A Deep Belief Network Based Brain Tumor Detection in MRI Images

Thahseen P[1], Anish Kumar B[2]

[1]MEA Engineering College, State Highway 39, Nellikunnu-Vengoor, Perinthalmanna, Malappuram, Kerala

[2]Assistant Professor, MEA Engineering College, State Highway 39, Nellikunnu-Vengoor, Perinthalmanna, Malappuram, Kerala

Abstract: *Glioma is a common and malignant tumor, which may lead to short life span in their highest grade. Thus to improve the quality of life of cancer patients is the early diagnosis of brain tumor, which is a stage of treatment. MRI (Magnetic Resonance Imaging) is widely used medical imaging technique used to assess tumors, but large amount of data produced by MRI may vary greatly. Thus manual detection will be a challenge. To detect brain tumor in magnetic resonance imaging many automated diagnostic systems play an important role. The system may mainly include three steps namely preprocessing, classification and post processing. A DBN (Deep Belief Network) based classification method is used to identify brain tumor in MRI images which can yield the result more accurately.*

Keywords: Glioma, Oncological, Magnetic Resonance Imaging, Deep Belief Network

1. Introduction

Glioma is a common and malignant tumor, which may lead to short life span in their highest grade. Based on the type and location of tumor the symptoms may vary. Along with tumor fatigue, fever, loss of appetite, sweating, loss of body weight is the other symptoms of brain tumor.

The majority of research in developed countries shows that the number of people having brain tumors was killed due to the fact of inaccurate detection. Among cancer brain tumor is the most devastating and aggressive type of cancer and therefore its correct and early identification will be a challenge. Brain tumor is a major type of cancer. Abnormal growth of cells in brain may lead to brain tumor. If the growth and death of brain cells is not balanced, brain tumor may occur. These tumors may be broadly

classified into two namely malignant (cancerous) or benign (non-cancerous). Glioma is a common and malignant tumor, which may lead to short life span in their highest grade. Based on the type and location of tumor the symptoms may vary. Along with tumor fatigue, fever, loss of appetite, sweating, loss of body weight is the other symptoms of brain tumor.

Gliomas can be broadly classified into LGG (Low Grade Gliomas) and HGG (Highest Grade Gliomas) where the former is less aggressive [1], [2]. After diagnosis the patients may not survive on an average of 14 months even under the treatment [3]. Therefore to improve the quality of life of oncological patients is the early diagnosis of brain tumor, which is a major step of treatment. Various techniques are existing to detect cancer but most of them are in advanced stage, so the chance of recovery of the patient will be less.
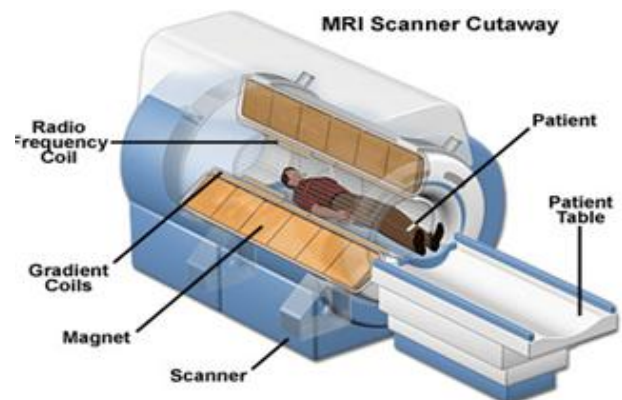

Figure 1: MRI Scanning [2]

Due to overlapped structure of cancer cells the early detection of tumor detection is challenging. Surgery, chemotherapy, radiotherapy, or combination of them is the treatments used nowadays to cure brain tumor in their advanced stage.
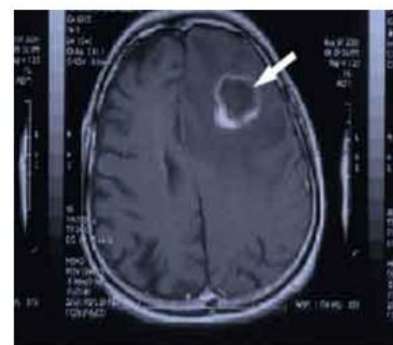

