DOI: 10.15849/IJASCA.220720.04

Int. J. Advance Soft Compu. Appl, Vol. 14, No. 2, July 2022
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Artificial Intelligence Scheme for Medical Images Classification and Prediction

Basil Al-Kasasbeh
Faculty of Computer Studies, Arab Open University, Saudi Arabia
e-mail: bkasasbah@arabou.edu.sa

Abstract

Medicine is the industry where smart technologies and artificial intelligence are most commonly used. Medical imaging is usually used for tumor diagnosis; this includes Computer Tomography and Magnetic Resonance Imaging. Early tumor detection in various organs based on such images is important. This study intended to present an Adaptive Convolution Neural Networks (ACNN) based method for tumor detection in the brain. The ACNN will utilize a modified stochastic gradient descent (MSGD) training algorithm with adaptive momentum and learning rates to speed up the convergence of the error, which will speed up the classification process and improve the accuracy. MSGD is implemented such as when the loss increases, the learning rate increase, and vice versa. The proposed modifications allow the network to increase the learning rate at the beginning of the training process and slow down as the network outcomes reach stabilized conditions. The proposed method results were compared against the performance of several conventional combinations of CNN with several machine learning classifiers. The test results show that the proposed method outperformed the performance of the CNN with all the above-said adaptations. Accordingly, the contributions of this study are (1) improving the ACNN training algorithm for the tumor classification problem and (2) proposing original CNN architectures specialized for tumor classification.

Keywords: artificial intelligence; convolution neural network; medical images processing; tumor classification, tumor detection.

1 Introduction

Generally, cancer is an abnormal growth of cells considered the third prominent cause of death among humans throughout the globe, accounting for approximately 9.96 million deaths and 19.29 million new cases in 2020. Tumors can be classified as benign, noncancerous, and malignant tissues that grow by a splitting process. Brain tumor (BT), according to International Agency for Research on Cancer (IARC), is the 19th most commonly occurring cancer, accounting for 1.6% of global cancer cases and accounting for 2.52% of global deaths in 2020. Thus, BT is the twelfth most dangerous cancer in mortality rate. Moreover, it is also noted from IARC factsheets that approximately 3.7 women and 2.6 men have BT for every hundred thousand. BT is a type of cancer caused by the neoplastic formation of tumor cells due to improper mitosis operation that affects the complex architectural arrangement of cells in the spinal cord or brain region.

The early discovery of malignant tissues is of great importance. The tumor needs to be measured through various characteristics, including the size, shape, color, and distribution [1]. Until today, the biopsy is considered standard among all other primary testing and diagnosis techniques for estimating the grade and stage of the tumor. The biopsy is famous and well-practiced due to its conclusiveness. Still, it is a time-consuming modality as it has to be performed by a lengthy procedure to reduce its aggressiveness. Since the diagnosis through biopsy shows inter-observer variability [2], medical specialists initially prefer magnetic resonance imaging (MRI) and computer tomography (CT) as imaging modalities to determine the architectural structure and stage of the tumor [3]. As the preference of treatment and procedures hinge on the tumor stage, its grade, and pathological type, the medical community and researchers employed various computer vision and image-processing techniques to assist neuro-oncologists. Tumor types and grade diagnosis are usually performed by neurological examination, biomarkers and biopsies, and imaging [4].

Generally, BT is categorized into primary and secondary based on its origin. The primary BT is the tumor originates from the central nervous system (CNS) tissues. In contrast, secondary BT, also known as metastasis, is a cancerous tumor that starts elsewhere in other body parts (such as the kidney, lungs, breast, etc.) but then spreads and migrates to the brain. According to the American Association of Neurological Surgeons (AANS), more than 150 types of BTs are differentiated based on their origin, cell nature, and growth rate. Due to such variations, World Health Organization (WHO) classified BT into four various categories, starting from low (labeled as I) to high (labeled as IV) [5]. This grading is determined by a molecule's morphological properties, such as features and histology [6]. Grade I, II, and III tumor cells are well-, moderately-, and poorly- differentiated, respectively. While grade IV tumor cells are undifferentiated thus, tumor at advanced stages minimizes the life expectancy [7]. A classification of the brain tumor is given in Fig.1.

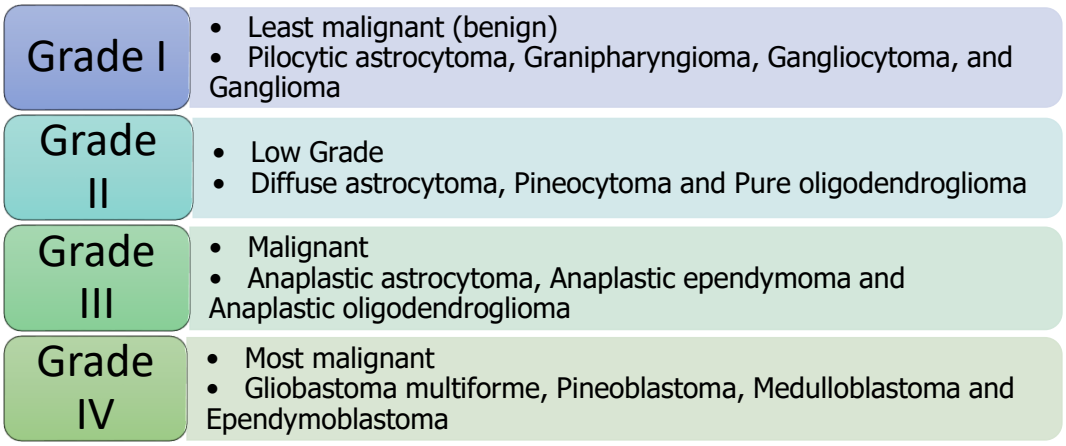

Fig. 1: Brain tumor classification

Brain tumor localization is significant for the treatment process. The treatment depends on the tumor's size, type, and location, which can be identified based on medical imaging, specifically, MRI [7]. Like the brain, various other human organs may suffer from tumors and abnormalities, which depend mainly on diagnosis based on medical imaging. Chest X-ray is one of the most common diagnosis mediums in the medical domain.

