Deep learning model for glioma, meningioma and pituitary classification

Toqa A. Sadoon, Mohammed H. Ali
Department of Electronic and Communications Engineering, College of Engineering, Al-Nahrain University, Baghdad, Iraq

Article Info
Article history:
Received Sep 15, 2020
Revised Dec 23, 2020
Accepted Jan 13, 2021

Keywords:
Brain tumor
CNN
Deep learning
Image classification
Image processing

ABSTRACT
One of the common causes of death is a brain tumor. Because of the above mentioned, early detection of a brain tumor is critical for faster treatment, and therefore there are many techniques used to visualize a brain tumor. One of these techniques is magnetic resonance imaging (MRI). On the other hand, machine learning, deep learning, and convolutional neural network (CNN) are the state of art technologies in the recent years used in solving many medical image-related problems such as classification. In this research, three types of brain tumors were classified using magnetic resonance imaging namely glioma, meningioma, and pituitary gland on the based of CNN. The dataset used in this work includes 233 patients for a total of 3,064 contrast-enhanced T1 images. In this paper, a comparison is presented between the presented model and other models to demonstrate the superiority of our model over the others. Moreover, the difference in outcome between pre- and post-data pre-processing and augmentation was discussed. The highest accuracy metrics extracted from confusion matrices are; precision of 99.1% for pituitary, sensitivity of 98.7% for glioma, specificity of 99.1%, and accuracy of 99.1% for pituitary. The overall accuracy obtained is 96.1%.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Toqa A. Sadoon
Department of Electrical and Communication Engineering
Al-Nahrain University
PO Box 64040, Baghdad, Iraq
Email: toqa.am.wd@gmail.com

1. INTRODUCTION
The second most important cause of death according to the world health organization is cancer, and there are about 9.6 million deaths in 2018 because of cancer. Globally, about 1 out of every 6 deaths from cancer. On the other hand, the brain considers one of the most complex organs in the human body that works with billions of cells. The accumulation of abnormal cells in the brain leads to the so-called brain tumor. A brain tumor is divided into two categories, primary and secondary. The first one arises in the brain, while the second one arises from other parts of the body. The tumors can be cancerous (malignant) or non-cancerous (benign). Cancerous brain tumors grow rapidly and spread to other areas of the brain compared to non-cancerous tumors. Glioma, meningioma, and pituitary are other different types of brain tumors [1]. On a larger scale, glioma tumor which is the most common type of primary brain tumor [2] are classified into four grades, and the higher the grade, the more malignant the tumor, and originate in the glial cells of the brain [3]. Meningiomas, which originate from a layer of tissue called the meninges, are sometimes considered benign tumors. The growth of this species is slow and less widespread. While the pituitary tumor grows on the pituitary gland. These tumors are also benign and less widespread [4, 5].

Journal homepage: http://ijaas.iaescore.com
Artificial intelligence (AI), machine learning (ML), and deep learning (DL) are three concepts relative to each other. In general, DL is a type of ML, and the last is a type of AI. To begin with, artificial intelligence refers to any technology that has some intelligence. On the other hand, ML is a technique used to form a model out of data, and has many technologies; one of these technologies is deep learning, which is our interest in this research. Depending on the application that used machine learning, and the training method ML techniques are classified into three kinds, supervised learning, unsupervised learning, and reinforcement learning. In the first one, all the training data set must consist of inputs with the labels together, and what the model must be found in the testing step for a given input is the predicted output. For example, regression, and classification which is our interest. While the training data in unsupervised learning contain inputs alone without valid outputs. In this type of learning, it is used for clustering. Reinforcement learning uses input, output, and grade in the training data. For example, such as control and gameplays. This paper only covers supervised learning. Moreover, three concepts must be understood in supervised learning which are; firstly, training data refers to the data used through the training step to train the model. Secondly, validation data, this collection of data used to compare the training data with predicted output. Finally, testing data used to measure the model's performance after the training process [10].

DL is one of the ML techniques [10], it consists of multiple layers for learning data. These methods have greatly improved the latest in speech recognition technology, Image recognition, object detection, and many other areas such as disease detection. Deep learning consists of two types, convolutional neural networks (CNN) has made breakthroughs in speech, image, and video processing, and recurrent neural networks (RNN) highlighted sequential data such as text and speech [11].

