IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 25, 2019, accepted May 16, 2019, date of publication May 27, 2019, date of current version June 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919122

Multi-Classification of Brain Tumor Images Using Deep Neural Network

HOSSAM H. SULTAN, NANCY M. SALEM, AND WALID AL-ATABANY
Department of Biomedical Engineering, Faculty of Engineering, Helwan University, Cairo 11792, Egypt

Corresponding author: Hossam H. Sultan (hossam.sultan@h-eng.helwan.edu.eg)

ABSTRACT Brain tumor classification is a crucial task to evaluate the tumors and make a treatment decision according to their classes. There are many imaging techniques used to detect brain tumors. However, MRI is commonly used due to its superior image quality and the fact of relying on no ionizing radiation. Deep learning (DL) is a subfield of machine learning and recently showed a remarkable performance, especially in classification and segmentation problems. In this paper, a DL model based on a convolutional neural network is proposed to classify different brain tumor types using two publicly available datasets. The former one classifies tumors into (meningioma, glioma, and pituitary tumor). The other one differentiates between the three glioma grades (Grade II, Grade III, and Grade IV). The datasets include 233 and 73 patients with a total of 3064 and 516 images on T1-weighted contrast-enhanced images for the first and second datasets, respectively. The proposed network structure achieves a significant performance with the best overall accuracy of 96.13% and 98.7%, respectively, for the two studies. The results indicate the ability of the model for brain tumor multi-classification purposes.

INDEX TERMS Brain tumor, convolutional neural network, data augmentation, deep learning, MRI.

I. INTRODUCTION

Brain tumor can be defined as unnatural and uncontrolled growth in brain cells. Since the human skull is a rigid and volume limited body, consequently, any unexpected growth may affect a human function according to the involved part of the brain; moreover, it may spread into other body organs and affect human functions [1]. According to the world cancer report published by the World Health Organization (WHO), brain cancer accounts for less than 2% of human cancer; however, severe morbidity and complications are produced [2]. Cancer research corporation in the United Kingdom mentioned that there are about 5,250 deaths annually by the act of brain, other Central Nervous System (CNS) and intracranial tumors in the UK [3].

Brain tumors can be classified in many ways, for instance, primary and secondary tumors. The former represents about 70% of all brain tumors, while secondary tumors are the residuals 30%. This classification is determined according to tumors origin just as tumors first originate in the brain are called primary tumors. On the other side, tumors first arise in any other part of the body and then transferred to the brain are called secondary tumors, and most of them are malignant [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

Numerous imaging techniques can be used to detect and classify brain tumors. However, MRI is one of the most common non-invasive techniques. MRI popularity comes from the fact of using no ionizing radiation during the scan as well as its superior soft-tissue resolution plus the ability to acquire different images using various imaging parameters or by employing contrast-enhanced agents [5], [6].

Gliomas are the most prevalent type of brain tumors that originate in the glial cells of the brain [7]. Gliomas include 30% of all brain tumors and CNS, and 80% of all malignant brain tumors [7]. Gliomas classified into four grades according to the WHO starting from type I to IV [8]. Grade I tumors are benign and have a much similar texture of the normal glial cells, Grade II is a slightly different in texture, Grade III tumors are malignant with abnormal tissue appearance while grade IV is the most severe stage of gliomas and tissue abnormalities that can be visualized by naked eye [1]. Meningioma is a tumor that forms on the membrane that covers the brain and spinal cord inside the human skull and grows placidly. Most of meningioma tumors are benign [8]. However, pituitary tumor starts from the pituitary glands that control hormones and regulate functions in the body. It can be benign, benign that expands to bones, and malignant. Complications of pituitary tumors may cause permanent hormone deficiency and vision loss [1].

By cause of the information mentioned above, early detection and classification of brain tumors turn into a vital task in case assessment and accordingly help in selecting the most convenient treatment method to save patients life [8]. Furthermore, the classification stage may be a confusing and tedious task for physicians or radiologists in some complicated cases. These cases need experts to work on, localize the tumor, compare tumor tissues with adjacent regions, apply filters on the image if necessary; to make it more clear for human vision, and finally conclude; whether it is a tumor besides its type and grade if available. This task relatively consumes time, and that's why there is a need for a Computer Aided Diagnosis (CAD) system to early detect brain tumors in much less time without human intervention.

Machine Learning (ML) is the study of algorithms and statistical models that can be used to perform a specific task without using outright instructions, relying on patterns instead of that [9]. ML algorithms have been widely emerged in the medical imaging field as a part of artificial intelligence [6]. It can be divided into two main categories, supervised and unsupervised. In supervised techniques, an algorithm is used to find a mapping function of input variables and their related output labels to predict new subjects labels. The primary goal is to learn inherent patterns within the training data using algorithms such as Artificial Neural Network (ANN) [10], Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) [11]. In contrast, unsupervised learning is based only on the input variables as in fuzzy c-means [12] and Self-Organization Map (SOM) [13]. There is a must to extract features of the training images that are usually grayscale, texture and statistical features to establish learning and that perhaps require segmenting the tumor in most cases before features extraction stage. These features are called handcrafted features in which an expert who has a strong knowledge and the ability to figure the most meaningful features is needed. Moreover, this job consumes much time and is prone to error while handling large scale of data [14].

Deep Learning (DL) is a subdivision of ML that is based on learning data representations and hierarchical feature learning. DL algorithms utilize arrange of numerous layers of nonlinear processing identities for feature extraction. The output of each sequential layer is the input of the next one, and that helps in data abstraction as we go deep within the network [15]. Convolutional Neural Network (CNN) is a class of DL and commonly used in analyzing visual imagery and designed to require minimal preprocessing [16]. It is inspired by biological processes in human brain [17] and utilized to handle data that come in multiple arrays [18]. The first use of the deep convolutional neural network with a similar form of its current one was in the recent years of the last century when Lecun introduced a deep neural network "lenet" that was used in document recognition applications in 1998 [19]. After many years it came much more popular right after using a deep convolutional neural network to classify images of (ImageNet LSVRC-2010), by utilizing a model called (AlexNet) [20]. AlexNet shows outstanding results in comparison with other used network structures at this period. Afterward, its success led to initiate consecutive successes of CNNs in the deep learning community.

