An enhanced deep learning approach for brain cancer MRI images classification using residual networks

Sarah Ali Abdelaziz Ismael, Ammar Mohammed, Hesham Hefny

Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt

ABSTRACT

Cancer is the second leading cause of death after cardiovascular diseases. Out of all types of cancer, brain cancer has the lowest survival rate. Brain tumors can have different types depending on their shape, texture, and location. Proper diagnosis of the tumor type enables the doctor to make the correct treatment choice and help save the patient's life. There is a high need in the Artificial Intelligence field for a Computer Assisted Diagnosis (CAD) system to assist doctors and radiologists with the diagnosis and classification of tumors. Over recent years, deep learning has shown an optimistic performance in computer vision systems. In this paper, we propose an enhanced approach for classifying brain tumor types using Residual Networks. We evaluate the proposed model on a benchmark dataset containing 3064 MRI images of 3 brain tumor types (Meningiomas, Gliomas, and Pituitary tumors). We have achieved the highest accuracy of 99% outperforming the other previous work on the same dataset.

1. Introduction

Cancer is a significant health problem in the world nowadays. Every sixth death in the world is due to cancer, making it the second leading cause of death after cardiovascular diseases [1]. Among different types of cancer, brain tumors are considered one of the deadliest forms due to its aggressive nature, heterogeneous characteristics, and low survival rate. Brain tumors can have different types depending on several factors such as the shape, texture, and location of the tumor (e.g., Acoustic Neuroma, Meningioma, Pituitary, Glioma, CNS Lymphoma, etc.) [2]. In clinical practice, the incident rates of Glioma, Meningioma, and Pituitary tumors are approximately 45%, 15%, and 15%, respectively, among all brain tumors [3]. Depending on tumor type, doctors are able to diagnose and predict patient survival, also they can decide on the appropriate choice of treatment which can range from surgery, followed by chemotherapy and radiotherapy, to a "wait and see" approach which avoids invasive procedures. Hence, tumor grading is an important aspect of treatment planning and monitoring [4].

Magnetic Resonance Imaging (MRI) is a noninvasive, pain-free medical imaging procedure that is used for delivering excellent pictures of the human body organs in 2D and 3D formats. It is widely used and considered to be one of the most accurate techniques for cancer detection and classification, due to its high-resolution images on the brain tissue [5]. However, cancer type recognition based on MRI images is a challenging, error-prone, highly specialized task that depends on the experience of the radiologist, and more importantly, it is a time-consuming procedure. Additionally, tumors may have various shapes and there may not be enough visible landmarks in the image to contribute

to an accurate decision. Therefore, we can conclude that human diagnosis is typically unreliable. Moreover, the misdiagnosis of brain tumor type can be a serious issue since it will prevent the effective response to medical intervention and decrease the chance of survival among patients. On the other hand, the correct diagnosis will help the patient start the correct treatment immediately, and live a longer life. Accordingly, this necessitates an urgent need in Artificial Intelligence (AI) field to develop and design a new and innovative Computer Assisted Diagnosis (CAD) system, aiming to relieve the workload of the diagnosis and classification of tumors and act as a helping tool for doctors and radiologists.

Recently, CAD has been an active area of research and development in medical imaging and diagnostic radiology, especially for cancer classification and diagnosis. Typically, a CAD system is consisting of three steps; segmenting the lesions from the image, extracting characteristics of the segmented tumors on basis of statistical or mathematical parameters analysis obtained during the learning process of a set of labeled MRI images, then applying a proper machine learning classifier to predict the class of abnormality [6]. Many traditional machine learning approaches required lesion segmentation prior to the classification step. The process of segmentation is a computationally intensive step, that may be inconsistent depending on the variance in image contrast and intensity normalization, and may affect the performance of the classification [7]. Feature extraction is a crucial procedure used to generate interesting parts – called features – that can describe the contents of a raw image. However, this stage had some negative effects; it is a time-consuming task that requires prior knowledge about the problem domain. Classifying tumor types based

on morphological features can easily lead to misclassification because of the similar appearance of various types of tumors. Features are then used as inputs to machine learning classifiers which in turn assign a class label to the image based on those important features [8]. There are several approaches for MRI brain tumor classification using the later-described machine learning approach; such as the work of [6,9–23]. Deep learning on the other hand [24–36], is a subset of machine learning that does not require handcrafted features. Deep learning has been successfully proven to reduce the gap between human vision and computer vision in pattern recognition and can achieve a higher classification performance than conventional machine learning methods [7]. Thus, several works have been adopting deep learning techniques for brain cancer diagnosis [25–36]. The essence of these works is to find the best deep learning model and architecture that optimally classifies brain cancer. However, optimization processes are applied to find the right architectures and hyper-parameters of the model.

This paper proposes an enhanced deep learning model for brain tumor classification from MRI images. The proposed model is trained on a benchmark brain tumor MRI images dataset [37] and is evaluated under several performance metrics. The paper is doing so by extending the work done previously using the Residual Network (ResNet) architecture [38] in order to improve performance. Also, we evaluate the proposed model based on different metrics such as accuracy, precision, recall, f1-score, and balanced accuracy. By these types of metrics, we can have a clear indicator of the generalization of the model over the imbalanced MRI images dataset. The experimental result shows that the proposed deep learning model is competitive when compared with the other approaches.

The rest of this paper is organized as follows: Section 2 describes other related work. Section 3 gives a review background on deep learning architecture. Section 4 describes the proposed model. Section 5 introduces the experimental results and the evaluation of the model. Section 6 compares our results with the other published work. Section 7 finally concludes the paper.