Generally, the manual inspection of medical images is hectic and prone to error. Therefore, a computerized diagnostic is required. The performance of computer-aided medical diagnosis (CAMD) has been advanced with the recent advances in artificial intelligence (AI) techniques. Various CAMD applications have proven their applicability and efficiency in various medical domains, such as COVID-19 identification, breast cancer detection [8], tuberculosis diagnosis [9], and assessing the risk of lungs cancer [10]. The artificial neural network (ANN) classifier is used with medical images to classify regions into normal and anomalies. Similarly, several AI-based applications have been developed to assist BT diagnosis using deep learning (DL) and machine learning (ML). These applications use AI for feature selection mechanisms and tumor grading used in a non-invasive approach. While the invasive process of surgeons' intervention is required to collect tumor cell samples from the human body, AI is used to examine the collected samples through histopathological images. Besides, the imaging modalities, including CT and MRI, are utilized in a non-invasive approach to predict tumor grading based on previous knowledge embodied in a trained AI classifier. Besides being a non-invasive technique that uses non-ionizing radiation, MRI generates a sequence of images for a particular cell/tissue region but with varying contrast visualization like T1, T2, and FLAIR. Therefore, MRI is widely adopted for brain tumor classification and grading.

The convolution neural network (CNN) is an ANN used for feature extraction and classification of images in one phase. Many applications have been developed based on ANN and CNN in recent years. However, the problem remains challenging due to the complexity and variation in contrast, texture, and shape of

the resulted MRI images [11-13]. Accordingly, this paper proposes a neural network approach for the primary diagnosis of brain tumors and chest X-rays. A new adaptive hybrid training algorithm provides the novelty of the proposed approach for a convolutional neural network classifier. The proposed system is based on the use of CNN for feature extraction. The features are then used as an input to an adaptive gradient descent machine learning classifier, and the two adaptive hyperparameters are adaptive momentum and learning rate [14-16]. The value is adjusted adaptively according to these adaptive parameters' error variations. The adaptiveness of the proposed method is used to reduce overall loss and increase the convergence speed of the algorithm. The proposed method results are compared against the output of several algorithms, including CNN, with several different classifiers such as Adam, RMSprop, Adadelta, and AdaGrad.

In this study, two different data sets are used to compare the performance of the proposed algorithm; these are: the **Rembrandt** dataset [17], which contains 110,020 MRI images, while the second is **Covid-19 NIH Chest X-ray** [18], the NIH Chest X-ray dataset contains 112120 X-rays. This introduction highlighted the objectives covered in this paper, while the remaining parts of this study are organized as follows: Section 2 presents the related work. Section 3 discusses the proposed method and highlights the mathematical basics behind the proposed solution. Section 4 presents the results and the discussions. Section 5 gives a conclusion.

2 Related Work

Various approaches and frameworks were developed for tumor classification purposes. Al-Saffar and Yildirim [19] automated glioma grading (brain tumor). The framework used supervised and unsupervised techniques to classify input images into normal, high glioma grade (HGG), and low glioma grade (LGG) images. First, multiple eigenvalues selection schemes extract useful features from MRI images using an unsupervised technique. The support vector machine (SVM) and multi-layer perceptron (MLP) are used for classification. The experiments over 1467 MRI images obtained from The Cancer Imaging Archive (TCIA) achieved 91.02% accuracy for glioma grading. Shafi et al. [20] developed an approach for identifying four types of BT using an ensemble network of SVM classifiers and feature extraction and ranking using Shannon entropy. The results of the ensemble SVM using 2399 MRI images were achieved a 97.74%. Marghalani and Arif [21] used a bag of features with SVM to differentiate between BT and Alzheimer's disease and accomplished 97% accuracy on MRI data.

Although the above approaches achieved good accuracy in limited classification tasks, the ML-based model is hard to use in complex domains like BT with huge and diverse datasets. Accordingly, various approaches were developed based on DL and CNN networks. In [22], Tandel et al. presented an ensemble framework of multiple classifiers that performed classification using a majority-voting approach. The network experiments five pre-trained CNN networks (GoogleNet, AlexNet,

ResNet-18, ResNet-50, and VGG-16), which were referred to as DL-based models in comparison with five ML models (SVM, Naïve Bayes, decision tree (DT), k-nearest neighbors (KNN), and linear discrimination). The DL-based majority-voting model outperformed the ML-based model with an improvement of 10.12% accuracy when trained on four data sets. The results of the experiments using Rembrandt obtained an accuracy of 96.51%. Deepak and Ameer [23] developed a transfer-learning (TL) of pre-trained CNN networks in the GoogLeNet technique to differentiate three types of BT (glioma, pituitary, and meningioma) using an MRI dataset from figshare. The developed DL-based framework was compared with SVM and KNN for classification. The results revealed that KNN with deep CNN features outperformed the other classifiers by attaining an accuracy of 98%. Furthermore, a model with limited data was developed and obtained 97.1% accuracy with SVM when trained with 56% of the dataset. Pashaei et al. [24] proposed a classification framework consisting of a convolutional neural network (CNN) for feature extraction and a kernel extreme learning machine (KELM) for the classification of three types of BT using T1-weighted contrast-enhanced MRI images. The results of the proposed classification over a dataset of 3064 images achieved an accuracy of 93.68%.

