International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-3S, February 2019

Brain Tumor Image Classification and Grading Using Convolutional Neural Network and Particle Swarm Optimization Algorithm

N. Hema Rajini

Abstract--- Brain tumor defines the aggregation of abnormal cells in certain tissues of the brain area. The earlier identification of brain tumor plays a significant part in the treatment and recovery of the patient. The identification of a brain tumor and its grade is generally a difficult and time consuming task. For effective classification and grading of brain tumor images, in this paper, we present a convolutional neural network (CNN) and particle swarm optimization (PSO) algorithm of Glioma by the use of magnetic resonance imaging (MRI). The presented CNN-PSO model make use of PSO algorithm to select the deep neural network architecture which are generally depends on trial and error or by employed fixed structures. A detailed experimentation of the CNN-PSO method is carried out on several benchmark MRI brain images and verified its effectiveness on the applied test images with respect to different classification measures.

Keywords--- Brain Tumor, CNN, Deep Learning, Particle Swarm Optimization.

I. INTRODUCTION

Brain tumor is a kind of abnormal cell aggregation is certain brain regions. Based on the occurrence of brain tumor, it can be classified into primary and metastatic. The starting point of primary brain tumors is in the brain, whereas metastatic brain tumors startfrom other body parts. Tumors can be cancerous (or malignant) or noncancerous (or benign). Malignant brain tumor spread to other regions of the brain and spine rapidly. A fine detailed categorization divides the tumor into four grades and the higher grade tumors will be more malignant. Because of the existence of brain tumors in the center of the nervous system, benign tumors also might harm the brain and cause irrecoverable effects. In adults [1], the general kind of main brain tumor is assumed as Gliomas. From the grades of I to IV severity, gliomas are classified in order to grading system of World Health Organization (WHO) [2]. Cell in Grade I tumors are benign and they look more or less normal. Little abnormal looks are established by cells in Grade II tumors. Purely abnormal and malignant cells are comprised in Grade III tumors. The cell that is totally abnormal and fast-spreading is assumed as Grade IV. Glioblastoma multiforme (GBM) are typical kind of tumor that occurs from the tissue layers known as meninges which results in Meningioma tumors. The spinal cord and brain are covered by Meninges and play as protector. As it increasesin a slow manner, they are highly assumed as benign tumors and it has low probability to spread. The tumor that occurs at pituitary gland are known as Pituitary tumors and it is major cause of about 14% of intracerebral tumors, with few are because of defects in inherited genetics [3] with some are because of continuous mutation. It might provide severe health issues through the tumors and are assumed as benign because of its existence in sensible brain regions [4]. In recovery and treatment of the patient [5], earlier detection acts as a significant role. It is a time taking and complicated procedure to detect the brain tumor and its grade. Generally, while the brain tumor has developed adequately and different harassing symptoms have established, the patient will be referred to magnetic resonance imaging (MRI). When the presence of tumor is detected after analyzing the images of brain, the patient is subjected to brain biopsy. For a correct answer in biopsy, it may take to month unlike MR in few cases. Perfusion is the method that the MRI specialists performs for grade tumor and biopsy confirmation. With a view to brain tumors grading, except biopsy, there are few new techniques had been introduced in present times. Differentiating low-grade and high-grade glioma employing perfusion MRI has been capable of solving few disadvantages in biopsy. The computer-aided system implication is helpful for detection. In prior stages of tumor growth, an efficient and automatic system for classifying brain tumor supports physicians to interpret the medical images and aids in specialist's decision. Through spending reduced time, the grading of brain tumor is performed in this study and it provides enhanced accuracy. Additionally, the entire classification procedure is non-invasive. For analyzing the medical images, enough focus had been given for the purpose of diagnosing. The interest in the domain of health-related techniques and topics are emerging now as the presence of modern ML techniques have proven its efficacy in resolving different issues[6]. Several studies have been conducted in different tumor classification employing MRI, mainly evolutionary algorithms, MR brain images and artificial neural networks (ANN) [7], support vector machine (SVM) and hybrid intelligent techniques are the shallow ML algorithms that are used to differentiate the abnormal and normal classes of images in brain MR that are denoted through existing works.