The main advantages of CNNs are feature learning and providing unlimited accuracy rather than traditional machine learning and vanilla neural networks which may be achieved by increasing training samples and therefore leads to a more robust and accurate model [6]. In CNN architecture, the convolutional filters are acting as features extractors, and as we go deep, we extract more and more complex features (spatial and structural information). Feature extraction happens through convolving small filters with the input patterns followed by selection of the most distinguishing features and then start to train the classification network [18].

Brain tumors classification has been performed using many machine learning techniques and imaging modalities over the years. In 2009, Zacharaki *et al.* [21] proposed a system to classify different grades of glioma besides a binary classification for high and low grade using SVMs and KNN. Accuracy of 85% is obtained for multi-classification and 88% for binary classification. El-Dahshan *et al.* [22] introduced a method to classify 80 brain tumor normal and abnormal images using Discrete Wavelet Transform (DWT) to extract features, Principal Component Analysis (PCA) to reduce features, and then ANN and KNN to classify images with overall accuracy of 97% and 98% respectively. In 2015, Cheng *et al.* [23] proposed a method to enhance the brain tumor classification performance by augmenting the tumor region via image dilation and then by splitting into subregions. They used three approaches to extract features; intensity histogram, Gray Level Co-occurrence Matrix (GLCM) and Bag of Words (BOW) and finally achieved best accuracy of 91.28% by using ring form partition in addition to tumor region augmentation.

In the work proposed by Ertosun and Rubin [24], the authors used CNN to classify different grades of gliomas pathological images (Grade II, Grade III and Grade IV) and another task to classify Low-Grade Glioma (LGG) vs. High-Grade Glioma (HGG). They obtained accuracies of 71% and 96% respectively. Paul *et al.* [25] used axial brain tumor images to train and develop two main approaches for classification (fully connected neural network and a convolutional neural network), CNN architecture was formed of two convolutional layers with two corresponding max-pooling layers followed by two fully connected layers and achieved maximum accuracy of 91.43%. Posteriorly, Afshar *et al.* [26] presented a capsule network (CapsNet) that integrates both the MRI brain image and the coarse tumor boundaries to classify the brain tumor. Accuracy of 90.89% has been obtained in this study. In another study, Anaraki *et al.* [27] proposed a models of two combined regulations to classify brain tumor images based on CNN and Genetic Algorithms (GA-CNN), in the first case study, accuracy of 90.9% has been attained

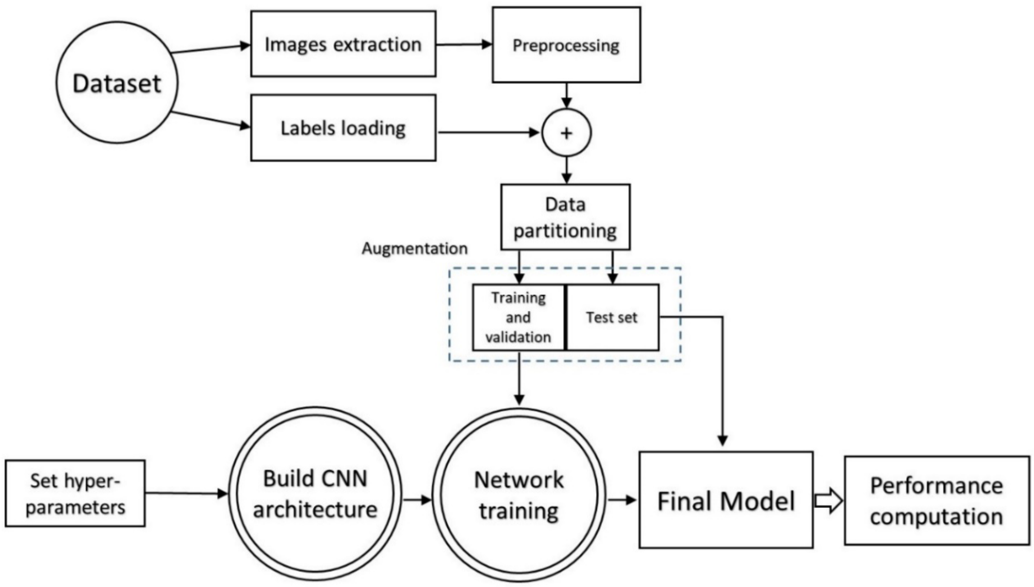


FIGURE 1. **Block diagram of the proposed method.**

in classifying three grades of glioma, while 94.2% accuracy for classification of glioma, meningioma and pituitary tumor have been fulfilled in the second case study.

In this paper, a CNN architecture is proposed to classify different types and grades of brain tumors. The architecture of the network is evolved using different configuration to acquire the most appropriate structure. The paper is organized as follows; in section 2, the proposed methodology is discussed in details starting from the original dataset and how manipulation occurs to adapt the CNN model to the tools and hardware resources used in this research. Section 3 and, 4 are dedicated to results and discussion respectively followed by a conclusion in Section 5.

II. METHOD

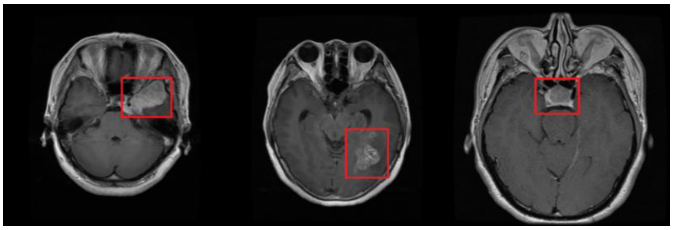
FIGURE 1 shows the block diagram of the proposed method, in which the system starts to load and extract images and labels from datasets raw files and then make a preprocessing and augmentation techniques just after splitting the dataset into training, validation and test sets. Then, the structure of the proposed method is introduced, followed by setting the hyper-parameters, regularization techniques, and optimization algorithm. Finally, network training and performance computations are presented.