Afshar et al. [25] used a modified CNN model called capsule network (CapsNet) to capture the tumor's spatial relation and surroundings for accurate tumor classification. The problem is associated with the pooling layers in the CNN, where pooling is only implemented if it will improve the results. These processes are leaned in the learning process and associated with edges between the various pooling layers in the network. The experiments of the proposed network resulted in an accuracy of 90.89%. Kabir Anaraki et al. [26] suggested a different approach for grading Glioma (BT) and differentiating it from Pituitary and Meningioma BT. The classification was built based on bagging CNN networks classification combined with a genetic algorithm (GA). The experiments over the IXI dataset, which contains 600 MRI images, resulted in an accuracy of 90.9% for estimating Glioma grade and 94.2% accuracy for tumor classifying.

Similarly, Jude and Anitha [27] proposed a modified GA framework for generating offspring with minimal randomness compared to the conventional GA. The proposed modification on GA is used to enhance the backpropagation learning of the ANN. The Gray Level Difference Method (GLDM) is used for feature extraction, and the GA is used for feature selection, which is fed into the ANN. The modified GA combination with the ANN framework obtained an accuracy of 98% on a customized dataset of 330 images. Khairandish et al. [28] implemented a hybrid model composed of CNN and SVM with a threshold segmentation approach. The hybrid model obtained an overall accuracy of 98.49% while classifying BT into malignant and benign tumors based on MRI. The experiments were conducted on the BRATS dataset of 330 cases. Ghassemi et al. [29] proposed a generative adversarial network (GAN) based on deep ANN to classify 3064 T1-CE MRI into three different types of BT, which yield an overall accuracy of 95.6%.

Yang et al. [30] utilized pre-trained GoogleNet and AlexNet to grade glioma (type of BT) using MRI based on deep ANN. Based on the experimental results on a dataset of 750 images, it was concluded that GoogleNet performed better than AlexNet by securing an average test accuracy of 90.9% over fivefold cross-validation. Swati et al. [31] developed a content-based image retrieval (CBIR) tool based on pre-trained VGG-19 as a feature extraction framework for a brain tumor. They adopted transfer learning and achieved a precision of 96.13% by testing the proposed model using fivefold cross-validation on the CE-MRI data set of 3064 images. Similarly, Talo et al. [32] fine-tuned the ResNet-34 model by modifying the dense layers to classify brain abnormality using 613 MRIs. They further enhanced the model performance to 100% accuracy by using a data augmentation approach.

Mehrotra et al. [33] designed a transfer learning-based DL framework to segregate malignant tumors from benign tumors using 696 T1-weighted MRI. In addition to pre-trained networks (AlexNet, SqueezNet, GoogleNet), a CNN architecture is arranged to accomplish 99.04% accuracy. Likewise, Srikanth and Suryanarayana [34] employed VGG-16 and achieve an accuracy of 98% on multi-class BT classification. Harish and Baskar [35] suggested a modified version of the Faster Region-based CNN model combined with ResNet-50 to segregate the tumor region from rest and classify BT (MRI) into malignant and benign using the AlexNet framework. Sajjad et al. [36] proposed a multi-grade BT diagnostic tool that first segments tumor regions using the DL approach. They then fine-tuned pre-trained VGG-19 with augmented data to attain overall accuracy of 90.67%. Ismael and Abdel-Qader [37] suggested a BT classification system by first practicing the Gabor filter and discrete wavelet transform (DWT) for statistical feature extraction. Later, they split the T1-weighted MRI data set into training-validation sets to train and test a multi-layer perceptron (MLP) classifier for a three-class classification problem. The model achieved an overall accuracy of 91.9% on three types of brain tumors (Glioma, Pituitary, and Meningioma). Similarly, Sabitha et al. [38] used DWT to obtain features and principal component analysis (PCA) for dimensionality reduction. Finally, kernel SVM attained an accuracy of above 90% while classifying the MRI images as benign, normal, or malignant. Table 1 summarizes the related work reviewed in this section. An accurate BT classification has been obtained using CNN with different deep ANN classifiers. Although such an approach achieved better results compared to the classical ML classification algorithms, the problem with such approach is the time-consumption.

Similar to the work reported in the literature, which suggested using CNN for feature extraction combined with a robust classification technique, such as deep ANN, the proposed framework combines CNN with adaptive deep ANN. The proposed method isolates the feature extraction process from the classification process to ease the complexity of the complex structure of the CNN and the ANN. The proposed method aims to enhance classification accuracy. Yet, the proposed framework also aims to improve the convergence speed. Accordingly, enable fast

classification of huge database. Yet, unlike the previous work, the proposed framework is built by modifying the learning technique.

Table 1: Summary of the Related Work on BT Classification

Ref.	Task	Features	Technique	Dataset	Results
Al-Saffar and Yildirim [19]	Glimo grading (normal vs. HGG vs. LGG	Eigenvalues selection	SVM and MLP	TCIA	91.02%
Shafi et al. [20]	BT classification	Shannon Entropy	Ensemble of SVM	Customized	97.74%
Marghalani and Arif [21]	BT vs. Alzheimer classification	Bag of features	SVM	Customized	97%
Tandel et al. [22]	BT classification	CNN	Majority Voting	Rembrandt	96.51%
Deepak and Ameer [23]	BT classification	CNN	DL-transfer learning	Figshare	98%
Pashaei et al. [24]	BT classification	CNN	ELM	Customized	93.68%
Afshar et al., [25]	BT classification	CapsNet	CapsNet	Customized	90.89%
Kabir Anaraki et al. [26]	BT classification and Grading	CNN	GA	IXI	94.2%
Jude and Anitha [27]	BT classification	GLDM and GA for selection	ANN	Customized	98%
Khairandish et al. [28]	BT classification	CNN	SVM	BRATS	98.49%
Ghassemi et al. [29]	BT classification	GAN	Deep ANN	Customized	95.6%
Yang et al. [30]	BT classification	CNN	GoogleNet and AlexNet	Customized	90.9%
Swati et al. [31]	BT classification	CNN	VGG-19	CE-MRI	96.13%
Talo et al. [32]	BT classification	CNN	ResNet34	Customized	100%
Mehrotra et al. [33]	BT classification	CNN	AlexNet, GoogLeNet, ResNet50, ResNet101, SqueezeNet	TCIA	99.04%
Srikanth and Suryanarayana [34]	BT classification	CNN	VGG-16	Customized	98%
Harish and Baskar [35]	BT classification	R-CNN	ResNet50	Customized	99.25%
Sajjad et al. [36]	BT classification	CNN	VGG-19	Customized	90.67%
Ismael and Abdel-Qader [37]	BT classification	Gabor filter and DWT	ANN	Customized	91.9%
Sabitha et al. [38]	BT classification	DWT	SVM	Customized	90%