Figure 2: MRI image of tumor affected brain

MRI (Magnetic Resonance Imaging) is widely used medical imaging technique used to assess tumors, but large amount of data produced by MRI may vary greatly. Thus manual detection will be challenging. Segmentation and classification of tumor in MRI images are done in order to find whether tumor affected or not and segment the area affected. The prescriptive methods used for the detection of

brain tumor in their early stage are biopsy, manual inspection, expert opinion and etc. The biopsy method may take ten to fifteen days whereas human prediction will not be accurate always to give a result about tumor. Detection of tumor is more accurate than manual detection using image processing techniques. The
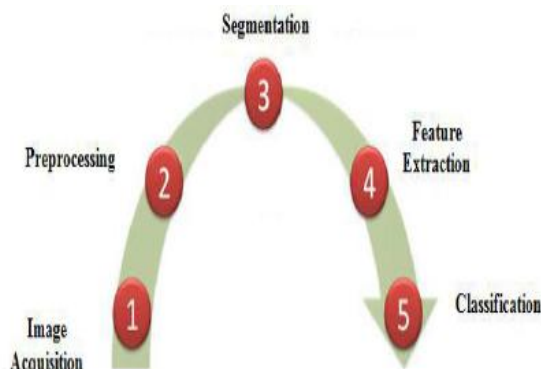
Figure 3: Major steps of image processing techniques [10]

The major benefit of using image processing technique is the time for detection will comparatively much lesser than manual detection. The result of such image processing technique is the tumor is detected and the correct position of the tumor is determined. the fig 1.3 shows the basic steps of image processing techniques to detect brain tumor. There also exist many automated diagnostic systems which plays a major role in detection of brain tumors in MR images. To segment and classify the features in MR images many classification techniques are used such as SVM (Support Vector Machine), k-NN (k- Nearest Neighbor), NN (Neural Networks), DBN (Deep Belief Network) etc. Before applying classification techniques, preprocessing techniques should be applied. Preprocessing of images is done in order to reduce the noise and thereby enhances the MR images.

2. Literature Survey

Comparing different classification methods used for brain tumor detection in MRI images.

1) SVM (Support Vector Machine)
Sahoo, Laxmipriya, et al. [4] Support vector machine is an algorithm that constructs a hyper-plane to analyze data and classify objects. SVM is also be used for regression. Fig. 4. Show the SVM analysis. SVMs are widely used in machine learning. They are also known as support vector networks. The associated algorithms analyze the data and recognize the patterns that are used for analysis. Classification and regression analysis. If a set of training examples were given, each example marked as belonging to one of two categories, an SVM training algorithm constructs a model that assigns new examples into one category or the other [5], making it a non-probabilistic binary linear classifier [4].
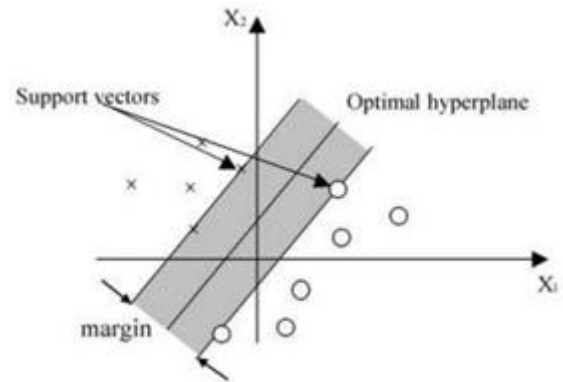
Figure 4: SVM Analysis

The major disadvantages of SVM are it is not easy to determine the optimal parameters when there is nonlinearly separable data is present[7] and the method is likely to give poor performance, if the number of features is much less than the number of samples[6].