ConvNet is a specialized kind of neural network for data processing. CNN refers to that the network uses a process from math namely convolution instead of the general matrix multiplication [12]. CNN became more famous after AlexNet [13] has been a record in 2012 which is designed by Krizhevsky et al, and showed on ILSVRC excellent performance [14]. AlexNet's success in setting the way for the invention of different CNN models [15] in addition to applying those models in various areas of natural language processing and computer vision [16]. But actually, It was an ancient technology that was developed in the 1980s and 1990s. In other words, CNN is a deep neural network with many hidden layers. It is worth mentioning, it is not good to use the original images during training, because this leads to poor results. Therefore, to extract the features the images must be processed. To do that, there are many techniques used for this purpose, which are independent of machine learning, and this is taking a great deal of time and cost, while on CNN the story is different. There are some of the hidden layers in ConvNet that are responsible to extract the feature through the training process, and the other responsible for example classification. Moreover, when CNN contains more deep hidden layers, leads to more feature extraction, as a result, better performance.

These days, the power of deep learning and graphics processing unit (GPU) can be an important tool to develop many networks that use to solve different problems. These networks can be run through a high-level programming interface based on NVIDIA GPUs accelerated libraries. TensorFlow, PyTorch, MATLAB, MXNet, NAVIDA Caffe, PaddlePaddle, Chainer are common deep learning frameworks. In this study, we use MATLAB 2018 framework and the power of GPU and CUDA from NAVIDA GeForce 920M.

MRI Brain tumor classification based on a machine learning technique has been performing over many years. In [17], S. Charfi et al, presented a technique for brain tumor classification into normal or malignant and abnormal or benign. They used the histogram dependent thresholding for image segmentation. Moreover, for feature extraction the authors used discrete wavelet transform, for reducing the dimensionality of the wavelet coefficients used principal component analysis and the feed-forward back-propagation neural network for classification. The classification accuracy on both training and test images is 90%. J. Cheng et al. [1], presented a model to enhance the performance of MRI classification of brain tumors with three categories, meningioma, glioma, and pituitary. Firstly they used a region of interest (ROI) from the augmented tumor
region via image dilation. Secondly, they split the augmented tumor into fine ring-form subregions. Finally, they have been used bag-of-words (BoW), intensity histogram, and gray level co-occurrence matrix (GLCM) feature extraction. The best accuracy has been evaluated is 91.28%.

While Paul et al. [18], presented a model to classify MRI brain tumors into three categories, pituitary, meningioma, and glioma. They used just axial slices and two types of neural networks (fully connected neural network and CNN) which contain two layers of (convolutional layers, max-pooling layers followed by fully connected layers) and finally achieved maximum accuracy of 91.43%. In [19], Parian Afshar et al. used CapsNet architecture to classify MRI brain tumor into glioma, meningioma, and pituitary. They take the tumor coarse boundaries as extra inputs for the training process. The accuracy of this model is 90.89%.

Amin Kabir et al. [20], employ the genetic algorithm (GA) to achieve the best performance of the CNN model for classification MRI into four glioma grades and three types of the brain (glioma, meningioma, and pituitary), unlike other methods rest on trial and error. The accuracy of this model is, study I: 90.0%, study II: 94.2%. Zar Nawab Khan Swatia et al. [21], proposed a model to classify MRI of brain tumors into three types (glioma, pituitary, and meningioma). They are used VGG19 to initialized weights. After that, they applied fine-tuned VGG19 on the dataset. The average accuracy was 94.82%.

In [22], Muhammed Taloo et al. employ the Vgg-16, AlexNet, ResNet-34, ResNet-18, and ResNet-50 pre-trained models to classify MRI brain into five classes which are normal, inflammatory, degenerative, neoplastic, and cerebrovascular diseases classes. The best accuracy of classification obtained was 95.23% ± 0.6 in the case of the ResNet-50 model among other models. In this paper, a model of CNN has been presented to classify MRI of brain tumors into three types, glioma, pituitary, and meningioma. The network architecture with various numbers of layers and parameters is developed on a trial-and-error basis to arrive at the best model. The proposed method consists of the following stages: Data preprocessing, data augmentation, localization of brain tumors, CNN for feature extraction, and classification. The remainder of the paper is organized as follows; the materials and methods section to describe the proposed model in detail. After that, the results and discussion section to summarize all the result and the comparison obtained from the model. Finally, the conclusion to describe all the work briefly.