2. Related work

There have been many attempts over the past years to create an automated system for classifying brain tumors from MRI images. Several authors have used traditional machine learning approaches, going through sequential stages starting with preprocessing the images, then feature extraction process, followed by feature selection process to reduce the feature size, and finally using a classification algorithm to get the final output. Different schemes for feature extraction were used such as Discrete Wavelet Transform (DWT) [9–12,19,22,15,6], Gray Level Co-occurrence Matrix (GLCM) [6,13,20,16,17,23,39], Histogram of oriented gradients (HOG) [21], Genetic Algorithm [40,14] and Zernike Moments [22]. For the feature selection process, several authors used Particle Swarm Optimization (PSO) [9] and Principle Component Analysis (PCA) [10,17,12,13,15,18]. For the classification task, the most popular classification method was Support Vector Machines (SVM) and it was used by several authors like [6,9–18]. Other authors used different classification methods such as Random Forest [19], Adaboost machine learning algorithm [20], Instance-based K-Nearest using Log and Gaussian weight Kernels [21,18], Extreme learning machine (ELM) [22] and Sequential Minimal Optimization (SMO) [23]. However, the main problem in the previous machine learning studies is that the manual feature extraction is a tedious, time-consuming, and error-prone process. It also requires previous knowledge about the domain. Some of the previous approaches also required region-based tumor segmentation prior to feature extraction and classification. Also, feature selection requires to further reduce the selected features before preforming classification and no single feature extraction method is applicable in general.

Deep learning [24] is a subset of machine learning that does not require handcrafting features; however, it needs to be applied optimally on preprocessed data and using the right architectures and hyper-parameters to obtain better results. An example of the deep learning method is to use Convolutional Neural Networks (CNNs). The later is used in most brain tumor related research due to the advanced image processing and faster computation feasibility. Several efforts have proposed CNN for brain cancer diagnosis [25–31]. These efforts aim at finding the best model and network architecture that automatically increases the performance of brain cancer classification. However, the previous efforts have been done using relatively smaller datasets compared to the work presented in this paper, as medical datasets are limited and very hard to collect.

Similar to the work of this paper, several authors have used the same benchmark brain tumor dataset [37] (discussed in Section 4.2) to obtain a good brain tumor classifier. The data was first used by the authors in [41]. They used a traditional machine learning approach by using three feature extraction methods: intensity histogram, gray level co-occurrence matrix (GLCM), and bag-of-words (BoW) model to increase accuracies of classifying brain tumors to up to 91.28%. Other authors in [42,33] have used Convolutional Neural Networks to classify the brain tumor. They achieved an accuracy of 91.43% and 84.19% respectively. The authors in [35,34] have proposed a Capsule Networks architecture based on Convolutional Neural Networks. They incorporated both the raw MRI brain images and the tumor course boundaries in order to classify the brain tumors. They have achieved an accuracy of 90% and 86% respectively. Finally, A. Pashaei et al. [36] combined a Convolutional Neural Network with other machine learning methods as KELM achieving accuracy of 93.8%.

3. Background

Deep learning has shown promising performance over the recent years in different domains such as computer vision [43], speech recognition, text recognition, robotics [24] and computer-aided diagnosis [44]. As deep learning models are fed with raw data, they have the ability to learn multiple levels of abstraction, representation, and information automatically from a large set of data. Moreover, deep learning models have a huge advantage over traditional machine learning algorithms which are limited in processing natural images in their raw form, time-consuming, based on expert knowledge and require a lot of effort for tuning the features. Broadly speaking, there are several architectures for deep neural networks. Convolutional Neural Networks (CNN) is one type that is widely used in image and video recognition, speech processing and natural language processing. CNN has been inspired biologically by the working of the visual cortex in cats and spider monkeys [43]. CNN has been successfully applied mostly in image processing due to its ability to recognize patterns in images [45]. Generally, a CNN contains mainly three types of layers namely convolutional, pooling and fully-connected layer. These layers are stacked on top of each other to form the architecture of CNN as shown in Fig. 1.

The convolutional layer is the main layer in a CNN. This layer is responsible for feature extraction such as edges and colors of the image. It is what gives us a network with a visual understanding of the dataset images. This layer focuses on the use of learnable kernels. These kernels are usually small in spatial dimensionality but spread along the entirety of the depth of the input. Usually, the convolutional layer convolves each filter across the spatial dimensionality of the input to produce a feature map as its output. The second layer type is the pooling layer, which is responsible for reducing the dimensionality of the convolved features, and thus reducing the number of parameters and the computational power complexity of the model. The last type of layer is the fully-connected layer which contains neurons of which are directly connected to the neurons in the two adjacent layers, without being connected to any layers within them. Every neuron in this layer is connected to every neuron in the previous layer [43]. The purpose of the fully-connected layer is to achieve linearity in the networks for high-level reasoning.
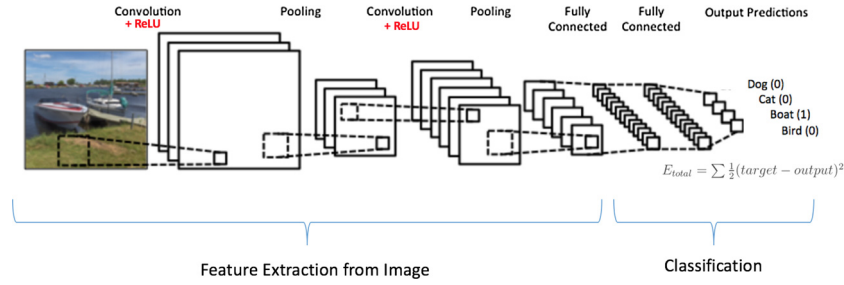
Fig. 1. Convolutional neural networks layers.

Like other neural network architectures in the training phase, CNN uses a learning or optimizer algorithm to update the weights of the neural network. The learning algorithm takes a classification error or loss as input and back propagates the error into the network to update the filters and the weights. A SoftMax activation function is used at the output layer to normalize the output sum such that all numbers at the output neuron will add up to one.