2) k-NN (k- Nearest Neighbor):
Arriaga-Gomez and Miguel F et al [9] k-NN is one of the major distance-based algorithms; where given k as a positive integer and a sample feature vector (sample template), the k training features with the smallest distance to the sample is selected. The sample is identified as the most repeated among the selected k feature vector.

Ramteke R. J. and Y. Khachane Monali et al [10] proposed automatic classification method of medical images into Normal and Abnormal based on automatic abnormality detection and image features. The major advantages of k-NN are simpler to implement and understand. The result will not be accurate always as it determines its class assignment by either getting a majority vote for them or averaging the class numbers of nearest k points.

3) CNN (Convolutional Neural Network)
S. Pereira, A. Pinto, V. Alves and C. A. Silva et. al [8] says that to achieve some breakthrough results and win well- known contests CNN were used. Convolving an image or a signal with kernels to obtain feature maps is the main application of CNN. In paper [8] single layered CNN is used so that the features used for classification is less when compared to deep networks.
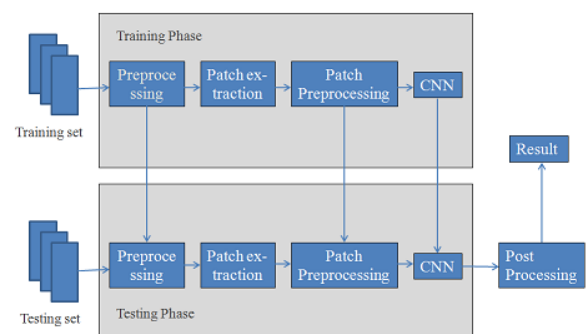
Figure 5: CNN Architecture

4) DT (Decision Tree)
Decision tree is a tree-like graph of decisions. Each branch represents the decisions to make graphically. It is a nonparametric, supervised approach. DT calculates the class

membership by repeatedly dividing one data set uniform sub-sets. Hierarchical classification permits acceptances and the rejection of class labels in each intermediate stage. This method consist of 3 parts: partition of nodes, look for the terminals and class assignment label the terminal nodes. DT is easy to interpret and explain but its classification error rate is high.

5) DBN (Deep Belief Network)

Das, Deepjoy, and Alok Chakrabarty et al [13] proposed an automated classification method which mainly focuses on recognition of facial emotion using a dataset comprises of four kinds of emotions (happy, angry, neutral And sad) with different models of deep neural networks and compare its performance.

3. Methodology

1) Preprocessing

Preprocessing of MR images is the primary step of brain tumor detection. Many preprocessing techniques exist which may include intensity normalization, filtering, histogram equalization etc. Preprocessing of images is done in order to reduce the noise and thereby enhances the MR images. The main purpose of these methods is mainly to improve the quality of image thereby, improves image to get more surety and ease in detecting the tumor. In the proposed technique intensity normalization is applied on each sequence to make the contrast and intensity ranges more similar across patients and acquisitions. We compute the mean intensity value and standard deviation to normalize the MRI images across all training patches extracted for each sequences. On each sequence, we normalize the patches to have zero mean and unit variance.

2) CNN- The Existing System

S. Pereira, A. Pinto, V. Alves and C. A. Silva et. al says that to achieve some breakthrough results and win well-known contests CNN were used [8]. Convolving an image or a signal with kernels to obtain feature maps is the main application of CNN. Through the weights of the kernels, a unit in a feature map is connected to the previous layer. By back propagation, the weights of the kernels are adapted during the training phase, in order to enhance the characteristics of the input. Convolutional layers have fewer weights to train than dense FC layers, making CNN easier to train and less prone to over fitting, since the kernels are shared among all units of the same feature maps. Since the same kernel is convolved over the entire image, independently the same feature is detected i.e., translation invariance. Information of the neighborhood is taken into account, by using kernels, which is an useful source of context information. On the output of each neural unit, usually a non linear activation function is used.
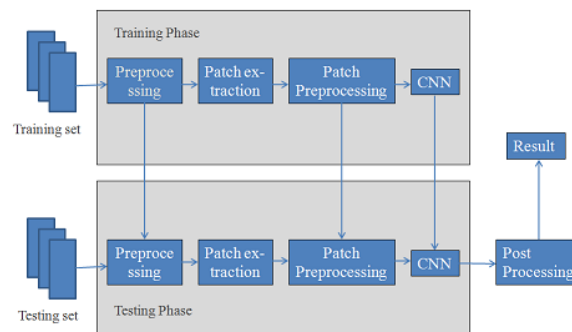
Figure 6: CNN Architecture