With a view to classify different brain tumors kinds and Gliomas grades, SVM is examined in [8].

Manuscript published on 28 February 2019.
***** Correspondence Author (s)
N. Hema Rajini, Department of Computer Science and Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi, Tamilnadu, India. (e-mail: auhemasmith@yahoo.co.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license https://creativecommons.org/licenses/by-nc-nd/4.0/

The interesting region is primarily described and then few features like tumor shape are derived in the projected method from MR images. SVM with the elimination of recursive features had been employed with a view to choose the suitable feature. High accuracies are gained through the projected technique of binary classifications and result observation. However, multi-class classification accuracy is low in order to the given confusion matrix. Employing convolutional neural networks (CNN) and its associated models, the three tumor kind's classification performance is compared in [9]. Different CNN structure performances has been examined and relative shallow network with two max-pooling layers, two completely associated layers and two convolutional layers are employed for classification. In classification accuracy, the Vanilla preprocessing method is found to be efficient. CNN is employed in classifying unhealthy and healthy brain images as well as low-grade and high-grade Glioma tumors in [10]. As its network architecture, a well-known AlexNet's modified version is employed. High efforts are needed in modeling an automatic and real-time technique for brain MRI classification, even though, there are some valuable works being performed.

While resolving the difficult issues in machine learning, CNN comprises of several notable achievements and presently they are assumed as better technique for image processing [11]. Convolution operators are employed in large number of network layers in spite of matrix multiplication. This serves as a resolver of issues in convolutional networks with high computational cost. As the dataset in MRI-based diagnosis involves images of thousands various quality and kinds, this is highly significant. The automatic feature extraction is the other benefit of the method when comparing with shallow ML techniques. For deriving features, a technique was generally projected and to dimensionality reduction, a technique was employed to choose features that are dominant in traditional methods. At present timeZ, CNN had been employed extensively in medical images processing operations like brain tumor images skull stripping, grade classification [12] and segmentation [13-14].

In this paper, to perform Gliomas grades classification, a technique based on CNN has been projected for MR images. It is a difficult process to select a suitable architecture for deep neural network (DNN) for a certain work, that is always performed through using a general framework or through trial and error. Here, CNN framework is derived by the use of particle swarm optimization (PSO) algorithm. The network with various layer counts and parameters are examined through PSO and for additional processing, best performing network for the dataset was chosen. To validate the study, the projected technique is employed in two case studies. For establishing the projected technique strength in subsequent case, various kinds of tumors from other MRI database was employed as input towards CNN to verify the final diagnosis performance. With a view to support physicians in prior detection, the experimental outcome confirms that projected technique is suitable over various datasets of brain MRI.

The remainder of the paper is organized as follows. A detailed explanation of the presented work is given in Section 2 and the PSO based architecture selection is given in Section 3. The result analysis takes place in Section 4 and conclusion is provided in Section 5.

II. PROPOSED METHOD

This section describes about layers, parameters and structure of CNN. In artificial intelligence world, the method of Deep learning (DL) is a subgroup of ML methods. DL allows the system to characterize, generate and identify difficult concepts by employing easier concepts. At the same time, with multi-levels of abstraction [11], it allows multilayer models to learn data representations. In the domain of image processing, CNN are the DL based supervised methods that create notable enhancements. Commonly, the three major convolutional network layers are pooling, convolutional and fully connected. To convolute the images in input to produce different feature maps in convolutional layers, various kernels are employed by the network. The parameter count gets decreased with the use of this layer and correlation among the adjacent pixel is learnt through the network [14]. In each CNN, there exist two training phases. The input images are fed into the network in feedforward phase. At the same time, every neuron's input vector dot product and parameters vector is done. Then, the output will be estimated. With the expected output, the network output is compared through a loss function and the error rate is estimated depending on the error. By employing the chain rule, every parameter gradient estimation is performed in this phase and at the end the entire parameters are modified. For enough iteration counts, the process will be repeated. The brief explanations are provided below.