A. DATASET

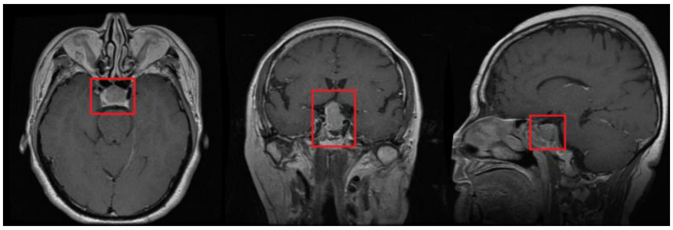
We use in this work two different datasets. The first one is acquired from Nanfang Hospital and General Hospital, Tianjing Medical University, China from 2005 to 2010 [23], and then published online with various versions since 2015 reaching to its last release in 2017. The database includes T1-weighted contrast-enhanced images acquired from 233 patients with three types of brain tumors that are meningioma, glioma, and pituitary tumor [28]. Brain tumors can be different in shape, location, and size according to their types and grades as figured in FIGURE 2(a). The dataset includes three different views: axial, coronal and sagittal views as shown in FIGURE 2(b). The second dataset is obtained from The Cancer Imaging Archive (TCIA) public access repository [29]. The Repository of Molecular Brain Neoplasia Data (REMBRANDT) contains MRI multi-sequence images from 130 patients with different diseases, grades, races, and ages [30]. We selected images on T1-weighted contrast-enhanced that include different grades of glioma (Grade II, Grade III, and Grade IV) as shown in FIGURE 3. Table 1 and Table 2 show supplementary details about the description of the two datasets respectively.

B. PRE-PROCESSING STAGE

Before feeding the images into the proposed structure, a pre-processing step is performed. The first process is to downsize the original image from $512 \times 512 \times 1$ pixels into $128 \times 128 \times 1$ pixels in order to decrease dimensionality, computations and help the network to show a better performance in lower time and more straightforward calculations. Then, data is shuffled before splitting them to maintain the system to train on unsorted data and prevent focusing on a narrow band of the entire dataset. Data is divided into three sections; training, validation, and test sets all with their individual target labels (68% for training and 32% for system test and validation). Finally, we augment the images of study I so that the system can identify them as new ones, and that is usually used to avoid overfitting and increase model robustness [20], [31]. In addition to this geometric augmentation, a grayscale distortion (salt noise) is added to the images. FIGURE 4 shows the presented augmented images vs. the



(a)



(b)

FIGURE 2. (a) Different three axial brain tumor types as follows; Meningioma, Glioma and Pituitary tumor from left to right respectively, (b) Pituitary tumor is demonstrated in three different acquisition views (Axial, Coronal, and Sagittal) from left to right respectively. Tumors are localized inside a red rectangle.

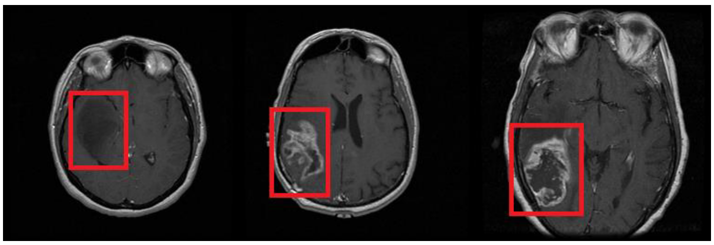


FIGURE 3. Different glioma grades included in REMBRANDT dataset (Grade II, Grade III and Grade IV from left to right respectively). Tumors are localized inside a red rectangle.

TABLE 1. Number of slices for each brain tumor type (meningioma, glioma and pituitary) in dataset I and their corresponding number of patients.

Tumor Category	Number of Patients	Number of Slices
Meningioma	82	708
Glioma	91	1426
Pituitary	60	930
	233	**3064**

TABLE 2. Number of slices for glioma grades in dataset II and their corresponding number of patients.

Tumor Category	Number of Patients	Number of Slices
Grade II	33	205
Grade III	19	129
Grade IV	21	182
	73	**516**

original one; changes include flipping around the x-axis, right/left mirroring, adding salt noise and image rotation by 45 degrees for study I. By doing this augmentation process, we have increased the original 3064 images by a factor of 5, so the final dataset became 15,320 images for study I and 516 images for study II.

C. PROPOSED CNN ARCHITECTURE

FIGURE 5 shows the proposed CNN structure. It includes 16 layers starting from the input layer which hold the augmented images from the previous pre-processing step passing through the convolution layers and their activation functions that used in features selection and down-sampling (convolution, Rectified Linear Unit (ReLU), normalization and pooling layers). To prevent overfitting, a dropout layer is used and followed by, a fully connected layer and a softmax layer to predict the output and finally a classification layer that produces the predicted class.

The description of each layer is as follows; First, the input layer is used to confirm the size of input images and applying a data normalization [20]. In the proposed architecture, three convolutional layers are included. A 2D convolutional layer applies sliding **K** convolutional filters (kernels) of size (**M** $\times$ **N**) over the input images by moving the filters along the input and compute the dot product of the weights (kernels

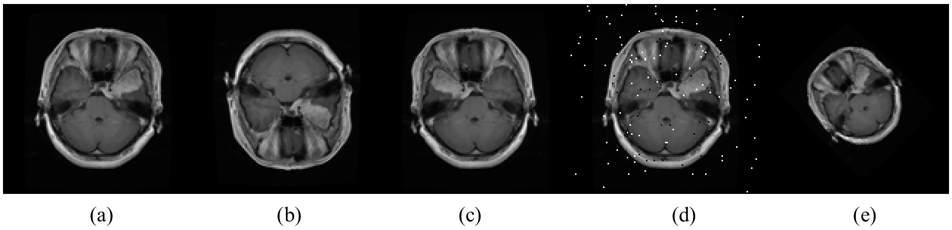


FIGURE 4. **(a) The original image, (b) up-down flipping, (c) right/left mirroring, (d) add salt noise to the image, (e) rotating by 45 degree.**