The extracted feature becomes more abstract with the increasing depth when we stack several convolutional layers. The first layer helps to enhance features such as edges, which are grouped in the following layers as motifs, parts, or objects.

In the context of CNN the following concepts are important:-

a) *Initialization*:- We use the Xavier initialization [8] to achieve convergence. The activations and gradients are maintained with this; otherwise back-propagated gradients could vanish or explode.

b) *Activation function*:- it is responsible for the transformation of data non-linearly. Rectifier linear units (ReLU) is defined as:-

$$f(x)=max(0,x),\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

to achieve better results than classical sigmoid, or hyperbolic tangent function were found. However, we can the gradient flowing and consequent adjustment of the weights by imposing a constant θ can be impaired. We cope with these limitations using leaky rectifier linear unit (LReLU), a variant on the negative part of the function that introduces a small slope. This function is defined as:-

$$f(x) = max(0,x)+\alpha\,min(0,x),\ldots.(2)$$

where α implies leakyness parameter. We use softmax, in the last FC layer.

c) *Pooling:*- spatially nearby features in the feature maps were combined. This combination of possibly redundant features makes the representation more compact and invariant to small image changes, such as insignificant details; it also decreases the computational load of the next stages. It is more common to use max-pooling or average-pooling [8] to join features.

d) *Regularization:*- it reduces overfitting. In the FC layers, we use Dropout. It removes nodes from the network with probability p in each training step. It forces all nodes of the FC layers to learn better representations of the data, preventing nodes from co-adapting to each other is done in this way. All nodes are used at test time. Since each network is trained with a portion of the training data, dropout can be seen as an ensemble of different networks and a form of bagging.

e) *Data Augmentation:*- to increase the size of training set and reduces overfitting we use data augmentation. We restricted the data augmentation to rotating operations, since the class of patch is obtained by central voxel.

f) *Loss function:*- during training, it is the function to be minimized.
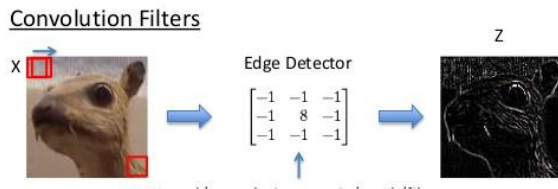
Figure 7: Convolution Filters

3) DBN - The Proposed System

Deep Belief Network was introduced by Professor Geoffrey Hinton in order to overcome the limitations of earlier neural networks. A deep belief network is a generative graphic model, or even a network of neurons deep in machine learning type, consist of several layers of latent variables, connections within the layers, but not within the units, between each layer [14].
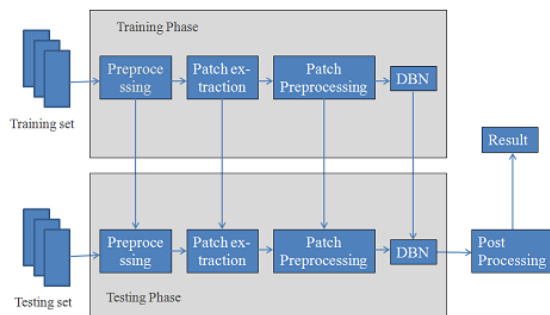
Figure 8: DBN Architecture

A DBN can learn to probabilistically reconstruct its inputs when trained on a set of examples in an unsupervised way. The layers then act as feature detectors on inputs. After this learning step, to perform classification, a DBN can be further trained in a supervised way.