2. RESEARCH METHOD
2.1. Data set preparation
Medical images that have been used in this work consist of 3064 T1-weighted contrast-enhanced MRI (CE-MRI) from 233 patients of either sagittal, axial, or coronal views. These data sets were used in Cheng et. al. [1] for classification and was collected from Tianjing Medical University, Nanfang Hospital, General Hospital, and Guangzhou in China from 2005 to 2012.

2.2. Data pre-processing
Image processing tools have been used extensively in medical imaging techniques and can improve the accuracy of diagnostic processes, and typically include image enhancement to reduce the effects of corruption that can contaminate medical images during the acquisition or transfer process [23]. In this thesis, the pre-processing on MRI brain scan slices involves implementing many algorithms as a preparation for the feature extraction in the convolutional layer. This preparation includes MRI dimensions resizing and using a Gaussian filter for MRI slice enhancement. In the case of resizing, the scan of MRI was resized to 128 × 128 pixels, as a result, the algorithm proposed in this work was implemented based on using squared slices. On the other hand, image enhancement is a complex task that is highly dependent on the nature of the image. Several types of noise can be found in images that require different image enhancement techniques. The visual quality of the medical image plays an important role in the accuracy of the clinical diagnosis because doctors are usually trained and have experience in specific, high-quality medical images. A low-pass filter Gaussian filter was applied for noise removal [24].

2.3. Brain tumor location
For faster and more accurate diagnosis automatically, the tumor has been detected. This operation helps the model to focus on a specific region (a tumor just) in the image and not all the image dimensions. By feeding the neural network with the image of detected tumors, the structure can be better learned and steps are taken to distinguish brains with and without tumors.

2.4. Data augmentation
When using multi-layered deep nets or handling a limited number of training images, there is a risk of overfitting. The standard solution to reduce overfitting is to increase data that artificially extend the data set [10]. Common augmentation techniques that have been used in this work include sub transformations such as
rotation with 45 angles, flipping, mirror, and noise addition. Figure 1 show an example of the data set after the above processing. In addition, Table 1 summarize the number of datasets before and after augmentation.

![Figure 1](image)

Figure 1. Sample of different types of brain tumor, (a) glioma, (b) meningioma and (c) pituitary after pre-processing and data augmentation which are the original image, rotating by 45 degree, add salt & paper noise, mirror and up-down flipping from left to right.

| Category   | Number of slice before augmentation | Number of slices after augmentation |
|------------|------------------------------------|------------------------------------|
| Glioma     | 1426                                | 7130                               |
| Meningioma | 708                                 | 3540                               |
| Pituitary  | 930                                 | 4650                               |
| Total      | 3064                                | 15320                              |

2.5. CNN classification model

As we mentioned in the previous sections, CNN is a neural network using for classifying images and displayed good performance in categorizing different supervised learning tasks [25]. We have been tested many layers and parameters in this work, and the best one for this model was the following. The model includes 28 layers that began with the input layer that takes the image with 128 x 128 x 3 size after pre-processing and augmentation. These images are passing through six layers of (convolutional layer, Rectified Linear Unit (ReLU), and max-pooling layer) respectively. Absolutely, we use five dropout layers to prevent from overfitting. The last three layers are in sequence fully connected layer, the softmax layer, and finally classification layer. Moreover, we use after the first convolutional layer batch normalization layer. The following sentences describe the behavior of each layer in details. The input layer is used to enter the training data to the model with input size 128 x 128 x 3.
The convolutional layer is used to extract the feature from the input image. In this layer, there is a filter called kernel that has been convolved with the input image. And we must mention to that, the kernel in the early layers used to extract low-level features from the image such as lines and edges. While the kernel in the advanced layers is used to extract complex features [26].

The output of this layer is a new set of images called feature maps, which is equal to the number of a kernel that has been used in this layer [10]. In this work, the numbers of filters are 64, 64, 96, 96, 128, and 128 with kernel size 7 x 7, 9 x 9, 9 x 9, 9 x 9, 11 x 11, and 11 x 11 respectively. Stride is moving along the vertical or horizontal position of the image by one or more step size through a convolutional operation. The stride size is one for all the convolutional layers. But when we give a border size of the image more importance this is called by padding and this is done by adding extra row and column around the image matrix. The padding size used in this work is 0, 1, 1, 1, 1, and 1. Figure 2 show an example of a convolutional layer operation with a kernel size of 3 x 3.