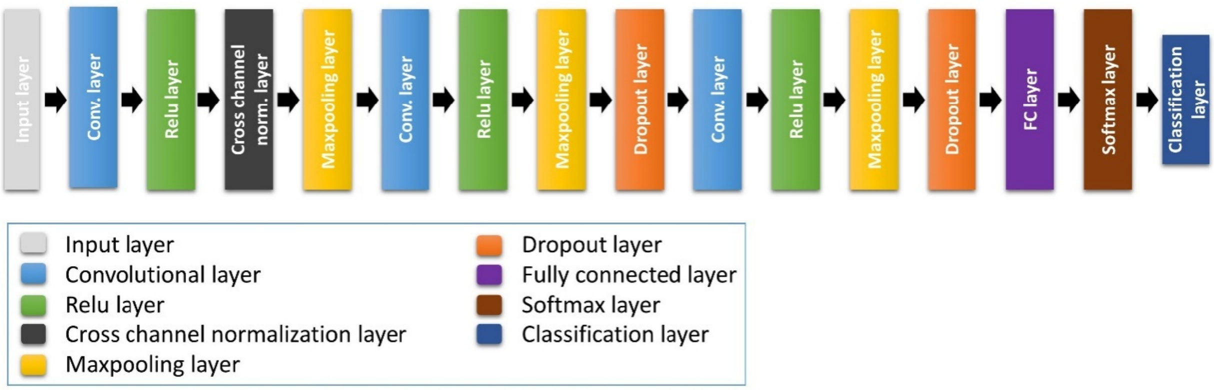


FIGURE 5. **The proposed CNN architecture.**

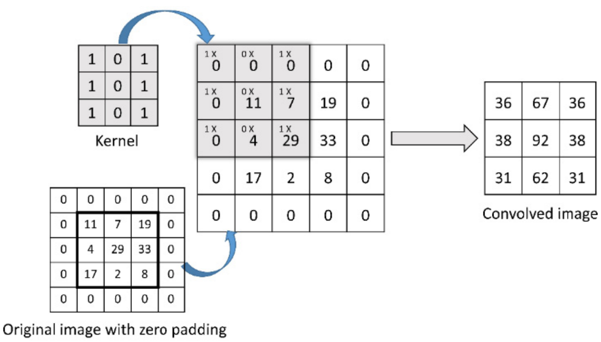


FIGURE 6. **Convolutional layer example (input: 3×3, zero padding: 1, kernel size: 3×3, Stride: 1, output: 3×3).**

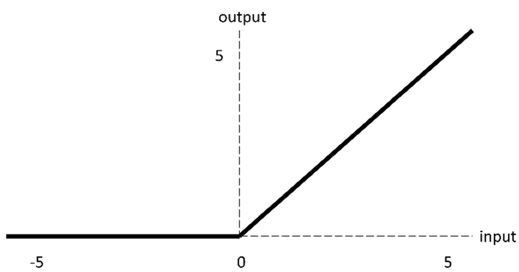


FIGURE 7. **ReLU activation function.**

weights) and the input. The filters slide over the image with vertical and horizontal steps called stride (**S**). Padding (**P**) of the original images may happen before sliding the filters in order to maintain information at the edges. These kernels are used as features identifiers; such that kernels in the early layers detect only low-level features like (edges, lines and blobs), while advanced ones are used to detect more and more complex features [19].

FIGURE 6 shows an example of applying a kernel of size 3×3 (appears in grey) over a 3×3 image producing the same input dimensions of 3×3 after kernel sliding and dot product. The involved parameters we have used are; $\mathbf{K} = 64$, 128 and 128, $\mathbf{M} \times \mathbf{N} = 10 \times 10$, 3×3 and 2×2, $\mathbf{S} = [1, 1]$, $\mathbf{P} = [0,0,0,0]$, [2,2,2,2] and [2,2,2,2] for the convolutional layers 1, 2 and 3 respectively. Every convolutional layer is followed by a non-saturated activation function called ReLU that is mainly used to decrease the training time dramatically comparing to other activation functions [20], [32], [33]. The following equation describes the ReLU model as a function of $\mathbf{x}$ in which the output equals the input when $\mathbf{x}$ is positive and 0 for other values [33]. ReLU function is graphically represented in FIGURE 7.

$$f(x) = max(0, x) \tag{1}$$

Then, a cross-channel normalization layer is used to normalize the input layer by scaling and adjusting the related activations. It makes a local response normalization layer based on a channel-wise with a window of a particular size (it has been arbitrarily chosen to be 5). The normalization can be used in backpropagation and network training acceleration [20], [32].

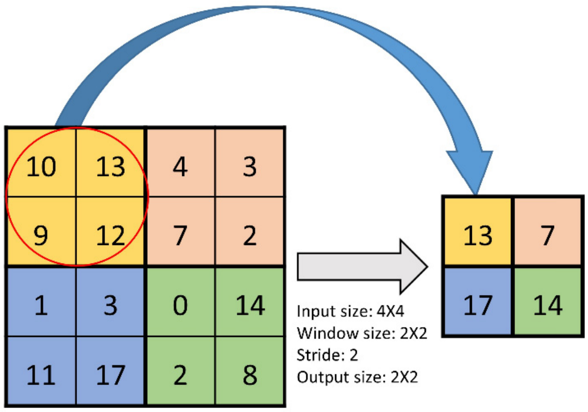


FIGURE 8. Example of a max-pooling layer (the maximum value out of a specific window (appears in the same color) is only considered).

As for maximum Pooling layer, it is a way of down-sampling used to achieve spatial invariance by splitting the whole image into small rectangles (2×2 in the proposed structure) that are moving over the image with a determined step (2×2) and then consider only the maximum value of the four elements. The pooling layer is used to reduce numbers of parameters and consequently computations in the network [34], [35]. An example of a max-pooling layer is shown in FIGURE 8.

One of the most common methods to reduce overfitting is to use a dropout layer (an example is shown in FIGURE 9). In this layer, some activations (nodes) are dropped out randomly which significantly helps also in speeding up the training phase [36]. In the proposed structure, we have found that 10% and 20% dropout probabilities were the most suitable values for dropout layers 1 and 2 respectively.