DBNs can be considered as a simple composition of unsupervised network, such as restricted Boltzmann machines (RBMs) or auto encoders, where each sub layer hidden network serves as the visible layer for the next. This also leads to a quick, layer by layer, the unsupervised training procedure where the contrastive divergence is applied to each subnet in turn, starting with the 'down' pair of layers (the lowest visible layer being a training set).
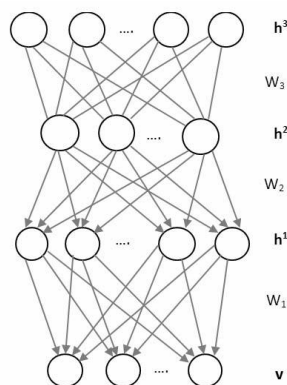
Figure 9: Illustration of a DBN with three hidden layers [15]

Restricted Boltzmann Machines and Belief Networks are the two different types of Deep Belief Networks.

a) Belief Networks

A basic belief network comprised of stochastic binary units layers with weighted connections. This network is acyclic graph which allows us to note that at the leaf node what type of data the belief network believes.

The goal of a belief network is to infer the states of the unobserved stochastic binary units. And it can adjust the weights between these units. Due to this the network can generate similar to the observed data. The stochastic binary units in belief networks consists of state of 0 or 1 .and the probability of becoming 1 is determined by a bias and weighted input from other units.

In belief networks, it is difficult to obtain the posterior distribution. This is called 'explaining away' issue. This tells that, if there is an effect of both units gets influence, then two independent hidden units can become dependent, which is also called conditional dependence. This means, the occurrence of either of hidden units can explain away the occurrence of the unit these are connected from both hidden units. If belief networks have multi-layered neural networks, then the posterior distribution depends on the prior, and likelihood of upper hidden layers. And there are number of ways of possible configurations among these layers.

Therefore, Hinton proposes an idea of learning one layer at a time, and restricting the connectivity of stochastic binary units, to make the learning efficient and simple.

b) Restricted Boltzmann Machines

Restricted Boltzmann Machines Boltzmann Machine is a stochastic recurrent neural network. RBM is a stochastic neural network which consists of stochastic binary units and undirected edges between units. The RBM has a scalability issue. So that the learning for this machine is impractical. The RBM has one layer of hidden units. And restricts connections between hidden units. Therefore RBM allows efficient learning algorithms.
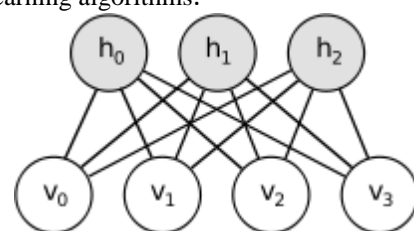
Figure 10: RBM Structure [14]

The probability distributions over hidden and/or visible units are in terms of the energy function on these configurations:-

$$P(v, h) = 1/Z \, exp \, (-E \, (v, h))………(3)$$

Where Z:

$$Z = \Sigma_{v, h} \, exp \, (-E \, (v, h))$$

Contrastive divergence algorithm is used to fasten the learning for a RBM. Updating all the hidden units in parallel starting with visible units, reconstruct visible from the hidden units, and finally update the hidden units again, is the general idea.

Deep Belief Networks (DBN) is a multi-layer belief network, in which each layer is Restricted Boltzmann Machine. To construct DBN they are stacked to each other. For training DBN first, learn a layer features from the visible units using Contrastive Divergence (CD) algorithm. Then treat the activation of previously trained as visible units and learn features of features in a second hidden layer. When the learning for the final hidden layer is achieved the whole DBN is finally trained.