It has been observed that the deeper a Convolutional Neural Network gets, the more complex tasks it can solve and improve the model accuracy. However, as the network depth increases, the more its accuracy starts to saturate and degrade and the problem of exploding/vanishing gradients appears. Residual Network (ResNet) is a type of Convolutional Neural Network architecture introduced by Kaiming He et al which won first place at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 [38]. One major advantage of ResNet was that it allows the training of extremely deep neural networks by introducing the concept of shortcut connections to skip one or more layers, therefore solving the problem of degrading accuracy and vanishing gradient. Shortcut connections act as an identity mapping function, having its output added to the outputs of the stacked layers. ResNet is a modularized architecture that stacks multiple shortcut connectors building blocks [46], a building block is illustrated in Fig. 2 along with the difference between plain and Residual Network. The residual mapping function is described in Eq. (1) where x and y are the input and output vectors and F (x, Wi) is the residual mapping to be learned.

$$y = F(x, Wi) + x \tag{1}$$

4. Proposed approach

This section describes the design and implementation of the proposed model.

4.1. Proposed approach pipeline

We have proposed a pipeline for creating our deep learning model. The pipeline is composed of multiple stages starting with receiving the raw data and ending with producing the model classification output.

Each stage output in the pipeline is given as input to the subsequent stage. The pipeline of the proposed approach is shown in Fig. 3. And mentioned in detail below.

1. Data Acquisition: Data defines the task and plays a big part in the model performance. After obtaining the brain tumor dataset, we have applied the required conversions and preprocessing of the images to help the model perform at its best.
2. Data Preprocessing: Medical images datasets are smaller in size than the datasets available in other domains due to their limitations. We used different augmentation techniques to maximize our data on runtime, allowing the model to generalize and achieve high results.
3. Data Visualization: We have plotted the training data during preprocessing and augmentation phases to get a sense of the patterns in it.
4. Model Creation: A model is an algorithm that accepts input data X and predicts outcome Y. We have used Residual Network architecture for our model.
5. Model Training: Training is the process the algorithm undertakes to optimize its parameters and update weights for the classification model.
6. Model Evaluation: We evaluated our model using accuracy, precision, recall, f1-score and balanced accuracy.

4.2. Dataset description

Our work will be using a publicly available brain tumor dataset of 3064 T1-weighted contrast-enhanced MRI images containing three types of brain tumors with the highest percentage among brain tumors (Meningiomas, Gliomas, and Pituitary tumors). The dataset was acquired from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjing Medical University, China, during the period from 2005 to 2010. It contains images collected from 2D MRI scans from 233 cancer patients anonymously. The dataset first was processed by Cheng et al. [41] to create a classification model for brain tumor types. We have plotted some sample images from the dataset each with their class label printed above it in Fig. 5. The brain tumor dataset has the brain tumor type as ground truth: Meningioma (708 slices), Glioma (1426 slices), and Pituitary tumor (930 slices). A representation of the

Fig. 2. Plain and Residual Networks, a residual connection building block is illustrated at the top.
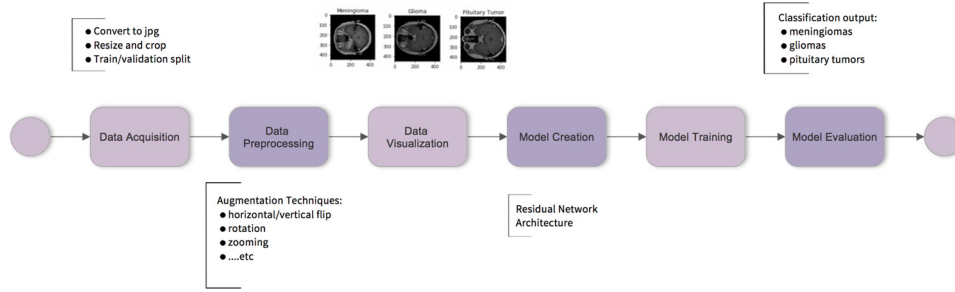
Fig. 3. Proposed approach pipeline.

distribution of the images per class is shown in Fig. 4. The brain CE-MRI images dataset is available at: (https://figshare.com/articles/brain_tumor_dataset/1512427/5).

4.3. Images preprocessing

The dataset was originally provided in MATLAB data format, each file stored a "struct" containing information about the image, such as a label that specifies the type of tumor as ground truth (1 for Meningioma, 2 for Glioma, 3 for Pituitary tumor), patient id, image data, tumor border coordinates and a binary mask image with 1s indicating tumor region. We created a python script using the "h5py" library to read ".mat" files stored in the "HDF5" format and extract the required attributes. Images were saved at their original size 512×512 pixels, each image was placed into a folder corresponding to its ground truth (tumor type). We also used the patient id attribute on each image for dataset organization as we are going to discuss next.

We split the dataset into training and validation sets with a ratio of 80% and 20% respectively as per the Pareto principle [47]. The 80-20 split ratio is one of the most common ratios in the deep learning field and was used in similar work in medical images such as [29,39]. We
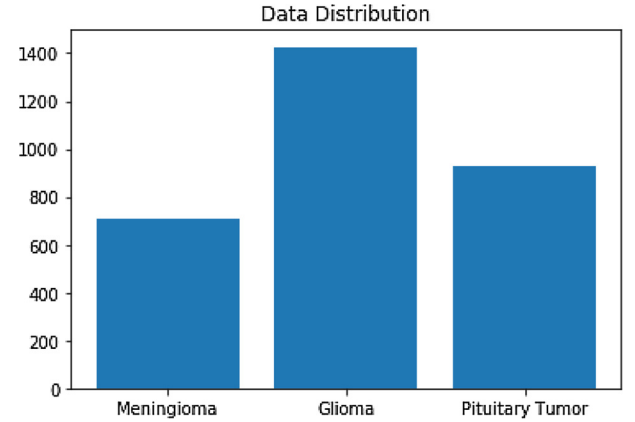

Fig. 4. Distribution of the brain tumor dataset class sizes.