The volume of input is loaded with zeros with a view to manage the output size unchanged over the border. Same padding and valid padding are the two methods in padding. In this study, for entire convolutional layers, same padding is employed. An instance of implicating 3×3filter over a 4×4input matrix is demonstrated in fig. 1. The output matrix size still looks similar as 4×4 matrix, through inserting zeros.

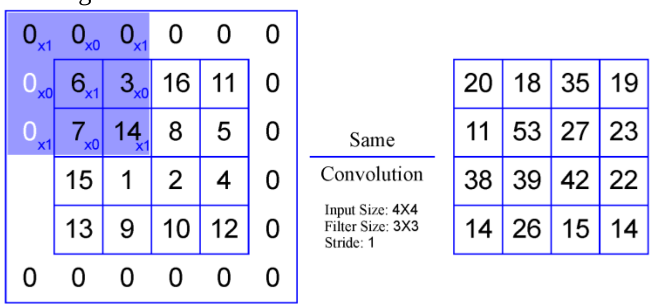

Fig. 1: Zero-padding method in CNN

2.1 Weight initialization

The network convergence can be speeded up through correct initial weights. Here, different methods for initializing weights might be introduced. After the investigation of employing various initializers, it is noted that the good performance is gained with normal distribution [15] through 'He' initializer.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

2.2. Activation function

Commonly after convolutions, an activation function or nonlinear operator is employed in deep networks. The existence of this function enhances the model. It is popular that implying rectified linear unit (ReLU) activation function in DNN enhances the speed of the training. ReLU processes the negative rates towards zero and is described in (1).

$$relu(x) = \begin{cases} x & if x \geq 0 \\ 0 & if x < 0 \end{cases} \quad (1)$$

When comparing with ReLU, Leaky ReLU has performed well as given in Eq. (2). While the function is not active, it enables little, non-zero gradient.

$$leaky_{relu(x)} = \begin{cases} x & if x \geq 0 \\ ax & if x < 0 \end{cases} \quad (2)$$

where$\alpha = 0.3$. Nowadays, exponential linear units (ELUs) tend to enhance the classification accuracy [16] and training speed. However, with the low computation cost, ELU gains negative rate enabling it to give mean unit activations nearby zero such as batch normalization.

$$elu(x) = \begin{cases} x & if x \geq 0 \\ a(e^x - 1 & if x < 0 \end{cases} \quad (3)$$

In scaled exponential linear unit (SELU) activation function [17] existence, evolving network performance has been analyzed [17] in Eq. (4). Through a small ELU twist, the SELU is presented here. The respective equation function are provided where $\lambda = 1.0507$ and $\alpha = 1.6732$.

$$selu(x) = \lambda \begin{cases} x & if x \geq 0 \\ ae^x - \alpha & if x < 0 \end{cases} \quad (4)$$

2.3. Pooling

After Convolutional layer, pooling layer generally exists, it decreases the parameter count and feature maps size that result in the reduction of computational cost. Pooling layers are subjected to little modifications because of the assumption of adjacent pixel estimations. Max-pooling is the extensively employed pooling techniques. After every convolutional layer, the max-pooling layer usually exists with 2×2 as filter size implied with 2 strides and takes around four at maximum.

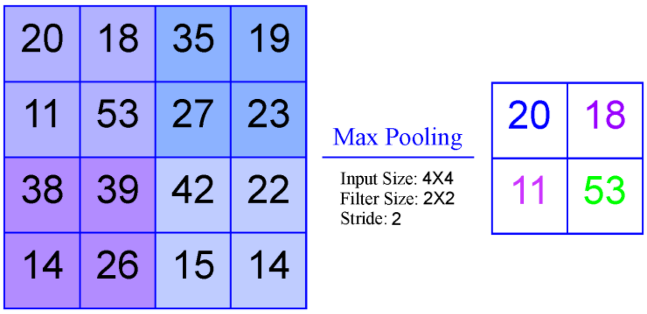

Fig. 2: Max-pooling layer

2.4. Regularization

The significant problem of ML is to produce a method which works well for also new entries not only with training data. For DL, various regularization methods had been projected. A dropout is employed that gives stigmatically low-cost, still an influential method for regularization. In the training level, it eliminates few nodes arbitrarily of the entire associated layer to avoid overfitting. Dropout is assumed as an ensemble method, at the same time, as it gives various networks while training.