3 The Proposed Work

The structure of the proposed framework, as given in Fig. 2, consists of three sages. The first stage processes the input images in the dataset by applying preprocessing steps: image cropping, resizing, and normalization. The second stage implements the CNN process for feature selection. In the third stage, the selected deep features are used as an input to the adaptive deep learning classifier to predict the correct classification of the input images.

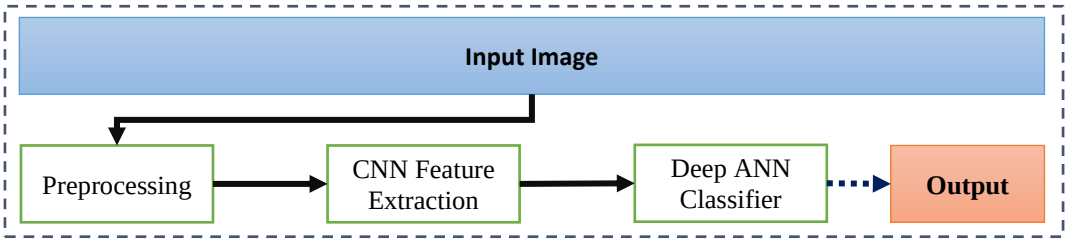

Fig. 2: The structure of the proposed solution for BT classification

3.1 Data Preprocessing

Commonly, medical imaging is noisy, unbalanced, and has different sizing. Accordingly, these images require some preprocessing to achieve better results. Generally, the CNN stage, which followed the preprocessing stage, produced more accurate results with balanced and big datasets. Yet, because the available datasets are not considered big, the preprocessing implements an augmentation of the dataset. Accordingly, noise data is added to the dataset as adding noise to inputs of the ANN leads to significant improvement in generalizing the dataset.

Moreover, adding noise acts as some augmentation of the dataset and balances the datasets. Three augmentation techniques were used, Random Horizontal Flip, Random Resized Crop (to get deeper relation among pixels), and finally, augmenting images with varying intensity. Augmentation also leads to avoiding overfitting of the ANN. After the dataset is augmented, the images are normalized, resized, and converted into unified stream code. The images are normalized as given in Equation 1.

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where y_i is the normalized intensity value for the pixel at the position x_i, $i = \overline{1,N}$. The min(x) and max (x) values represent the underlying input image's minimum and maximum intensity values. Once the normalization process is completed, the images are then resized to 64 x 64. A cropping technique was used for all images. To crop the part that contains only the brain of the image, we used a contour detection algorithm [18] to find the extreme points of the brain image. Finally, the data was converted into a byte-stream structure hierarchy.

3.2 Feature Extraction using CNN

As a deep network, the Convolution Neural Network (CNN) uses multiple layers to extract the spatial and temporal features in the input image. The constructed CNN consists of five convolution layers and five max-pooling layers in the proposed framework, and each follows each convolution layer for size reduction. The convolution layers apply convolution filters to the input for extracting the spatial and temporal features. The max-pooling layers are subsampling layers that reduce the dimensionality of the input. Finally, a fully connected layer is used to extract the final features. Specifically, the proposed CNN uses five convolution layers with filter sets of 64,128, 256, 512, and 1024 respectively, with kernels of size 3x3 with padding. A drop out of 0.3 with batch normalization was adopted throughout the layers. The ReLU activation function is used in all layers. The convolution layers were followed by a flattened layer and four dense layers. Fig. 3 illustrates the developed CNN.

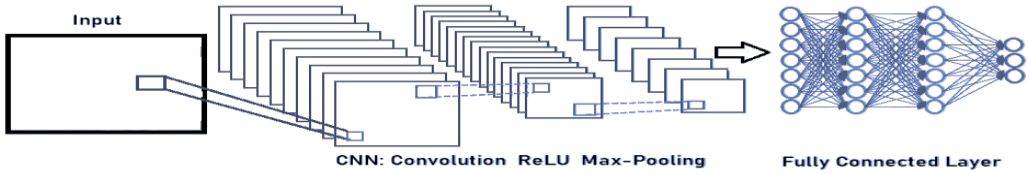

Fig. 3: Proposed CNN for feature extraction

3.3 Classification using Deep Learning

In the classification step, the feature extracted in the previous step is used as an input to a deep ANN classifier of fully connected layers with customized stochastic gradient descent (SGD) training algorithm developed based on adaptive learning and momentum terms. The adaptive terms are introduced to reduce the loss and increase the convergence of the network. The network consists of five fully connected layers with 64,128,256, 512 neurons with an output classification layer that depends on the number of output classes with the ReLU activation function, as illustrated in Fig. 4. A dropout and batch normalization is used. The cross-entropy loss function is employed in this study to calculate the error between predicted and true values and perform the training/adjustment of the trained model iteratively—the lower the loss, the better the model. For binary classification, the binary cross-entropy is defined as given in Equation 2.