have also tried re-sampling our data using other popular ratios like 70-30 and 75-25 getting an accuracy of 97% and 98% respectively. However, from our experiments, the 80-20 ratio showed the best result, by achieving an accuracy of 99%. Two approaches were used while
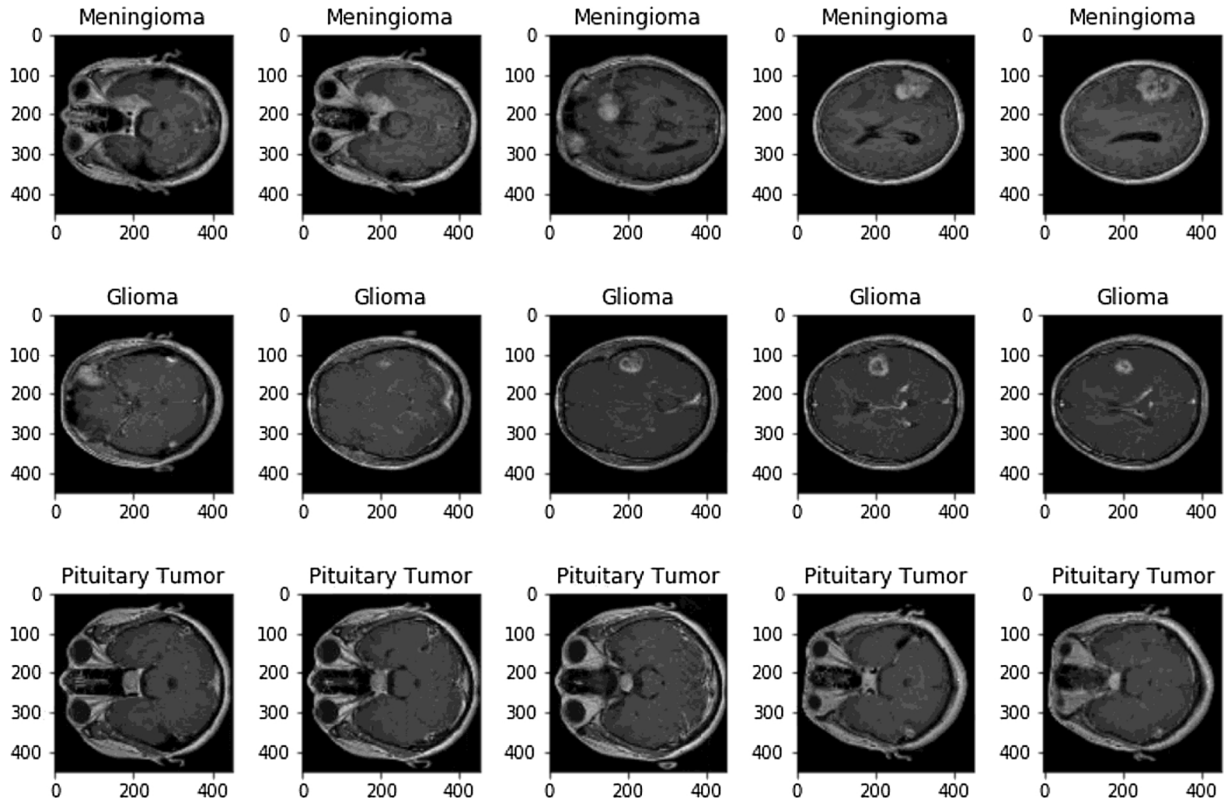

Fig. 5. Sample images plot from the brain tumor dataset with their class labels.

splitting the data, first, we split data according to patients, so that each patient images were used in either training or validation set but cannot be used in both. This is to make sure that the model result will indicate whether it can classify new patient data that it was not exposed to while training, we call this approach "patient-level approach". In the second approach, we split the images randomly into training and validation sets, this approach we shall refer to as "image-level approach". As a preprocessing step images were cropped at the center to 450×450. The images did not hold any important information at the removed part rather than background-color only.

4.4. Data augmentation

One way to overcome the problem of lack of sufficient amount of training data is data augmentation [48]. Data augmentation is a way of expanding the dataset by performing various manipulations such as lightning, scaling and flipping on the existing data to create new images while preserving the same label. Deep learning models are known to perform better on larger datasets. By augmenting images, the total number of images is increased at runtime allowing the model to train better. Data Augmentation is also considered to be a type of regularization [49] performed on the dataset level, as it allows to reduce overfitting and increases generalization performance by enhancing the training dataset itself which is the root of the problem, without relying on doing any changes to the model architecture. Furthermore, Data Augmentation solves the problem of class imbalance, through over-sampling the minority class it helps to achieve a balanced result on the training data [50]. Medical image datasets are limited in size and hard to collect. Data augmentation has proven to perform powerfully with Convolutional Neural Networks in skin lesion classification, liver lesion classification, brain scan analysis and various medical images problems [50].

In our work, images were augmented during training using various augmentation techniques including horizontal and vertical flips, rotation, shifting, zooming, ZCA whitening, shearing, and brightness manipulation. The classification accuracy results of each of the previous techniques were 97%, 98%, 97%, 98%, 98%, 98%, and 98% respectively. We noticed that each augmentation technique affected the performance of individual classes differently, for example, brightness manipulation got an accuracy of 96% for class one, while applying rotation technique got 98% for the same class. Meanwhile using the same two techniques for class two got 99% for brightness and 98% for rotation. By combining all of the previously mentioned augmentation techniques we have managed to achieve an overall accuracy of 99% which is an improvement of 4% than results without augmentation. The results of applying individual augmentation techniques are displayed in Table 1. We also experimented with random cropping as well but it did not improve the accuracy much.

Table 1
Results of experiments using different augmentation techniques on the dataset in image-level.

Augmentation technique	Class 1	Class 2	Class 3	Balanced accuracy
No augmentation	95%	96%	95%	95%
Horizontal and vertical flips	97%	98%	97%	97%
Rotation	98%	98%	98%	98%
Shifting	97%	98%	98%	97%
Zooming	97%	99%	98%	98%
ZCA whitening	98%	99%	98%	98%
Shearing	97%	99%	98%	98%
Brightness	96%	99%	99%	98%
With augmentation	**98%**	**99%**	**100%**	**99%**

Table 2
Optimized hyperparameters used for training our proposed model.