2.5. Loss function

The loss function that has to be chosen to be reduced is the significant modeling aspect of a DNN. Categorical cross-entropy function (H) is commonly a better candidate and it had been widely employed. Over discrete variable x, it is described for two distributions (p and q) and is expressed as

$$H(p,q) = -\sum_x p(x) \ln(q(x)) \quad (5)$$

where p(x) and q(x) refers to the estimate for true distribution.

2.6. Training

The loss function has to be reduced through the method of gradient-based optimization with a view of DNN training. In DL, Stochastic gradient descent (SGD) is employed extensively as an optimizer.

Adaptive moment estimation (Adam) technique is employed for stochastic optimization. When compared to customary optimization methods, Adam performs well. Additionally, with the massive dataset, its computational efficiency is an advantage.

But, Adam algorithm estimates adaptive learning rates through calculating second moment and first moment of gradients for modifying the weights and the learning rate will be constant. In this study, the optimizers such as Adamax, Adagrad, Nadam and Adadelta were analyzed. The learning rate adjusts the parameters in Adagrad optimizer. When comparing with frequent parameters, it occurs by updating for infrequent parameters.

Adadelta constraints the window size of existing accumulated gradients, not like Adagrad that acquires entire previous squared gradients.

Depending on the infinity norm, an Adam optimizer variant is Adamax. Gradient optimizers are accelerated by Nadam through combining Nesterov and Adam. The momentum step parameters are modified prior to gradient computation in Nadam and create it probable to take accurate gradient direction steps. The values of all these optimizers' parameters except learning rate are chosen by the author.

III. NETWORK ARCHITECTURE DESIGN

Generally, a desirable network architecture is identified by performing testing of different common network structures.

This procedure needs large amount of trial and error, as well as high computational cost. Here, different CNN models for the process of MRI image classification are developed by the use of PSO algorithm. Rather than training and comparing more than one million diverse models, by employing PSO and comparing below 500 architectures an appropriate model is identified and the computation complexity is minimized. The process of choosing the network model is explained below.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

Table 1: Parameters involved in CNN selection process

Parameter	Value
No.of convolutional + max pooling layers	2-6
No. of fully connected + dropout layers	1-3
No. of filters	16, 24, 32, 48, 64, 96, 128
Kernel size	2-7
Number of fully connected neurons	128, 192, 256, 384, 512
Dropout rate	0.1-0.5

Particle swarm optimization algorithm

The process of natural selection is inspired from PSO algorithm. In each round, every particle's movement is controlled by its local best known position; however, it is also guided toward the best known positions in the search-space, which are updated as better positions by other particles. This is expected to move the swarm toward the best solutions. In this work, PSO algorithm is used to develop the best structure of the CNN by selecting appropriate variables for the network. These variables are number of convolutional and max-pooling layers, number of filters and size, number of fully connected layers, activation function, dropout probability, optimization method and learning rate. The values integrated with the variables are given in Table 1. Using the variables present in the table, numerous architectures are possible for the CNN. Instead of directly finding the probable architectures, PSO algorithm is used to identify the better one. The involved processes are displayed in Fig. 3. At the beginning, 50 networks with arbitrary variables are generated as initial particles. Every individual network has undergone training by 80% of data and testing by 20% of data. The classification accuracy is assumed as a will be trained by 80 percent of data and 20 percent of data are used as validation dataset. The accuracy is used as a parameter to retain or reject the network in the subsequent generation.