Finally, we have used three advanced layers; Fully Connected layer (FC), softmax layer and classification layer. The former one is used to connect every neuron in a layer to every neuron in another one (following and preceding); the output of this layer is three classes. Then, the FC layer is followed by a softmax layer that is also called (normalized exponential function). Softmax layer is used to squash all the predicted classes between 0 and 1, and the total sum of these values is equal to 1 (100%). The output of this layer can be calculated as follows:

$$y(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \tag{2}$$

The probability of any class (j) can be calculated over (k) different classes as a function y (z) and their total summations are equal to 1 [37] as shown in FIGURE 10. Finally, we use a classification layer which is based on cross-entropy loss to estimate the classification loss and provides the final predicted categorical label for each input image. Loss can be estimated from equation (3), where p is the target labels vector, and q (x) is the output vector from the softmax layer.

$$H(p, q) = -\sum_{x} (p(x) * log(q(x))) \tag{3}$$

D. REGULARIZATION TECHNIQUES AND OPTIMIZATION ALGORITHM

Regularization means fitting a solving function well during training while preventing system overfitting. Many techniques have been used to avoid overfitting during preprocessing and training phases. First, data augmentation is used to avoid overfitting by making a geometric and color distortion on the original images [23], [31], [38]. Then, different network architectures have been tested to avert network complexity. Then, dropout layers have been used to remove hidden units weights stochastically [36], [39]. L2 regularization is also used to add a penalty to the cost function and introduce weights decay as shown in the following equation [40].

$$Cost\ function' = Cost\ function\ (Loss) + \lambda \sum_{i=1}^{k} w_i^2 \tag{4}$$

where λ is the regularization parameter (hyper-parameter that is defined), w is the corresponding weight(s) for $i = 1, \ldots, k$. Finally "early stop technique" has been used in some trials which monitors the training and validation performance and stops the training process before full epochs training completion in case of system stability or before starting to overfit [40].

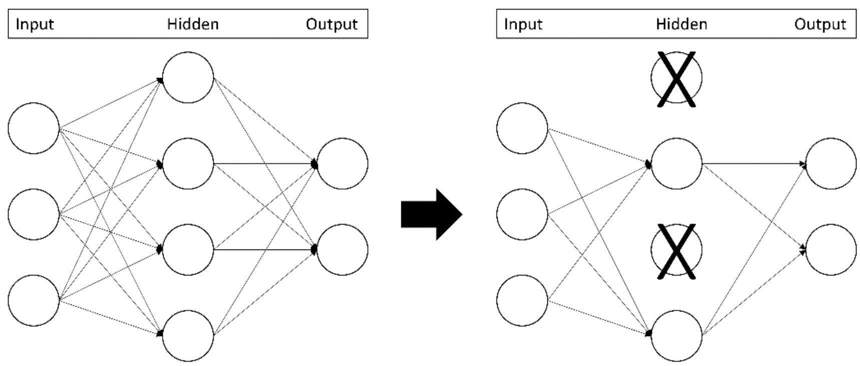


FIGURE 9. Example of dropout layer (probability of 50% appears on the right).

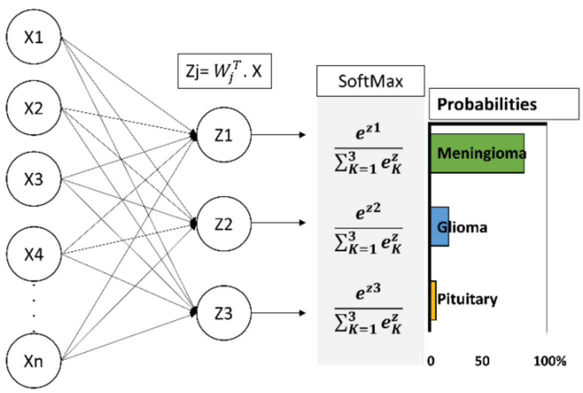


FIGURE 10. Example of Soft-max layer.

Optimization is used mainly to update network parameters and minimize loss function to reach the global minimum in the ideal case by taking small steps to the negative gradient direction (convergence) [41]. We have found that the "stochastic gradient descent with momentum" is the optimal optimizer for the proposed structure.

III. EXPERIMENTS AND RESULTS

For study I, FIGURE 11 shows both the accuracy progress and loss during the validation phase for our proposed network. FIGURE 11(a) shows that almost 100% accuracy has been achieved right after 5000 iterations. After the 8550th iteration, the accuracy shows a plateau of nearly 100%, and finally, the best overall accuracy obtained during the test phase is 96.13%. While in the mini-batches loss graph FIGURE 11(b), it is clear that the curve first starts to drop sharply, but some fluctuations appear due to using a small mini-batch size of 32 images. These fluctuations tend to disappear after 6400 iterations and the loss curve almost hits zero.

For study II, FIGURE 12 shows both the accuracy progress and loss during the validation phase. We can see from FIGURE 12(a) that accuracy of 100% has been achieved right after 100 iterations. Hence, the best overall accuracy obtained during the test phase is 98.7%. From the mini-batches loss graph in FIGURE 12(b), the curve first starts to drop sharply. This slope tends to disappear after 100 iterations and the loss curve almost hits zero.