The batch normalization layer is used to normalize the training data during training processing rather than normalizing all the data set in the pre-processing step and this process will decrease the training time [27]. Each convolutional layer is followed by an activation function used to determine the behavior of the connection node. In our model we use rectifier linear unit (ReLU), the output of this activation is a positive number and zero. The following equation is the mathematical representation for this function (1).

$$ f(x) = \max(0, x) $$

As for the max-pooling layer, it is used to reduce the feature maps dimension after a convolutional operation. Similar to the convolutional layer, pooling layer also has a filter moving on the feature map and as a result reduces the computation of the network [16, 28, 29]. The filter’s size is 2 x 2 with one stride size for all max-pooling layers. Figure 4 show the max-pooling layer with a 2 x 2 filter size.
Deep learning for glioma, meningioma, and pituitary classification (Toga A. Sadoon)

One of the common problems through training is overfitting. Overfitting is mean that, best learning performance against bad testing performance. In order to prevent the model from overfitting, we use a dropout layer. In this layer, some nodes have been selected randomly and set to zero, the number of selected nodes depended on a percentage value. In the proposed model, we found that the best dropout probabilities are 10%, 10%, 20%, 20%, and 20% respectively for the five dropout layers. An example of the dropout layer shown in Figure 5.

Finally, the last three layers which are fully connected layer, the softmax layer followed by the classification layer. The first layer (fully connected layer) is used to convert the two-dimensional image into 1D. In this layer, each neuron is connected to the previous neuron and the next neuron. The output of this layer is the same number of categories which are three classes in our case. The last layer of fully connected layers uses a special function to predict the probable outcome for each category, and the biggest value of probability represented the correct class. In this model, a softmax function has been used. To calculate the output of this layer we can use (2). Figure 6 show an example of the last three layers.

\[ f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \]  

Figure 4. Example of max pooling layer

Figure 5. Example of dropout layer [30]

Figure 6. Softmax layer with neural network [31]
Lastly, cross-entropy loss function has been used in the classification layer to determine the error (the difference between the actual output and the predicted output) of the classification and produce the final predicted class for each input image. The equation (3) used to calculate the error:

$$J = \sum_{i=1}^{M} \{-d_i \ln(y_i) - (1 - d_i) \ln(1 - y_i)\} + \lambda \frac{1}{2} \|w\|^2$$

(3)

Where is the output of the network, is the correct output, \(\lambda\) is a coefficient related to the connection weight and cost function, and represent the output node until M nodes. Moreover, to reduce the error we use stochastic gradient descent with momentum (sgdm) as an optimizer method. The proposed model shows in Figure 7.

Figure 7. Proposed work block diagram

3. RESULTS AND ANALYSIS

We divided the data set into 77% for training, 20% for validation, and the remainder used for model testing. In the training process for deep learning, momentum is set to 0.9, the maximum iteration 36800, the epoch is 100, the initial learning rate is 0.0001 and the mini-batch size is 32. Figure 8 show the accuracy and error for both training validation progress. Figure 8 show that after 10000 iterations the accuracy became near to 100% and in the final, the best validation accuracy obtained is 96.1%. While the loss function is less than 0.2. We must mention that because of using 32 images as a mini-batch size the curve firstly drops sharply with some fluctuations [32] but these tend to disappear after 10000 iterations for both curves. On the other hand, for model testing, we use 459 slices and the model shows the test accuracy of 93.2%.

Figure 8. Training process

3.1. Number of layer and hyper parameters

In this subsection, the different parameters and number of layers of the model that have been tested until reaching the best model are presented in Table 2.
Table 2. Various previously tested parameters and layers.

| Trial and error parameters | Pooling layer stride |
|----------------------------|----------------------|
| Number of (Convolutional layer+ReLU+Pooling layer) | 1, 2, 3, 4, 5, 6 |
| Pooling layer              | Max, average pooling |
| Dropout layer              | 1, 2, 3, 4          |
| Dropout ratio              | 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5 |
| Epoch                      | 30, 50, 70, 100     |
| Learning rate              | 0.01, 0.001, 0.0001, 1e-4, 1e-5 |
| Optimization methods       | SGD, Adam           |
| Mini-batch size            | 16, 32              |
| Number of kernel           | 48, 64, 96, 128     |
| Activation function        | ReLU, leakyReLU      |
| Kernel size                | 3, 5, 7, 9, 11      |
| Convolutional layer padding| 0, 1, 2, 3, 4       |
| Convolutional layer stride | 1, 2                |
| Pooling layer padding      | 0, 1                |
| Pooling layer stride       | 1, 2                |