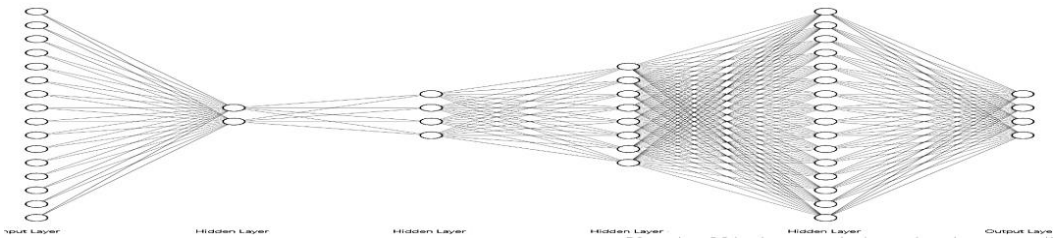

Fig. 4: Proposed ANN with five fully connected layer

$$J_i = -[t\ log\big(P(t)\big)\ log\big(1 - P(t)\big)] \quad (2)$$

where t is the actual label, $P(t)$ is the predicted label, and $i = 1, \dots, N$.

In the learning process, both the adaptive momentum and learning rates used in the proposed ANN are directly proportional to the error, meaning that the momentum and learning rate values will change according to changes in the error value. Accordingly, as the backpropagation (BP) algorithm provides an approximate trajectory in the weight and bias space, a modified SGD (MSGD) is used to compute the trajectories. In conventional SGD, it is required to use a small learning rate in the update weight equations of the SGD. The small learning rate leads to a slow update process; however, to speed up the learning process, the momentum term is used to supplement the learning process by integrating the momentum with weight update equations. Using momentum will speed up the convergence process and smooth the bias and weight update. The modified equations for updating the weights are calculated in Equation 3 and Equation 4.

$$\Delta w_{i,j}(z) = \mu \Delta w_{i,j}(z-1) + \sigma \frac{\partial E(w,b)}{\partial b_i^l} \quad (3)$$

$$w_{i,j}^l(z) = w_{i,j}^l(z-1) - \Delta w_{i,j}(z) \quad (4)$$

where μ is the momentum term or constant, such as $0 < \mu < 1$. The values of the $w_{i,j}(z)$ and $w_{i,j}(z\text{-}1)$ are the weight of the edge i,j at the z iteration and $z\text{-}1$ iteration, respectively. Accordingly, $w_{i,j}^l(z)$ and $w_{i,j}^l(z\text{-}1)$ are the weights at the l epoch. The value Δw represents the change in the weight. The parameter σ represents the learning rate, such as $0 < \sigma < 1$. The term ∂E(w,b) is the variation of the loss and bias w.r.t. ∂ b. Accordingly, as the loss increase, the learning rate increase and vice versa. This will allow the network to increase the learning rate at the beginning of the training process and slow down as the network outcomes reach stabilized conditions.

The momentum term used in the weight update, as given in Equation 2, has a great impact on the adaptions of the network; thus, it should be adaptive to the loss in the network. The adaptiveness of the term will, as mentioned, speed up the convergence by adjusting the momentum term according to the epoch current loss value. The variable adaptive momentum proposed in this study is given in Equation 5.

$$\mu(i) = \frac{\beta}{1 + 0.5e^{-|\sqrt{E_i E_{i-1}}|}} \quad (5)$$

where β is the forgetting factor, $0 < \beta < 1$, e is the positive number, and E_i and $E_{i\text{-}1}$ are the current and previous epochs errors, respectively. Besides, the learning rate σ is adaptively calculated as similar to the calculation of the adaptive momentum in equation (5). Accordingly, the two adaptive terms will be correlated, which will lead to a more effective, fast, and stable convergence of the BP algorithm. The proposed variable adaptive learning rate is calculated as given in Equation 6.

$$\sigma(i) = \frac{\beta}{\left(1+e^{-\frac{1}{E_i+E_{i-1}}}\right)} \tag{6}$$

where e is the learning constant, a positive number $0 < e \ll 1$, and E_i and E_{i-1} are the current and previous epochs error, respectively.

The SGD weight updating equation (3) includes both adaptive variable terms. The equation is given in Equation 7.

$$\Delta w_{i,j}(z) = \frac{\beta}{1+0.5e^{-\left|\sqrt{E_i E_{i-1}}\right|}}\Delta w_{i,j}(z-1) + \frac{\beta}{\left(1+e^{-\frac{1}{E_i+E_{i-1}}}\right)}\frac{\partial E(w,b)}{\partial b_i^l} \tag{7}$$

4 Experiments

Besides the proposed adaptive method, several training algorithms are used as a benchmark for the proposed one. The benchmark training algorithms are Adam, RMSprop, Adadelta, and AdaGrad. The benchmark training algorithms are chosen because they are all adaptive training algorithms. The adaptiveness of the above benchmark training algorithm is different from the proposed method.

4.1 Datasets

The datasets used for the curried research experimental are the REMBRANDT dataset created by Justin Kirby and Modified by Quasar Jarosz on Jun 03, 2020 [17]. The NIH X-Ray [18] was created by the National Institutes of Health datasets, as illustrated in Fig. 5. REMBRANDT provided the largest BT imaging dataset that contains 110,020 MRI images for 130 different patients. The dataset is generated from 874 glioma specimens comprising 566 gene expression arrays, 834 copy number arrays, and 13,472 clinical phenotype data points. NIH X-Ray contains 112,120 X-ray disease images from 30,805 different patients.