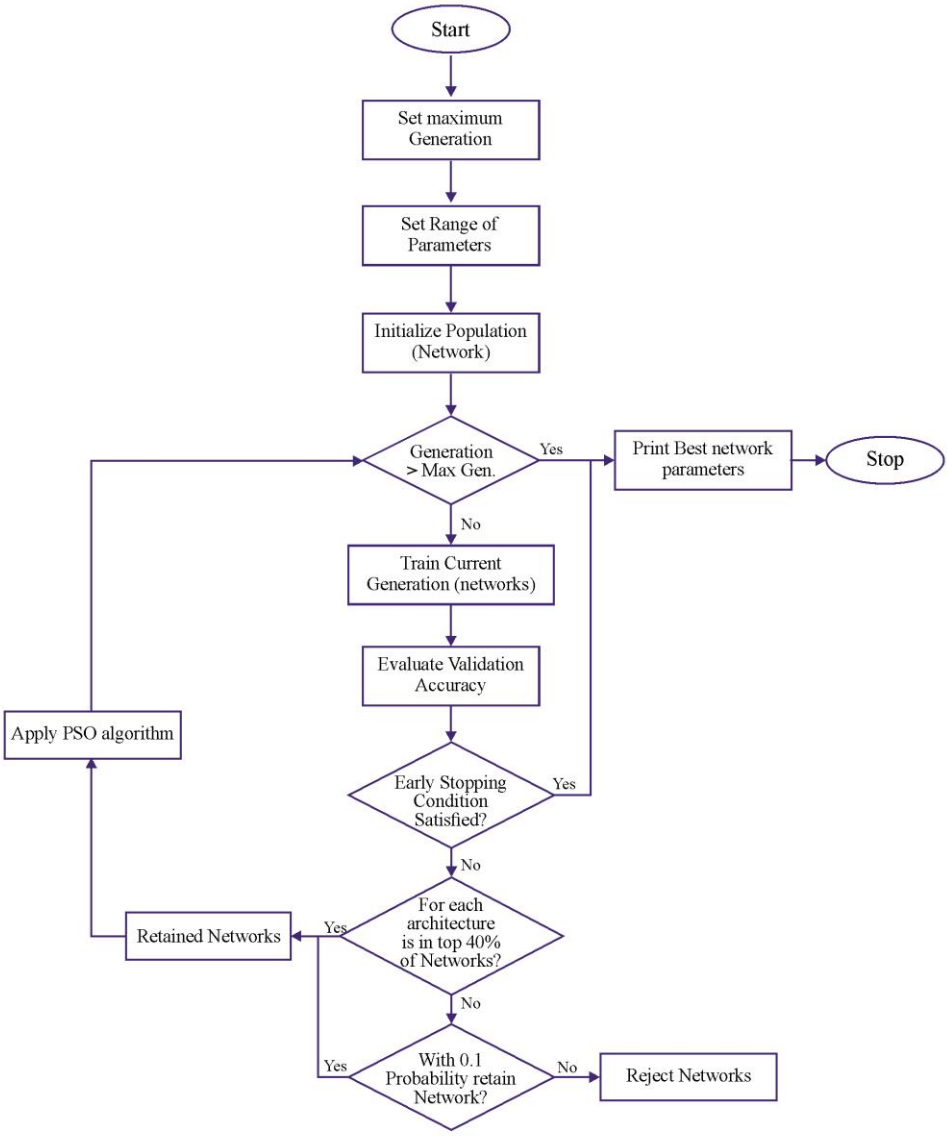

Fig. 3: Overall process of the CNN-PSO method

© Copyright: All rights reserved.

Using the presented method, PSO is terminated once the termination condition is fulfilled. The early termination condition is when no enhancement happens in the validation accuracy (loss function) of three sequent epochs. The early stopping condition is simulated for reducing the computational cost. This procedure is continued till it is terminated, and finally, the network architecture that has the best performance is chosen as the main network architecture for the classification.

IV. PERFORMANCE EVALUATION & RESULT

Dataset

The dataset applied in this study are accessed from 4 online databases. The normal MRI are attained from the brain development website (IXI dataset) [18] which comprises of almost 600 MRI from normal subjects (without any lesion). MR images of Glioma tumors are gathered from the cancer imaging archive databases. REMBRANDT dataset holds the pre-surgical magnetic resonance multi-sequence images from 130 patients that suffer from low or high grade Gliomas. TCGA-GBM data collection has glioblastoma multiform brain MRI of around 200 patients and TCGA-LGG dataset contains low grade Gliomas data, gathered from 299 patients. The data gathered from the above said databases are assumed as Case 1. Beside, the axial brain tumor images of Cheng et al. [19] are applied which contains a MRI with T1- weighted images from 233 patients with Meningioma, Glioma, and Pituitary brain tumor types. These images are assumed as Case 2. Fig. 4 shows the brain MRI of the normal person.