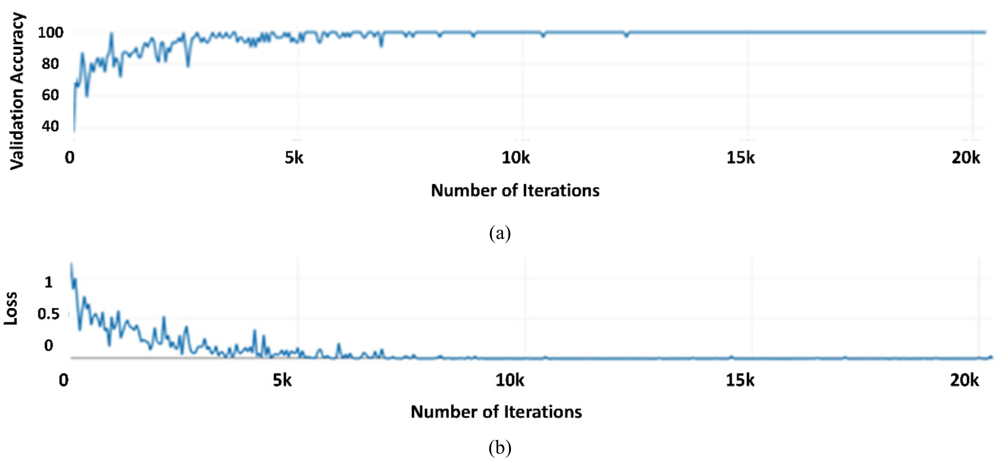


FIGURE 11. Validation accuracy and loss over the whole training iterations of study I: (a) Validation accuracy (higher is better), and (b) Loss (lower is better).

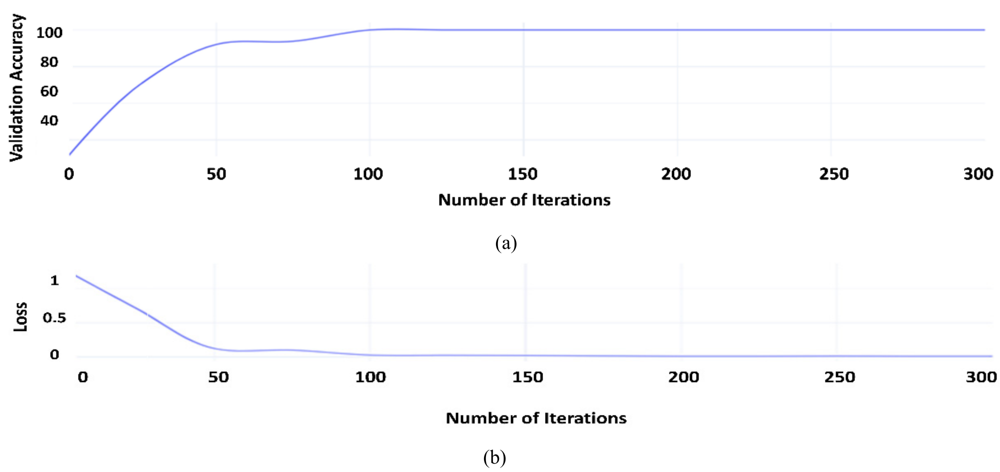


FIGURE 12. Validation accuracy and loss over the whole training iterations of study II: (a) Validation accuracy (higher is better), and (b) Loss (lower is better).

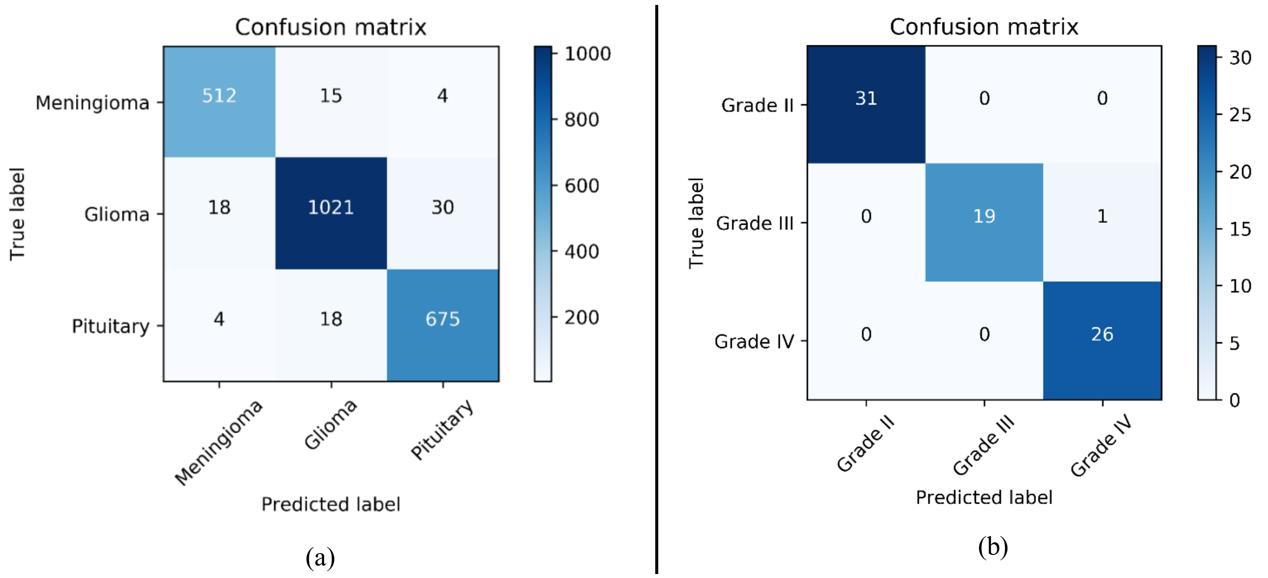


FIGURE 13. The confusion matrix of the proposed model: (a) for study I, and (b) for study II.

TABLE 3. Accuracy metrics in terms of TP, TN, FP, FN, precision, sensitivity, specificity, and accuracy.

Metrics / Method	Tumor Type	TP	TN	FP	FN	Precision	Sensitivity	Specificity	Accuracy	Total no.
Proposed Model for Study I	Meningioma	512	1744	22	24	0.958	**0.955**	**0.987**	**97.54%**	547
	Glioma	1021	1195	33	61	**0.972**	0.944	0.951	95.81%	1069
	Pituitary	675	1566	34	48	0.952	0.934	0.97	96.89%	697
Proposed Model for Study II	Grade II	31	46	0	0	**1**	**1**	**1**	**100%**	31
	Grade III	19	57	0	1	**1**	0.95	**1**	95%	20
	Grade IV	26	50	1	0	0.963	**1**	0.98	**100%**	26

A. CONFUSION MATRIX

FIGURE 13 shows the confusion matrices that summaries system's performance for both studies (I and II). The X-axis represents the predicted values (system output) while the Y-axis represents the true labels (ground truth). Precision, Sensitivity, Specificity, and Accuracy have been calculated as in equation 5.

$$\begin{aligned} Precision &= \frac{TP}{(TP+FP)} \\ Sensitivity &= \frac{TP}{(TP+FN)} \\ Specificity &= \frac{TN}{(TN+FN)} \\ Accuracy &= \frac{TP+TN}{(P+N)} \end{aligned} \tag{5}$$

where,

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

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

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

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

Table 3 shows the accuracy metrics that are extracted from the confusion matrices. The highest performance of precision, sensitivity, specificity and accuracy are bolded in Table 3. Accuracy of 97.54% is obtained to classify meningioma, 95.81% for glioma and 96.89% for

TABLE 4. Different architectures and hyper-parameters tested and tried before reaching the final model.