Hyperparameter	Optimized value
Optimizer	Adam
Number of Epochs	500
Batch size	32
Learning rate	0.00001
Learning rate decay	0

4.5. Model design

For our model, we used ResNet50 architecture which is a Residual Network variant composed of 50 layers. Our model was trained with 500 epochs and using Adam optimizer (adaptive moment estimation) which is a method for stochastic optimization with the learning rate set to 0.00001. Hyperparameter values are displayed in Table 2.

5. Results

Our proposed model was trained using NVIDIA GeForce RTX 2070 GPU with Intel i5-9600K @ 3.7 GHz and 16 GB RAM. The model was implemented on Python 3.6, using Keras 2.2.4 library with Tensor Flow 1.13 as a backend. It took an average of 30 s to complete each epoch. The model results are discussed and evaluated in this section.

5.1. Confusion matrix

A confusion matrix shows the summary of predictions made by the model, where each row represents an actual class and each column represents the predicted class [51]. A normalized confusion matrix is the result of the values divided by the number of elements in each class for a better visual interpretation of which class is being misclassified. Confusion matrix and normalized confusion matrix results are shown in Fig. 6 for image-level and patient-level data distributions respectively.

5.2. Accuracy and loss

The model training and validation accuracy results across epochs are illustrated in Fig. 7, as well as training and validation loss for image-level and patient-level data distributions.

5.3. Accuracy

Accuracy is one of the most commonly used measures for classification models. It represents the number of correct predictions divided by the number of total predictions. For an imbalanced dataset, we may get high accuracy that is biased towards the class with the most instance count. In an extreme case, the classifier might assign every single test case to the large class, thereby achieving an accuracy equal to the fraction of the more frequent labels in the test set. Therefore, accuracy can be a misleading performance measure. A better generalizability measure is the balanced accuracy shown in Eq. (2) where l represents the class count, it can be defined as the average accuracy obtained on either class [52]. We have achieved 99% accuracy and 99% balanced accuracy for image-level, furthermore, we got 97% accuracy and 97% balanced accuracy for patient-level.

$$\frac{\sum_i^l (TP_i + TN_i)(TP_i + FP_i + TN_i + FN_i)}{l} \qquad (2)$$

5.4. Recall

Macro-average is used for computing overall recall by calculating recall for each class separately then getting the average value [53] as

(a) Image-level confusion matrix

(b) Patient-level confusion matrix
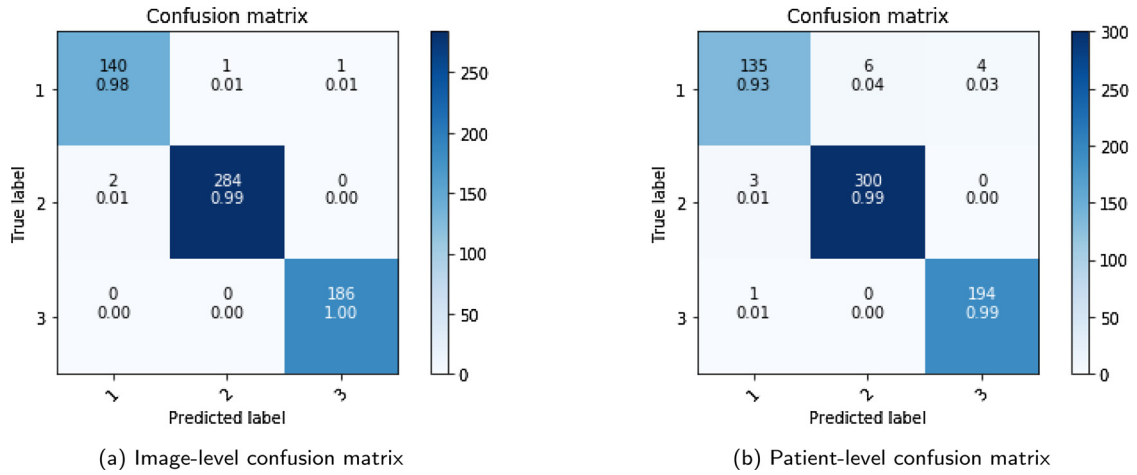
Fig. 6. Results for image-level classification model shown in Fig. 6a. Results for patient-level classification model shown in Fig. 6. In each cell, the number above is the count and the number below is the normalized count.

shown in Eq. (3). We got 99% recall for image-level and 97% recall for patient-level.

$$\frac{\sum_i^l (TP_i)(TP_i + FN_i)}{l} \qquad (3)$$

5.5. Precision

Macro-averaged precision is used as shown in Eq. (4), achieving 99% precision for image-level and 98% precision for patient-level.

$$\frac{\sum_i^l (TP_i)(TP_i + FP_i)}{l} \qquad (4)$$

5.6. F1 Score

Macro-averaged F1 score is used as shown in Eq. (5), achieving 99% f1-score for image-level and 97% f1-score for patient-level.

