learning neural networks for brain tumors.

Future Computing and Informatics Journal. 2018;3(1):68–71.
[19] Gull S, Akbar S, Khan HU. Automated Detection of Brain Tumor through Magnetic Resonance Images Using Convolutional Neural Network. Biomed Res Int. 2021;2021.
[20] Lu S, Lu Z, Zhang YD. Pathological brain detection based on AlexNet and transfer learning. J Comput Sci [Internet]. 2019;30:41–7. Available from: https://doi.org/10.1016/j.jocs.2018.11.008
[21] Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS. Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Applied Sciences (Switzerland). 2017;8(1).
[22] Arbane M, Benlamri R, Brik Y, Djerioui M. Transfer Learning for Automatic Brain Tumor Classification Using MRI Images. 2021;210–4.
[23] Polat Ö, Güngen C. Classification of brain tumors from MR images using deep transfer learning. Journal of Supercomputing. 2021;77(7):7236–52.
[24] Mehrotra R, Ansari MA, Agrawal R, Anand RS. A Transfer Learning approach for AI-based classification of brain tumors. Machine Learning with Applications [Internet]. 2020;2(July):100003. Available from: https://doi.org/10.1016/j.mlwa.2020.100003
[25] Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, et al. Brain tumor classification for MR images using transfer learning and fine-tuning. Computerized Medical Imaging and Graphics. 2019;75:34–46.
[26] Ghassemi N, Shoeibi A, Rouhani M. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. Biomed Signal Process Control. 2020;57:101678.
[27] Srikanth B, Venkata Suryanarayana S. Multi-Class classification of brain tumor images using data augmentation with deep neural network. Mater Today Proc [Internet]. 2021;(xxxx):1–7. Available from: https://doi.org/10.1016/j.matpr.2021.01.601
[28] Sasank VVS, Venkateswarlu S. An automatic tumour growth prediction based segmentation using full resolution convolutional network for brain tumour. Biomed Signal Process Control. 2022;71(PA):103090.
[29] Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. J Comput Sci. 2019;30:174–82.
[30] Alqudah AM, Alquraan H, Qasmieh IA, Alqudah A, Al-Sharu W. Brain tumor classification using deep learning technique - A comparison between cropped, uncropped, and segmented lesion images with different sizes. International Journal of Advanced Trends in Computer Science and Engineering. 2019;8(6):3684–91.
[31] Irmak E. Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework. Iranian Journal of Science and Technology - Transactions of Electrical Engineering [Internet]. 2021;45(3):1015–36. Available from: https://doi.org/10.1007/s40998-021-00426-9
[32] Nusrat I, Jang SB. A comparison of regularization techniques in deep neural networks. Symmetry (Basel). 2018;10(11):1–17.
[33] Smith LN, Topin N. Super-convergence: very fast training of neural networks using large learning rates. 2019;36.
[34] Han Z, Wei B, Xi X, Chen B, Yin Y, Li S. Unifying neural learning and symbolic reasoning for spinal medical report generation. Med Image Anal. 2021;67:101872.