Fig. 5 shows the MRI of 3 variant grades of Glioma tumor with gadolinium injection. And, Fig. 6 displays the instances of the three brain tumors images from Case II.

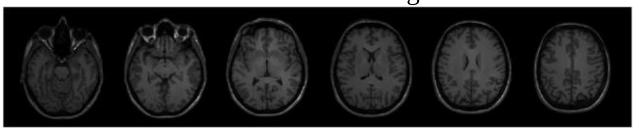

Fig. 4: Brain MRI of normal person

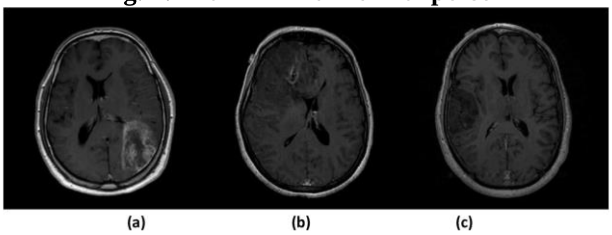

Fig. 5: MRI of 3 variant grades of Glioma tumor

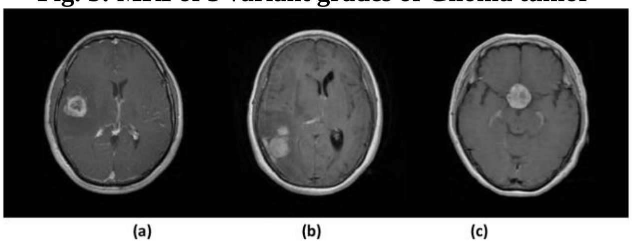

Fig. 6: Brain tumors images from Case II

Results analysis

The results of the employed classifier can be validated in numerous forms. In this study, confusion matrix is employed for checking the classifier results that provides useful data related to the real and predicted labels offered by the presented CNN-PSO model. By employing this information, the results can be validated in various dimensions. The confusion matrix of the two cases I and II are provided in Table 2. Using the computation of diverse quantities and ratios, the results of the CNN-PSO method is determined during the classification of brain MRI.

From Table 2, it is shown that the CNN-PSO method has properly identified the tumors with high precision. In addition, for precise evaluation, different criteria have been investigated. In Case I, Normal images are efficiently classified with the precision of 99.60%, recall of 99.80%, F-score of 99.70%, accuracy of 99.85%, AUC of 99.83% and kappa value of 99.60%.

Table 2: Confusion matrix obtained by CNN-PSO

Dataset	TP	FP	TN	FN
Case 1				
Normal	503	2	1494	1
Grade II	446	28	1476	50
Grade III	440	28	1456	76
Grade IV	490	63	1437	10
Case 2				
Glioma	117	5	221	1
Meningioma	104	2	232	7
Pituitary	114	2	228	1

TP-True Positive, TN-True Negative, FP-False Positive, FN-False Negative

Besides, Grade II are the frequently occurring malignant brain tumors which are effectively classified with the precision of 94.09%, recall of 89.02%, F-score of 91.96%, accuracy of 96.10%, AUC of 94.02% and kappa value of 89.39%. Similarly, for the grade III images, images are properly classified with the precision of 94.01%, recall of 85.27%, F-score of 89.43%, accuracy of 94.80%, AUC of 91.69% and kappa value of 85.99%. In the same way, for grade IV images, effective prediction is carried out with the precision of 88.60%, recall of 98%, F-score of 93.07%, accuracy of 96.35%, AUC of 96.90% and kappa value of 90.59%.