$$\frac{Precision_M * Recall_M * 2}{Precision_M + Recall_M} \qquad (5)$$

(a) accuracy

(b) loss

(c) accuracy

(d) loss
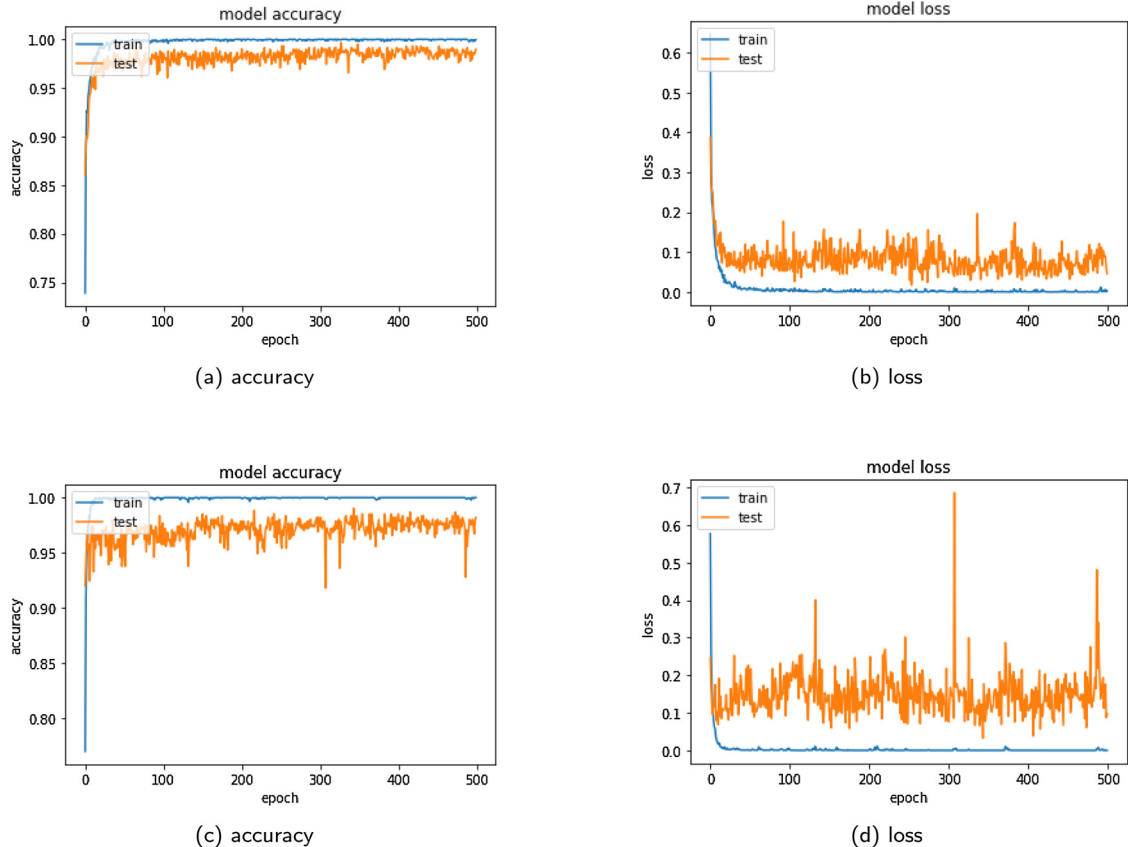
Fig. 7. Results for image-level classification model shown in Fig. 7a and Fig. 7b. Results for patient-level classification model shown in Fig. 7c and 7 d.

Table 3

Previous works for brain tumor classification using the same benchmark dataset [37].

Author	Year	Method	Performance
Cheng et al. [41]	2015	Feature extraction methods: Intensity Histogram GLCM BOW Classification Methods: SVM SRC KNN ROI	Accuracy:91.28%
J.Paul [32]	2016	CNN	Accuracy:84.19%
J.Paul et. al. [42]	2017	CNN	Accuracy:91.43%
Parnian Afshar et al. [35]	2018	CapsNets	Accuracy: 86.56%
Parnian Afshar et al. [34]	2018	CapsNets	Accuracy: 90.89%
N.Abiwinanda et. al. [33]	2018	CNN	Accuracy: 84.19%
A. Pashaei et al. [36]	2018	KELM + CNN	Accuracy: 93.68%
This work	**2019**	**ResNet50**	**Patient-level: Accuracy:97% Precision:98% Recall:97% F1-Score:97% Balanced Accuracy:97% Image-level: Accuracy:99% Precision:99% Recall:99% F1-Score:99% Balanced Accuracy:99%**

GLCM: Gray Level Co-Occurrence Matrix, BOW: Bag-of-Words, SVM: Support Vector Machine, SRC: Sparse Representation-based Classification, KNN: K-Nearest Neighbors, ROI: Region of Interest, CNN: Convolutional Neural Networks, CapsNets: Capsule Networks, KELM: Kernel Extreme Learning Machine.

5.7. Overall results

The overall average classification performance results are shown in Table 3 for both image-level and patient-level.

6. Discussion

In this paper, we have proposed an enhanced deep learning model for brain tumor classification from MRI images. Our model was trained on a benchmark brain tumor MRI images dataset [37] of 3064 MRI images. We have enforced various data augmentation techniques which increased the dataset size and enhanced the accuracy. We have used horizontal and vertical flips, rotation, shifting, zooming, ZCA whitening, shearing, and brightness manipulation as augmentation techniques. Our proposed method used Residual Networks (ResNet), a type of Convolutional Neural Network architecture which won first place at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 [38]. ResNet allows the training of extremely deep neural networks by introducing the concept of shortcut connections to skip one or more layers, therefore solving the problem of degrading accuracy and vanishing gradient which happened when models got deeper. We adapted the ResNet-50 model to our image dataset by modifying the fully-connected layer to have 3 neurons (matching our class count), then we fine-tuned the parameters of ResNet-50 by training it on our dataset.

The output of our classification model was analyzed and evaluated

using several evaluation metrics; such as accuracy, precision, recall, f1-score, and balanced accuracy. Usually, most of the previous studies used accuracy as a metric to evaluate the performance of the model. However, using this measure alone for comparison can be misleading, as it ignores the sensitivity to imbalanced data and the performance of some classes becomes better than others. We have used other measures such as precision, recall, f1-score, and balanced accuracy, so we could have a clear indicator of the generalization of the model over the imbalanced MRI images dataset. Since the later metrics are influenced by class imbalance, they are used to indicate overall model performance irrespective of the individual class count. In addition, two approaches were used while splitting the data, by shuffling and splitting the images randomly into training and validation sets, or what we refer to as an "image-level approach". And the "patient-level approach", meaning that each patient images were used in either training or validation set but not both. The second approach was used as an extra validation step to make sure that our proposed model can classify new patient data that it was not exposed to while training.