3.2. Confusion matrix

The confusion matrices have been used to measure the model's performance for our study. Precision, sensitivity, specificity, and accuracy have been determined using the following equations:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (4)
\]
\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)
\]
\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (6)
\]
\[
\text{Accuracy} = \frac{(TP + TN)}{\text{total}} \quad (7)
\]

Where, TP, FP, TN, FN are true positive, false positive, true negative, and false negative, respectively. To describe the confusion matrix we will mention to people with tumor by positive and people without tumor by negative. Moreover, true and false for correctly and incorrectly diagnose respectively.

So,
- True positive (TP): people with tumor correctly identified (correctly diagnose).
- False positive (FP): people without tumor incorrectly identified people with tumor (incorrectly diagnose).
- True negative (TN): people without tumor correctly identified as healthy (correctly diagnose).
- False negative (FN): people with tumor incorrectly identified as people without tumor (incorrectly diagnose).

Figure 9 show the accuracies that are found from the confusion matrix and summarized in Table 3. Precision of 99.1% for pituitary, sensitivity of 98.7% for glioma, specificity of 99.1%, and accuracy of 99.1% for pituitary are the highest performance.

![Confusion Matrix](image_url)
Table 3. Confusion matrix accuracy

| Tumor type | TP    | TN    | FP    | FN    | Precision % | Sensitivity % | Specificity % | Accuracy % |
|------------|-------|-------|-------|-------|-------------|---------------|---------------|------------|
| Glioma     | 1337  | 1620  | 89    | 18    | 93.8        | 98.7          | 94.8          | 96.5       |
| Meningioma | 685   | 2275  | 23    | 81    | 96.8        | 89.4          | 96.8          | 96.6       |
| Pituitary  | 922   | 2113  | 8     | 21    | 99.1        | 97.8          | 99.1          | 99.1       |

3.3. Training tools

The proposed model for brain tumor classification is trained on Intel (R) Core (TM) i3-4005U CPU @1.7GHz, RAM (4 GB), NVIDIA GeForce 920M GPU, NVIDIA CUDA 10.1.236, and Matlab 2018b.

3.4. Comparison with other classification models and discussion

For the purpose of comparison and to prove the superiority of our model over the rest of the models used similar images of brain tumor types, we used three cases for comparison: Firstly, we compare our model with the previous CNN models, and we see through compression that our model shows the best performance. Then we use two types of pre-trained models, ResNet-50 and AlexNet, and we see that the proposed model overcomes the previously trained models in the case of the data set that was used in this work. Table 4 shows the compression among the proposed model and the other models. Finally Figures 10 and 11 show the training progress and confusion matrix of the original dataset respectively. It is clear for us to see the large gap between results before and after date pre-processing and augmentation which is shown in Figure 8, Figure 9, Figure 10, and Figure 11. It is worth mentioning that; in the tumor localization step we have used the segmented image that is available with the data set, but in future, we can make a CAD to segment the image, detected the tumor, and finally classification.

Table 4. A comparison between the previous related works and the proposed our model.

| Model                                    | Accuracy % |
|------------------------------------------|------------|
| J. Cheng et al. [1]                      | 91.28      |
| Paul et al. [18]                         | 91.43      |
| Parnian Afshar et al. [19]               | 90.89      |
| Amin Kabir et al. [20]                   | 94.2       |
| Swati et al. [21]                       | 94.82      |
| Talo Muhammed, et al. [22]               | 95.23 ± 0.6|
| AlexNet                                  | 82.2       |
| Resnet-50                                | 75.6       |
| Proposed model                           | 96.1       |

Figure 10. The training process of the original data
In this research, a brain tumor classification model was proposed with magnetic resonance Imaging to classify the brain tumor into three types which are meningioma, glioma, and pituitary gland based on CNN. The proposed model consists of 28 layers starting with an input layer that takes the input images, 6 convolutional layers for feature extraction, a normalization layer for normalizing images, 6 ReLU layers function, 6 layers for max-pooling to reduce the dimensions of feature maps, 5 dropout layers to prevent from overfitting, a fully connected layer as a flattening layer, softmax layer to find for each class it’s probability and finally the classification layer to predict the output. Besides, data pre-processing and augmentation helped our model to show better accuracy and this has been illustrated in the paper above. Moreover, to prove the superiority of our model over the rest of the models, we presented a comparison among them. The accuracy of the proposed model is up to 96.1%.