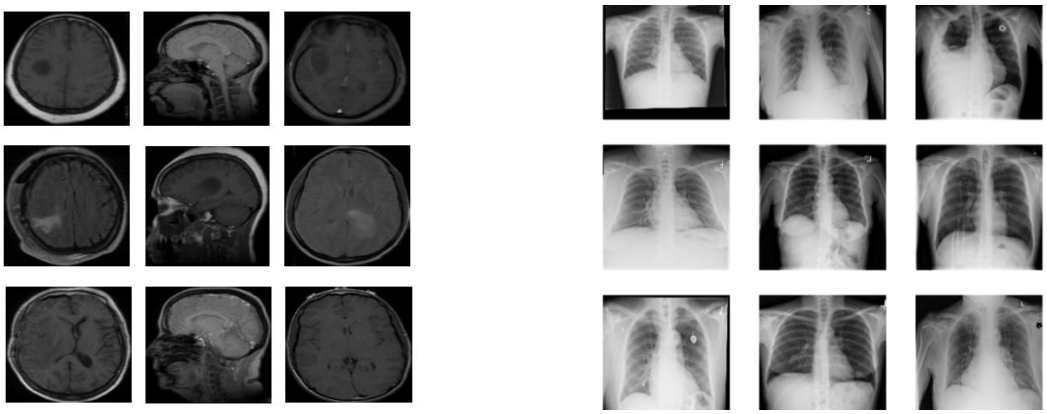

Fig. 5: Sample images from REMBRANDT (left) and NIH X-Ray (right) datasets

The dataset is divided into two parts for training and validation. The CNN model was trained using 70% of the dataset images, followed by a validation and testing process using the remaining 30% of the data set. The transfer learning method was not employed in our study since the primary concern of this study is to investigate the effect of using variable adaptive learning and momentum terms to improve both the converging speed and the accuracy. The same test used well-known adaptive training algorithms (Adadelta, RMSprop, Adam, and Adagrad). Fig. 4 illustrates the proposed algorithm results and the results of the other known training algorithms.

4.2 Results

The total number of training parameters of the proposed method is twenty-three million, three hundred ninety-three thousand, and ninety-two (23,393,092). The proposed and compared methods were tested, and these methods' convergence was reported. While the compared method achieved good accuracy in various domains, the noise nature of the medical imaging leads to a drop-down in the results. Moreover, the convergence rate of these methods is relatively slow. As the concern of the proposed method is the convergence rate, the convergence rate of the proposed and compared methods for REMBRANDT and NIH datasets are reported as given in Fig 6 and Fig. 7, respectively.

The proposed algorithm and the benchmark training algorithms are given in Table 2. The results are reported for the accuracy of both the training and validation process.