Furthermore, we have compared our results with the previous work done using the same benchmark dataset [37] in Table 3. Authors in [41] [42] [33] [35] [34] [36] have achieved accuracies of 91.28%, 91.43%, 84.19%, 86.56%, 90.89% and 93.68% respectively. It is found that our work has achieved state of the art results using the proposed model, achieving a higher accuracy by 5.32% from the previous literature [36] and clearly outperformed all previous work.

In our opinion ResNet architecture has achieved state of the art accuracy in classifying the MRI images dataset. MRI is the most popular non-invasive imaging technique used for detecting tumors; because of its good contrast resolution for the different tissues, and because it does not require radiation. However, tumor classification using MRI images is a challenging task because of the overlapping intensities, inconsistency in size, shape, and orientation. In addition to noise perturbations and low contrast of images [54]. Having a neural network with large enough depth as ResNet is suitable for managing this variety and learning the high-level features. It has proved able to handle the diversity and complexity of the MRI medical images better than the other types of shallower networks. Furthermore, medical datasets are often limited in size and hard to access. The use of data augmentation in our approach has improved the model accuracy by increasing the diversity of the available data without the need to collect more data.

Previous work in [41] used a traditional machine learning approach that required manual feature extraction, which is a time consuming and tedious task. Other previous approaches [42,33,35,34,36] which used Convolutional Neural Network architectures have used shallower networks. Therefore, they had limitations in learning the dataset high-level features and achieving high accuracy. In this work, our choice of ResNet architecture for the brain tumor classification task has proven to handle the MRI images effectively and achieve state of the art results.

For future work, we intend to extend this work by experimenting with larger brain tumor datasets to achieve higher accuracy. We also believe that the proposed model can be used with other types of medical images include radiography (X-ray), ultrasonography (ultrasound), endoscopic, dermoscopic and histology images.

7. Conclusion

In this paper, we have proposed an enhanced deep learning model for brain tumor classification from MRI images. The proposed method used Residual Network (ResNet), a type of deep learning architecture for the model design. We have performed various data augmentation techniques which increased the dataset size and improved the accuracy, such as horizontal and vertical flips, rotation, shifting, zooming, ZCA whitening, shearing, and brightness manipulation. The model was trained on a benchmark brain tumor MRI images dataset [37] of 3064 MRI images and was evaluated under several performance metrics such as precision, recall, f1-score, and balanced accuracy. We have achieved

accuracy of ".99" for image-level, and ".97" accuracy for patient-level using ResNet50 architecture which is state of the art result compared to previous literature. For future work, we intend to extend this work to experiment with larger datasets and more tumor types.

References

[1] Roser M, Ritchie H. Cancer. 2019 [Online]. Available at: https://ourworldinda-ta.org/cancer.

[2] Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 2016;131(6):803–20 [Online]. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27157931.

[3] Nawab ZAR, Swati K, Zhao Q, Kabir M, Ali Z, Ahmed S, et al. Content-based brain tumor retrieval for MR images using transfer learning. IEEE Access 2019;7. 17 809-17 822.

[4] Pereira S, Meier R, Alves V, Reyes M, Carlos A. Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment. iMIMIC - Workshop on Interpretability of Machine Intelligence in Medical Image Computing 2018.

[5] Kumar S, Dabas C, Godara S. Classification of Brain MRI Tumor Images: A Hybrid Approach. Procedia Comput Sci 2017;122:510–7. https://doi.org/10.1016/j.procs.2017.11.400.

[6] Vidyarthi A, Mittal N. Performance analysis of Gabor-Wavelet based features in classification of high grade malignant brain tumors. Proceedings of the 2015 39th National Systems Conference, NSC 2015 2016.

[7] Yuchen Qiu HL, Yan S, Gundreddy RR, Wang Y, Cheng S, Zheng B. A New Approach to Develop Computer-aided Diagnosis Scheme of Breast Mass Classification Using Deep Learning Technology. J Xray Sci Technol 2017;118(24):6072–8.

[8] Machhale K, Nandpuru HB, Kapur V, Kosta L. MRI brain cancer classification using hybrid classifier (SVM-KNN). 2015 International Conference on Industrial Instrumentation and Control. 2015. p. 60–5.

[9] Deepa A, Emmanuel S. MRI Brain Tumor Classification Using Cuckoo Search Support Vector Machines and Particle Swarm Optimization Based Feature Selection. Proceedings of the 2nd International Conference on Trends in Electronics and Informatics. 2018. p. 1213–6.

[10] Menaka Devi T, Ramani G, Xavier Arockiaraj S. MR Brain Tumor Classification and Segmentation Via Wavelets. 2018 International Conference on Wireless Communications. 2018. p. 1–4.

[11] Reema DBA, Mathew A. Tumor detection and classification of MRI brain image using wavelet transform and SVM. Proceedings of IEEE International Conference on Signal Processing and Communication. 2017. p. 75–8.

[12] Polly FP, Shil SK, Hossain MA, Ayman A, Jang YM. Detection and classification of HGG and LGG brain tumor using machine learning. Int J Adv Res Comput Sci 2018;9(2):585–91.

[13] Islam A, Hossain MF, Saha C. A new hybrid approach for brain tumor classification using BWT-KSVM. 4th International Conference on Advances in Electrical Engineering. 2017. p. 241–6.

[14] Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK. Multiclass brain tumor classification using GA-SVM. Proceedings - 4th International Conference on Developments in eSystems Engineering. 2011. p. 182–7.

[15] Abd-Ellah MK, Awad AI, Khalaf AA, Hamed HF. Design and implementation of a computer-aided diagnosis system for brain tumor classification. Proceedings of the International Conference on Microelectronics, ICM 2016:73–6.

[16] Kumar PMS, Chatterjee S. Computer Aided Diagnostic for cancer detection using MRI images of brain. IEEE Annual India Conference (INDICON) 2016.

[17] Vaishnavee KB, Amshakala K. An automated MRI brain image segmentation and tumor detection using SOM-clustering and Proximal Support Vector Machine classifier. ICETECH 2015-2015 IEEE International Conference on Engineering and Technology, no. March 2015.