Factor(s)	Values
Number of convolutional + ReLU layers	1, 2, 3, 4
Number of cross-channel normalization layers	1, 2, 3
Number of drop out layers	1, 2, 3
Maximum epochs	20,40,50,60,80,100
Number of fully connected layers	1, 2, 3
Number of convolutional kernels	8, 16, 32, 64, 128, 256
Kernel size	2, 3, 5, 7, 9, 10, 11
pooling layer	Max pooling, average pooling
pooling layer window size	2, 3, 4, 5
Optimizers	SGD, Adam, RMSprop
Mini-batch size	1, 4, 8, 16, 32, 64, 128
Dropout rate	0.1, 0.15, 0.2, 0.25, 0.5
Initial learning rate	0.01, 0.001, 0.0001
Learning rate drop factor	0.1, 0.2, 0.3

TABLE 5. Comparison between the proposed model and previous related works.

	Model	Best Accuracy for Study I	Best Accuracy for Study II	Classification Type	Classification Method
1	Cheng *et al.*[23]	91.28%	-	Multi	SVM and KNN
2	Paul *et al.*[25]	91.43%	-	Multi	CNN
3	Afshar *et al.*[26]	90.89%	-	Multi	CNN
4	Anaraki *et al.*[27]	94.2%	90.9%	Multi	GA-CNN
5	Zacharaki *et al.*[21]	-	85%	Multi	SVM and KNN
6	Zacharaki *et al.*[21]	-	88%	Binary	SVM and KNN
7	El-Dahshan *et al.*[22]	-	98%	Binary	ANN and KNN
8	Ertosun *et al.*[24]	-	71%	Multi	CNN
9	Ertosun *et al.*[24]	-	96%	Binary	CNN
10	Proposed structure	**96.13%**	**98.7%**	Multi	CNN

pituitary classification. However, we achieved an accuracy of 100% in classifying glioma Grade II, 95% for glioma Grade III and 100% for glioma Grade IV.

B. EMPIRICAL ARCHITECTURES AND HYPER-PARAMETERS

In this part, the different architectures parameters involved in the selection process are presented.

Table 4 shows the different tested parameters before reaching the final introduced structure that shows the best performance.

C. TOOLS AND TIME CONSUMPTION

The proposed deep neural network structure is trained on Intel i7-7700HQ CPU (2.8 GHz), NVIDIA GTX 1060 (6 GB) GPU, 16GB RAM, Matlab 2018b and Python. The training time was 289 minutes for (10,417 images) in study I and 2.5 minutes for (350 images) in study II. The average test execution time was 8.5 and 9.6 milliseconds per image for study I and II respectively.

IV. DISCUSSION

In this paper, an approach for brain tumors classification and grading of gliomas is proposed by applying a CNN model to radiological images. Numerous CNN models parameters are used to adapt the system before achieving the last architecture. Training a CNN from scratch is challenging as it may take weeks or months to fulfill satisfactory results for a dataset without being overfitted or underfitted. Results from previous literatures that have used the same brain tumor types with different architecture, hyper-parameters and depths are summarized in Table 5. It is clear that the proposed structure gives the best prediction results compared to other related previous studies which demonstrate the reliability of the proposed system. The proposed CNN method is a segmentation free approach as we load the brain tumor image to get the corresponding class directly.

In contrast, [21]–[23] used feature engineering to extract features and then reduced their dimensions to use them in another stage for classification. In [27], the authors have used GA to indicate the architecture of the network, however, GA didn't introduce the best possible prediction results. In [25], the authors have used only two convolutional layers with 64 kernels in each one. Moreover, they have used 4 dropout layers which are relatively high for the presented network. Despite Ertosun and Rubin [24] have used pathological images to train the network, weak results have been obtained after using a combination of 2 classifiers. In [26], the authors have used tumor coarse boundaries as an extra input to help the network in better results showing. The preceding stage needs another manual process to localize tumor before training a CNN. Although, we have achieved promising classification rates, however, the proposed system in this study needs to be tested on larger scale datasets that include different ages and races to increase its portability and extend it in other medical applications in the future. Additionally, system's structure cannot be reused to classify small number of images as it is one of the deep learning limitations, but instead of that, the system can be fine-tuned after training on a large dataset to manipulate small dataset.

V. CONCLUSION

In this work, we have presented a CAD system for the classification of brain tumor MR images into three types (meningioma, glioma, and pituitary) in one study, and further classifying of gliomas into different grades (Grade II, Grade III and Grade IV) using a custom deep neural network structure. The proposed network is constructed from 16 layers starting from the input layer which holds the preprocessed images passing through the convolution layers and their activation functions (3 convolution, 3 ReLU, normalization and 3 Maxpooling layers). Additionally, two dropout layers are used to prevent overfitting followed by a fully connected layer and a softmax layer to predict the output and finally a classification layer that produces the predicted class. Although the dataset is relatively not big (due to the variety of imaging views), data augmentation helped well to show better results and hence overcome this problem. Our proposed architecture has achieved the highest accuracy of 96.13% and 98.7% concerning the two datasets used in this paper.

REFERENCES

[1] L. M. De Angelis, "Brain tumors," *New England J. Med.*, vol. 344, no. 2, pp. 114–123, Jan. 2001.