Table 3: Evaluation of Different Measures with Proposed Method

Dataset	Precision	Recall	F-Score	Accuracy	AUC	KAPPA
Case Study 1						
Normal	99.60	99.80	99.70	99.85	99.83	99.60
Grade II	94.09	89.02	91.96	96.10	94.02	89.39
Grade III	94.01	85.27	89.43	94.80	91.69	85.99
Grade IV	88.60	98.00	93.07	96.35	96.90	90.59
Case Study 2						
Glioma	95.12	99.15	97.09	97.97	98.25	95.54
Meningioma	98.11	93.69	95.85	97.39	96.42	93.95
Pituitary	98.27	99.13	98.70	99.13	99.13	98.05

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

On the case II database, the presented CNN-PSO algorithm is found be effective with maximum classifier results. For the Glioma type, the CNN-PSO shows better results with the precision of 95.12%, recall of 99.15%, F-score of 97.09%, accuracy of 97.97%, AUC of 98.25% and kappa value of 95.54% respectively. For the pituitary type, maximum results are identified with precision of 98.27%, recall of 99.13%, F-score of 98.70%, accuracy of 99.13%, AUC of 99.13% and kappa value of 98.05% respectively.

Table 4: Comparison with Existing Methods

Case Study 1: Normal/ Grade II/ Grade III/ Grade IV	
Methods	**Accuracy**
Proposed	96.77
CNN + GA [20]	90.90
SVM + RFE [8]	62.50
Case Study 2: Glioma/ Meningioma/ Pituitary	
Proposed	98.16
CNN + GA [20]	94.20
Vanilla Pre-processing + shallow CNN [9]	91.43

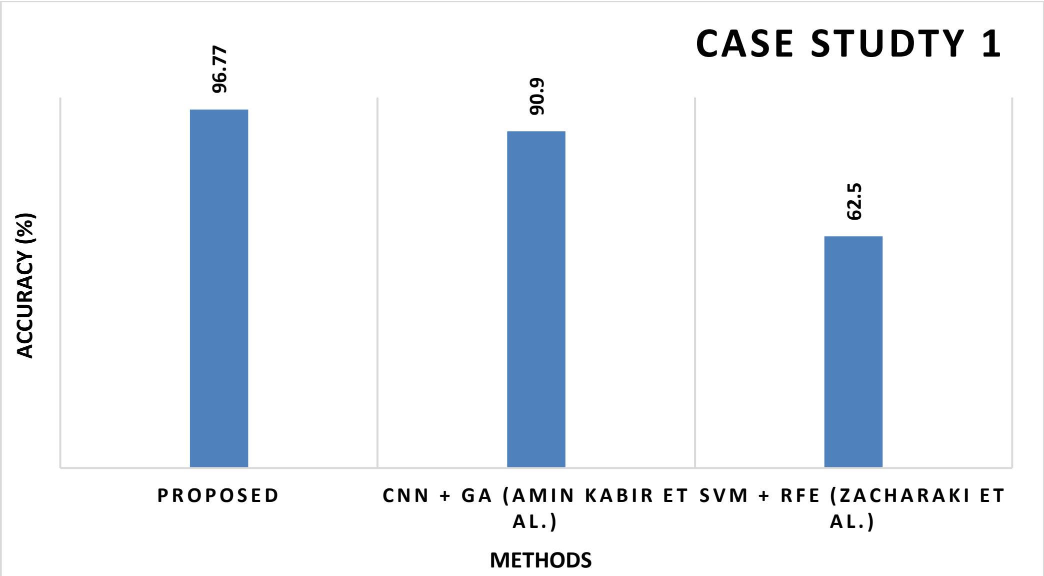

Fig. 7: Comparative analysis of Case I images interms of accuracy

After seeing the classifier performance, the effectiveness of the CNN-PSO method for classifying various MRI are verified. A comparative analysis with other methods are also made and provided in Table 4. Fig. 7 and Fig. 8, Fig. 7 and Fig. 8 show the comparative results analysis of various methods. From the figures, it is clear that the presented method attains a maximum classification accuracy of 96.77 on the case I database whereas the existing method exhibits lower performance of around 62% accuracy. Similarly, Fig. 8 also provides the enhanced results of the presented method over the compared methods. In overall, the CNN-PSO model shows effective results on the classification of brain MR images under several measures.