[18] Sundararaj GK, Balamurugan V. Robust classification of primary brain tumor in Computer Tomography images using K-NN and linear SVM. Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014 2014:1315–9.

[19] Latif G, Butt MM, Khan AH, Butt O, Iskandar DN. Multiclass brain Glioma tumor classification using block-based 3D Wavelet features of MR images. 2017 4th International Conference on Electrical and Electronics Engineering, ICEEE 2017 2017:333–7.

[20] Minz A, Mahobiya C. MR image classification using adaboost for brain tumor type. Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017 2017:701–5.

[21] Chauhan S, More A, Uikey R, Malviya P, Moghe A. Brain tumor detection and classification in MRI images using image and data mining. International Conference on Recent Innovations in Signal Processing and Embedded Systems. 2017. p. 223–31.

[22] Sornam M, Kavitha MS, Shalini R. Segmentation and classification of brain tumor using wavelet and Zernike based features on MRI. 2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016 2016:166–9.

[23] Deepa SN, Devi BA. Neural networks and SMO based classification for brain tumor. Proceedings of the 2011 World Congress on Information and Communication Technologies, WICT 2011 2011:1032–7.

[24] V. O. S. Nicholas G. Polson, Deep learning, Nature Methods, 13, 2018.

[25] Balasooriya NM, Nawarathna RD. A sophisticated convolutional neural network model for brain tumor classification. 2017 IEEE International Conference on Industrial and Information Systems. 2017.

[26] Seetha J, Selvakumar Raja S. Brain tumor classification using convolutional neural network. Biomed Pharmacol. J 2018;68:2018.

[27] Chang P, Grinband BWJ, Bardis MKM, Cadena G, Su M-Y, Cha S, Filippi C, Bota D, Baldi P, Poisson L, Jain R, Chow D. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. Am J Neuroradiol 2018;39(7):2018.

[28] Jude Hemanth D, Anitha J, Naaji A, Geman O, Popescu DE, Hoang Son L. A modified deep convolutional neural network for abnormal brain image classification. IEEE Access 2019;7:4275–83.

[29] Mina Rezaei HY, Hasso CM. Deep Neural Network with l2-Norm Unit for Brain Lesions Detection. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10637 LNCS 2017:798–807.

[30] Yan Xu ML, Jia Zhipeng, Ai Yuqing, Zhang Fang, Chang EI-C. Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation, ICASSP. IEEE International Conference on Acoustics. 2015. p. 947–51.

[31] Yan Liu LN, Stojadinovic Strahinja, Hrycushko Brian, Wardak Zabi, Lau Steven, Lu Weiguo, Yan Yulong, Jiang Steve B, Xin Zhen, Robert Timmerman, Gu X. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. PLoS ONE 2017;12(10):1–17.

[32] Paul J. Deep learning for brain tumor classification. 2016.

[33] Nyoman Abiwinanda STHAH, Hanif Muhammad, Mengko TR. Brain Tumor Classification Using Convolutional Neural Network. Springer Singapore; 2018.

[34] Afshar P, Plataniotis KN, Mohammadi A. Capsule networks for brain tumor classification based on mri images and course tumor boundaries. 2018. arXiv.

[35] Afshara P, Mohammadi A, Plataniotis KN. Brain tumor type classification via capsule networks. 2018. arXiv.

[36] Pashaei A, Sajedi H, Jazayeri N. Brain tumor classification via convolutional neural network and extreme learning machines. 2018 8th International Conference on Computer and Knowledge Engineering (ICCKE) 2018:314–9.

[37] Cheng J. brain tumor dataset 4. 2017 [Online]. Available: https://figshare.com/articles/brain_tumor_dataset/1512427.

[38] Kaiming He SR, Zhang Xiangyu, Sun J. Deep Residual Learning for Image Recognition, Kaiming He, Xiangyu Zhang. Shaoqing Ren and Jian Sun 2015;19(2):107–17.

[39] Sathees Kumar B, Anbu Selvi R. Feature extraction using image mining techniques to identify brain tumors. ICIIECS 2015-2015 IEEE International Conference on Innovations in Information. 2015.

[40] Bangare SL, Pradeepini G, Patil ST. Brain tumor classification using mixed method approach. 2017 International Conference on Information Communication and Embedded Systems. 2017.

[41] Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE 2015;10(10):1–13.

[42] Justin Paul BALS, Andrew Plassard J, Fabbri D. Deep learning for brain tumor classification, Biomedical Applications in Molecular. Structural, and Functional Imaging 2017:2017.

[43] Gorach T. Deep Convolutional Neural Networks - A Review. Int Res J Eng Technol 2018;56(5):1235–50 [Online]. Available: http://www.dbpia.co.kr/Journal/ArticleDetail/NODE07109492.

[44] Bakator M, Radosav D. Deep learning and medical diagnosis: a review of literature. Multimodal Technol. Interact. 2018;2(3):47.

[45] O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. 2015 arXiv, no. November. [Online]. Available: http://arxiv.org/abs/1511.08458.

[46] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9908 LNCS 2016:630–45.

[47] Rosie Dunford ET, Su Q, Wintour A. The Pareto Principle. The Plymouth Student Scientist; 2014.

[48] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW) 2018:117–22.

[49] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. http://www.deeplearningbook.org.

[50] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019.

[51] Sulaiman HMMN. A review on evaluation metrics for data classification evaluations, international. Journal of Data Mining & Knowledge Management Process (IJDKP) 2015;2015:1–11.

[52] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. Proceedings - International Conference on Pattern Recognition 2010:3121–4.

[53] Tharwat A. Classification assessment methods, Applied Computing and Informatics. 2018 [Online]. Available: https://doi.org/10.1016/j.aci.2018.08.003.

[54] Tandel GS, Biswas M, Kakde OG, Tiwari A, Suri HS, Turk M, Laird JR, Asare CK, Ankrah AA, Khanna NN, Madhusudhan BK, Saba L, Suri JS. A review on a deep learning perspective in brain cancer classification. Cancers 2019;11(1).