[2] B. W. Stewart and C. P. Wild, *World Cancer Report 2014*. Lyon, France: IARC, 2014.

[3] *Brain, Other CNS and Intracranial Tumours Statistics*. Accessed: May 2019. [Online]. Available: https://www.cancerresearchuk.org/

[4] A. Behin, K. Hoang-Xuan, A. F. Carpentier, and J.-Y. Delattre, "Primary brain tumours in adults," *Lancet*, vol. 361, no. 9354, pp. 323–331, 2003.

[5] A. Drevelegas, *Imaging of Brain Tumors With Histological Correlations*. Berlin, Germany: Springer, 2002.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[7] M. L. Goodenberger and R. B. Jenkins, "Genetics of adult glioma," *Cancer Genet.*, vol. 205, no. 12, pp. 613–621, Dec. 2012.

[8] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization classification of tumors of the central nervous system: A summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, Jun. 2016.

[9] C. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.

[10] T. Rajesh and R. S. M. Malar, "Rough set theory and feed forward neural network based brain tumor detection in magnetic resonance images," in *Proc. Int. Conf. Adv. Nanomaterials Emerg. Eng. Technol. (ICANMEET)*, Jul. 2013, pp. 240–244.

[11] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in *Proc. Int. Conf. Ind. Instrum. Control (ICIC)*, May 2015, pp. 60–65.

[12] M. Shasidhar, V. S. Raja, and B. V. Kumar, "MRI brain image segmentation using modified fuzzy C-means clustering algorithm," in *Proc. Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Jun. 2011, pp. 473–478.

[13] S. Goswami and L. K. P. Bhaiya, "Brain tumour detection using unsupervised learning based neural network," in *Proc. Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Apr. 2013, pp. 573–577.

[14] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *Proc. Sci. Inf. Conf.*, Aug. 2014, pp. 372–378.

[15] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.

[16] Y. LeCun. (2015). *Lenet-5, Convolutional Neural Networks*. Accessed: May 2019. [Online]. Available: http://yann.lecun.com/exdb/lenet

[17] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Netw.*, vol. 16, nos. 5–6, pp. 555–559, Jul. 2003.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Jan. 2012, pp. 1097–1105.

[21] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1609–1618, Dec. 2009.

[22] E. S. A. El-Dahshan, T. Hosny, and A. B. M. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digit. Signal Process.*, vol. 20, pp. 433–441, Mar. 2010.

[23] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0140381.

[24] M. G. Ertosun and D. L. Rubin, "Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks," in *Proc. AMIA Annu. Symp. Proc.*, vol. 2015, Nov. 2015, pp. 1899–1908.

[25] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, "Deep learning for brain tumor classification," *Proc. SPIE, Med. Imag., Biomed. Appl. Mol., Struct., Funct. Imag.*, vol. 10137, Mar. 2017, Art. no. 1013710. doi: 10.1117/12.2254195.

[26] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on MRI images and course tumor boundaries," 2018, *arXiv:1811.00597*. [Online]. Available: https://arxiv.org/abs/1811.00597

[27] A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybernetics Biomed. Eng.*, vol. 39, no. 1, pp. 63–74, Jan./Mar. 2019.

[28] J. Cheng. "Brain Tumor Dataset." Apr. 2, 2017. Distributed by Figshare. Accessed: May 30, 2019. [Online]. Available: https://figshare.com/articles/brain_tumor_dataset/1512427/5

[29] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.

[30] L. Scarpace, A. Flanders, R. Jain, T. Mikkelsen, and D. Andrews, "Data from REMBRANDT. The cancer imaging archive," 2015. doi: 10.7937/K9/TCIA.2015.588OZUZB.
[31] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" 2016, *arXiv:1609.08764*. [Online]. Available: https://arxiv.org/abs/1609.08764
[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[34] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks*. Berlin, Germany: Springer, 2010, pp. 92–101.
[35] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011, pp. 342–347.
[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[37] N. M. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, 2007, Art. no. 049901.
[38] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," 2012, *arXiv:1202.2745*. [Online]. Available: https://arxiv.org/abs/1202.2745
[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT press, 2016.
[41] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*. Physica-Verlag HD, 2010, pp. 177–186.

HOSSAM H. SULTAN was born in Giza, Egypt, in 1990. He received the B.Sc. degree in biomedical engineering from Helwan University, Egypt, in 2012, where he is currently pursuing the M.Sc. degree in biomedical engineering and related fields. He is currently a Senior Service Engineer. His research interests include image processing and machine learning for medical diagnostics.

NANCY M. SALEM was born in Egypt, in 1975. She received the B.Sc. degree and M.Sc. in electronics and communications engineering from Helwan University, Cairo, Egypt, in 1998 and 2003, respectively, and the Ph.D. degree in engineering from the Department of Electrical Engineering and Electronics, The University of Liverpool, U.K., in 2007. In 2008, she joined the Department of Electronics and Communications Engineering, Helwan University, as an Assistant Professor. She is currently an Associate Professor with the Department of Biomedical Engineering, Helwan University. Her current research interests include signal processing, medical image processing, machine learning, and watermarking and modeling. She is a member with the Institute of Electrical and Electronics Engineers and the Counselor of the Women in Engineering (WIE) and IEEE Helwan Student Branch. She reviewed many scientific papers, proposals and research projects.

WALID AL-ATABANY received the B.Sc. and M.Sc. degrees from the Biomedical Engineering Department, Cairo University, in 1999 and 2004, respectively, and the Ph.D. degree in biomedical engineering from Imperial College London, in 2010. In 2011, he was a Research Associate with Newcastle University for two years. He is currently an Associate Professor with the Biomedical Engineering Department, Helwan University, Egypt. His research interests include signal and image processing (particularly for visually impaired), medical imaging, and modeling of retinal processing. He received the 2nd Price Award from the 2nd Symposium of the Neuroscience Technology Network (NTN2009), the ARVO 2010 travel grant from the AFER/National Institute for Health Research Biomedical Research Centre for Ophthalmology, and two grants from the Newton institutional link grant from the British council, in 2015 and 2016.