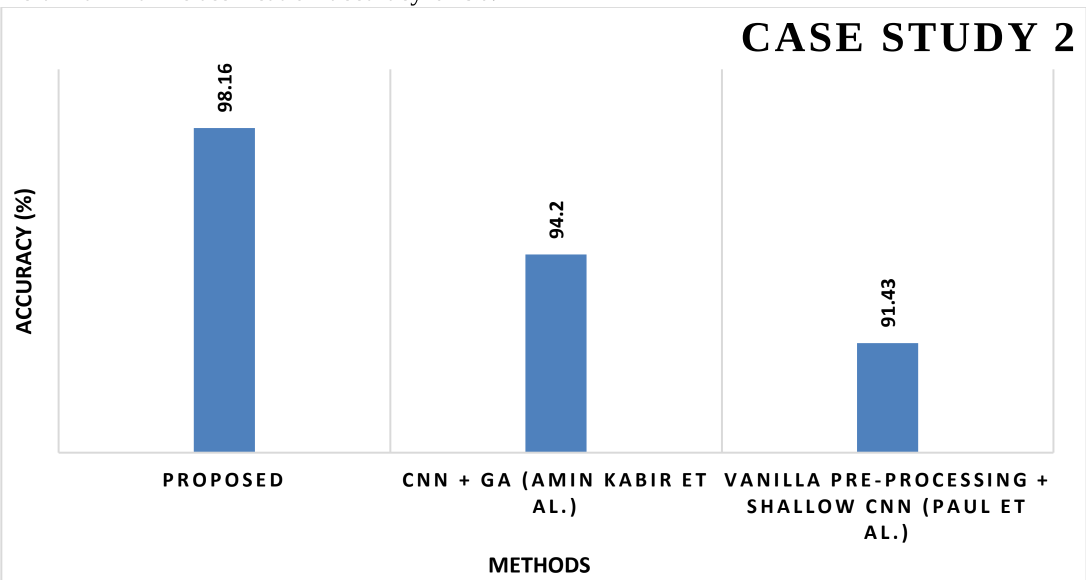

Fig. 8: Comparative analysis of Case II images interms of accuracy

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

V. CONCLUSION

With a view to classify different brain tumors kinds and Gliomas grades, in this paper, we have introduced a new classification model using CNN and PSO algorithm. In this paper, CNN framework is derived by the use of PSO algorithm. The network with various layer counts and parameters are examined through PSO and for additional processing, best performing network for the dataset was chosen. To validate the study, the data gathered from the various databases are considered under two cases I and II. Based on the detailed experimental analysis, it is confirmed that the CNN-PSO is the appropriate choice for brain MRI classification.

REFERENCES

1. Wen PY, Kesari S. Malignant gliomas in adults. N Engl J Med 2008;359(5):492–507.
2. Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 2016;131 (6):803–20.
3. Laws ER, Ezzat S, Asa SL, Rio LM, Michel L, Knutzen R. Pituitary disorders: diagnosis and management. John Wiley & Sons; 2013.
4. Black PM. Brain tumors. N Engl J Med 1991;324(22):1555–64
5. Kelly PJ. Gliomas: survival, origin and early detection. Surg Neurol Int 2010;1.
6. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017.
7. Lo C-S, Wang C-M. Support vector machine for breast MR image classification. Comput Math Appl 2012;64(5):1153–62.
8. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. MagnReson Med 2009;62(6):1609–18.
9. Paul JS, Plassard AJ, Landman BA, Fabbri D. Deep learning for brain tumor classification. Proc of SPIE. 2016. pp. 1013710–1.
10. Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS. Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Appl Sci 2017;8(1):27.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
12. Mohan G, Subashini MM. MRI based medical image analysis: survey on brain tumor grade classification. Biomed Signal Process Control 2018;39:139–61.
13. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 2016;35(5):1240–51.
14. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw 1995;3361(10):1995
15. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1026–34
16. Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus); 2015, arXiv preprint arXiv:1511.07289
17. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Selfnormalizing neural networks; 2017, arXiv preprint arXiv:1706.02515.
18. IXI Dataset. Available from: http://brain-development.org/ixi-dataset/.
19. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE 2015;10(10). e0140381
20. Anaraki, A.K., Ayati, M. and Kazemi, F., 2019. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. Biocybernetics and Biomedical Engineering, 39(1), pp.63-74.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.