REFERENCES
[1] J. Cheng, et al., “Enhanced performance of brain tumor classification via tumor region augmentation and partition,” PloS one, vol. 10, no. 12, Oct. 2015.
[2] Wen P. Y., Kesari S., “Malignant gliomas in adults,” N Engl J Med, vol. 359, no. 5, pp. 492-507, 2008.
[3] M. L. Goodenberger and R. B. Jenkins, “Genetics of adult glioma,” Cancer Genet, vol. 205, no. 12, pp. 613-621, Dec. 2012.
[4] Laws E. R., et al., “Pituitary disorders: diagnosis and management,” John Wiley & Sons, 2013.
[5] Black P. M., “Brain tumors,” N Engl J Med, vol. 324, no. 22, pp. 1555-1564, 1991.
[6] Aswathy, S. U., et al., “A survey on detection of brain tumor from MRI brain images,” 2014 international conference on control, Instrumentation, Communication and Computational Technologies (ICCICT), Jul. 2014.
[7] A. R. Abdulraaeq, et al., “An automated method for segmenting brain tumors on MRI images,” Biomedical Engineering, vol. 51, no. 2, pp. 97-101, Jul. 2017.
[8] R. Gurusamy and V. Subramaniam, “A machine learning approach for MRI brain tumor classification,” Comput., Mater. Continua, vol. 53, no. 2, pp. 91-108, 2017.
[9] Yazdani, et al., “Image segmentation methods and applications in MRI brain images,” IETE Technical Review, vol. 32, no. 6, 413-427, 2015.
[10] A. Ghosh, et al., “Fundamental concepts of convolutional neural network,” Recent Trends and Advances in Artificial Intelligence and Internet of Things, pp. 519-567, 2020.

Deep learning for glioma, meningioma, and pituitary classification (Toga A. Sadoon)
Said Charfi, Redouan Lahmyed, and Lalitha Rangarajan, “A novel approach for brain tumor detection using neural network,” *International Journal of Research in Engineering and Technology*, vol. 2, pp. 93-104, 2014.

J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabb, “Deep learning for brain tumor classification,” *Proc. SPIE, Med. Imag., Biomed. Appl. Mol., Struct., Funct. Imag.*, vol. 10137, Mar. 2017.

P. Afshar, *et al.*, “Capsule networks for brain tumor classification based on MRI images and course tumor boundaries,” 2018. [Online]. Available:https://arxiv.org/abs/1811.00597.

Amin Kabir Anaraki, Moosa Ayati, Foad 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.

Z. N. K Swati, *et al.*, “Brain tumor classification for MR images using transfer learning and fine-tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34-46, 2019.

M. Talo, *et al.*, “Convolutional neural networks for multi-class brain disease detection using MRI images,” *Computerized Medical Imaging and Graphics*, vol. 78, Dec. 2019.

Ali M. Hasan, *et al.*, “Combining deep and handcrafted image features for MRI brain scan classification,” *IEEE Access*, vol. 7, pp. 79959-79967, 2019.

N. Nabizadeh, “Automated brain lesion detection and segmentation using magnetic resonance images,” PhD, University of Miami, USA, 2015.

Y. LeCun, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

Y. LeCun, *et al.*, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *CoRR*, vol. abs/1502.03167, Feb. 2015.

D. Scherer, A. Müller, and S. Behnke, “Evaluation of pooling operations in convolutional architectures for object recognition,” *Artificial Neural Networks*, pp. 92 01, Jan. 2010.

J. Nagi, *et al.*, “Max-pooling convolutional neural networks for vision-based hand gesture recognition,” *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011.

Reza Bosagh Zadeh, Bharath Ramsundar, TensorFlow for Deep Learning, [Online]. Available: https://www.oreilly.com/library/view/tensorflow-for-deep/9781491980446/ch04.html.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, “Deep learning. Book in preparation for MIT Press,” 2016. [Online]. Available: http://www.deeplearningbook.org 1.

Hossam H. Sultan, Nancy M. Salem, Walid Al-Atabany, “Multi-classification of brain tumor images using deep neural network,” *IEEE Access*, 2019.