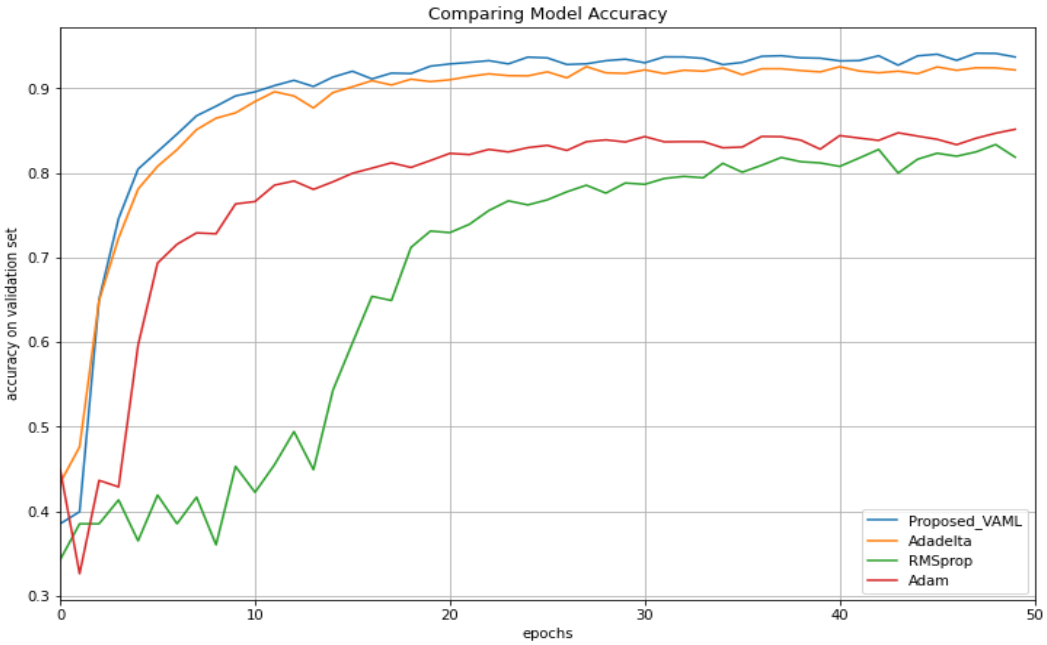

Fig. 6: The proposed method accuracy plot for the REMBRANDT dataset

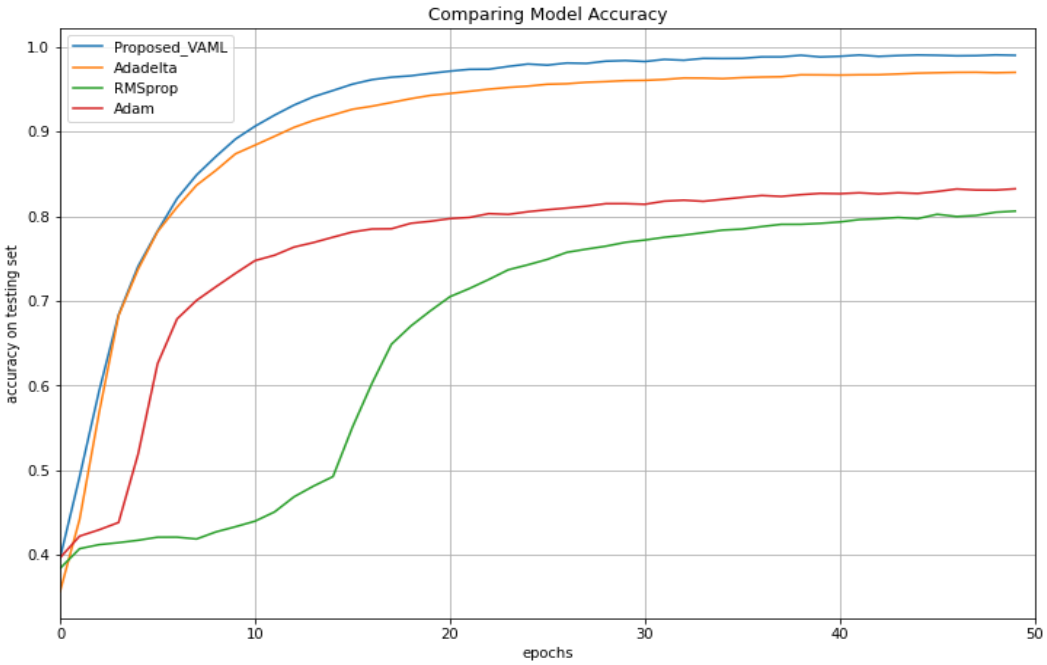

Fig. 7: The proposed method accuracy plot for the NIH Chest X-ray dataset

Table 1: The Accuracy results of all training algorithms for both datasets.

Dataset	**Accuracy %**							
	Proposed		**Adam**		**AdaGrad**		**RMSprop**	
	Test	*Vali.*	*Test*	*Vali.*	*Test*	*Vali.*	*Test*	*Vali.*
REMBRANDT	**99.04**	**94.23**	83.23	85.17	97.00	92.55	80.60	83.70
NIH Chest X-ray	**98.87**	**94.18**	82.90	85.69	97.45	93.70	81.07	82.30

The proposed method outperformed the other training algorithm in testing accuracy and validation accuracy, according to the reported results. The accuracy for both the testing and validation is 0.99 and 0.94, respectively. The performance of the Adam training algorithm is the best among the other training algorithms. Besides, the results show that the Adam training algorithm performed better than the proposed method up to the third epoch; the proposed method outperformed the Adam training algorithm after the fourth epoch. This can be referred to as the adaptive nature of the proposed method. As mentioned, the learning rate varies based on the learning stage, which in this case prof the fast adaptation of the proposed method after the 3rd epoch. The use of the adaptive terms in the SGD classifier provides the necessary improvement in terms of accuracy and fast convergence, as expected.

5 Conclusions

BT is the most dangerous cancer type, destabilizing the nervous system's normal behavior. BT imposes pressure within the skull, which in turn affects brain

functionality. In this study, a novel approach was developed based on variable adaptive momentum and learning rate to recognize and classify the different types of BT. The results of the proposed work were compared against similar well-known adaptive algorithms, which revealed that the proposed method outperformed the well-known algorithms during both the testing and validation. The timing required for the testing and validating of the proposed method was much faster than the other training algorithms needed. Future work intends to test the proposed algorithm against different datasets to confirm that the proposed algorithm is not data-dependent and performs well against all kinds of data. Further research is planned to be directed to an in-depth mathematical analysis of traditional problems of teaching neural networks: overcoming local minima, saddle points, and "ravines".

ACKNOWLEDGEMENT
The author would like to thank Arab Open University-Saudi Arabia for supporting this study.

References

[1] R. Kumar, R. Srivastava and S. Srivastava, "Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features," *Journal of Medical Engineering*, vol. 1, no. 1, pp. 1–14, 2015.

[2] K. Komaki, N. Sano and A. Tangoku, "Problems in histological grading of malignancy and its clinical significance in patients with operable Breast Cancer," *Breast Cancer*, vol. 13, no. 3, pp. 249–253, 2006.

[3] S. Bauer, R. Wiest, L.-P. Nolte and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine and Biology*, vol. 58, no. 13, pp. 97–129, 2013.

[4] A. Kotrotsou, P. O. Zinn and R. R. Colen, "Radiomics in brain tumors: an emerging technique for characterization of tumor environment," *Magnetic Resonance Imaging Clinics of North America*, vol. 24, no. 4, pp. 719–729, 2016.

[5] D. Louis, A. Perry, G. Reifenberger, A.Deimling, D. Figarella-Branger *et al.*, "The 2016 World health organization classification of tumors of the central nervous system: a summary,"*Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016.

[6] G. Tandel, M. Biswas, O. Kakde, A. Tiwari , H. S Suri *et al.*, "A Review on a deep learning perspective in brain cancer classification," *"Cancers*, vol. 11, no. 1, pp. 1-32, 2019.

[7] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain tumor segmentation using convolutional neural network in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240-1251, 2018.

[8] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang et al.," Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020.

[9] L. T. Duong, N. H. Le, T. B. Tran, V. M. Ngo and P. T. Nguyen, "Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning," *Expert Systems with Applications*, vol. 184, no. 1, pp. 1-48, 2021.

[10] S. Mishra, H. Thakkar, P. Mallick, P. Tiwari, A. Alamri," A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection," *Sustainable Cities and Society*, vol. 72, no. 1, pp. 1-14, 2021.

[11] Adnan A. Hnaif, Abdelfatah A. Tamimi, Ayman M. Abdalla and Iqbal Jebril, A Fault-Handling Method for the Hamiltonian Cycle in the Hypercube Topology, Computers, Materials & Continua, SCI (IF 4.89), Vol. 68, No. 1 pp. 505-519, 2021.

[12] Adnan A. Hnaif, Khalid Mohammad Jaber, Mohammad A. Alia, and Mohammed E. Daghbosheh, "Parallel Scalable Approximate Matching Algorithm for Network Intrusion Detection System", The International Arab Journal of Information Technology (IAJIT), ISI (IF 0.467), Vol. 18, No.1, pp. 77-84 https://doi.org/10.34028/iajit/18/1/9, 2020.

[13] Ahmad AA Alkhatib, Adnan A. Hnaif, and Tarek Kanan, " Proposed simple system for Road Traffic Counting", International Journal of Sensors, Wireless Communications and Control, 2018 Vol. 8(3).

[14] Haya Al-Masalha, Adnan A. Hnaif and Tarek Kanan, Cyber-Crime Effect on Jordanian Society, International Journal of Advances in Soft Computing and its Application (IJASCA), Vol. 12, No. 3, pp. 123-139, 2020.

[15] Ahmad A. A. Alkhatib, Mohammad Alia, Adnan Hnaif, " Smart System for Forest Fire Using Sensor Network", International Journal of Security and Its Applications, Vol. 11, No. 7 (2017), pp.1-16, 2017

[16] Ahmad A. A. Alkhatib, Mohammad Alia, Adnan Hnaif and Sufian Yousef, A novel method for localising a randomly distributed wireless sensor network, International Journal of Systems Assurance Engineering and Management, Vol. 9, No. 2, pp. 354–361, 2018.

[17] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-57, 2013.

[18] X. Wang, Y. Peng, L. Lu, Z. Lu, Mohammadhadi Bagheri *et.al.*, "ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *in Proc. IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA, pp. 2097-2106, 2017.

[19] Z. Al-Saffar and T. Yildirim, "A hybrid approach based on multiple Eigenvalues selection (MES) for the automated grading of a brain tumor using MRI," *Computer Methods and Programs in Biomedicine,* vol. 201, no. 1, pp. 1-10, 2021.

[20] A. Shafi, M. Rahman, T. Anwar, R. S. Halder and H. M. Emrul Kays, "Classification of brain tumors and auto-immune disease using ensemble learning," *Informatics in Medicine Unlocked,* vol. 24, no. 1, pp. 1-8, 2021.

[21] B. Marghalani and M. Arif, "Automatic Classification of Brain Tumor and Alzheimer's Disease in MRI," *Procedia Computer Science,* vol. 163, no. 1, pp. 78-84, 2019.

[22] G. Tandel, A. Tiwari and O. Kakde, "Performance optimisation of deep learning models using majority voting algorithm for brain tumor classification," *Computers in Biology and Medicine,* vol. 135, no. 1, pp. 1-10, 2021.

[23] S. Deepak and P. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine,* vol. 111, no. 1, pp. 1-10, 2019.

[24] A. Pashaei, H. Sajedi and N. Jazayeri, "Brain tumor classification via convolutional neural network and extreme learning machines," *in Proc. 8th International Conference on Computer and Knowledge Engineering (ICCKE),* Mashhad, Iran, pp. 314–319, 2018.

[25] P. Afshar, K. Plataniotis and A. Mohammadi, "Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Brighton, United Kingdom, pp. 1368–1372, 2019.

[26] A. Kabir, M. Ayati and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 63–74, 2019.

[27] D. Jude Hemanth and J. Anitha, "Modified genetic algorithm approaches for classification of abnormal magnetic resonance brain tumor images," *Applied Soft Computing*, vol. 75, no. 1, pp. 21-28, 2019.

[28] M. Khairandish, M. Sharma, V. Jain, J.M. Chatterjee and NZ. Jhanjh, "A Hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of mri brain images," *IRBM,* vol. 1, no. 1, pp.1-10, 2021.

[29] N. Khassemi, A. Shoeibi, and M. Rouhani, "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images," *Biomedical Signal Processing and Control,* vol. 57, no. 1, pp.1-8, 2020.

[30] Y. Yang, L.-F. Yan, X. Zhang, Y. Han1 and HY Nan, "Glioma grading on conventional MR images: a deep learning study with transfer learning, frontiers," *Neuroscience*, vol. 12, no. 1, pp.1-10, 2018.

[31] Z. Swati, Q. Zhao, M. Kabir, F. Ali and Z. Ali, "Content-Based brain tumor retrieval for mr images using transfer learning," *IEEE Access*, vol. 7, no. 1, pp. 17809-17822, 2019.

[32] M. Talo, U. Baloglu, O. Yıldırım and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cognitive Systems Research*, vol. 54, no. 1, pp. 176–188, 2019.

[33] R. Mehrotra, M. Ansari, R. Agrawal and R. Anand, "A Transfer learning approach for ai-based classification of brain tumors," *Machine Learning with Applications*, vol. 2, no. 1, pp. 1-14, 2020.

[34] B. Srikanth and S. V. Suryanarayana, "Multi-Class classification of brain tumor images using data augmentation with deep neural network," *Materials Today: Proceedings*, in press, 2021.

[35] P. Harish and S. Baskar, " MRI based detection and classification of brain tumor using enhanced faster R-CNN and Alex Net model," *Materials Today: Proceedings*, in press, 2020.

[36] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah *et al.*, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, no. 1, pp. 174–182, 2019.

[37] M. Ismael and I. Abdel-Qader, "Brain Tumor classification via statistical features and backpropagation neural network," *in Proc. IEEE International Conference on Electro &Information Technology (EIT)*, Rochester, MI, USA, pp. 252–257, 2018.

[38] V. Sabitha, J. Nayak and P. Reddy, "MRI brain tumor detection and classification using KPCA and KSVM," *Materials Today: Proceedings*, in press, 2021.

Notes on contributors

Basil Al Kasasbeh is an associate professor at the Faculty of Computer Studies, Arab Open University, Saudi Arabia. His research interests include Computer Networks, Mobile system, AI and Cyber security.