Implementation
We choose to implement DBN in MATLAB, since it can easily represent visible layers, hidden layers and weights as matrices and execute algorithms efficiently. In addition, we choose the MNIST dataset to perform calculations so that we can compare the performance of different classifiers.

Since DBN consist of many layers of RBMs, implementing a DBN includes training each layer of RBM. We first randomly initialize the units and parameters for training RBM. Two phases-positive and negative are there in Contrastive Divergence algorithm. The binary states of the hidden units are determined by calculating the probabilities of weights and visible units during the positive phase. It is called positive phase so, because it increases the probability of training data. On the other hand, the probability of samples generated by the model is decreased by the negative phase. A complete positive-negative phase indicates in epoch and at the end of iteration the error between generated samples by the model and actual data vector is calculated. Finally, the derivative of the probability of visible units with respect to weights is taken, which is the expectation of the difference between positive phase contribution and negative phase contribution to update the weights. After the error calculation, the full update rule with momentum and learning rate is implemented in MATLAB. To train the whole DBN, the following algorithm is a greedy learning algorithm.

Next is training DBN using greedy learning which trains one RBM at a time and continues until the last RBM. And the visible layers has undirected connections with the hidden layer. In MATLAB, by simply initializing parameters with previously obtained parameters we can implement greedy algorithm. Then we again call the algorithm to train next RBM layer.

4) Post Processing
After performing the CNN or DBN classification method in post processing technique we are removing the unwanted clusters from the resultant image.

4. Experimental Results

The DBN based brain tumor detection algorithm is successfully implemented and applied on dicom images collected from hospitals. CNN already exists and since it is single layered CNN the result will not be always accurate (as shown in below screenshot).
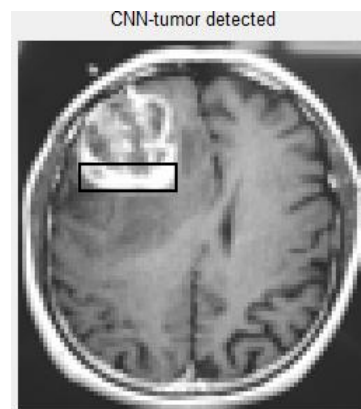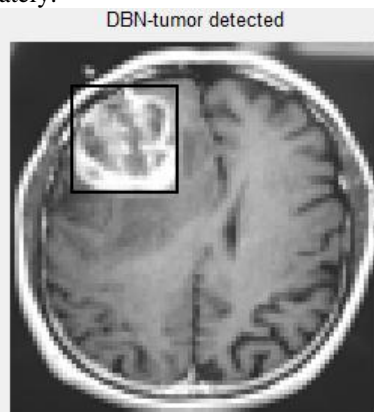

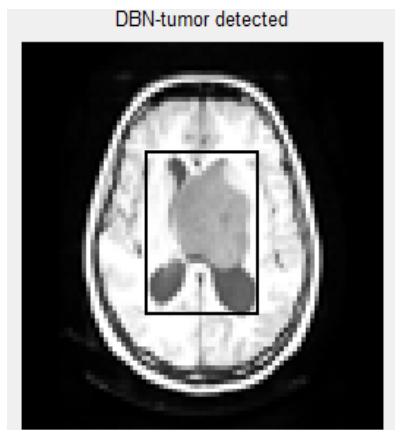**Figure 11:** Tumor detected using CNN in an MRI image
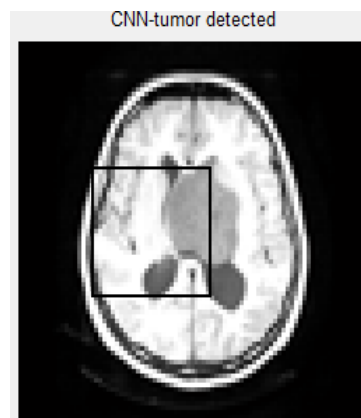
Figure 12: Tumor detection using CNN in a dicom MRI image

Since DBN is a type of deep learning method brain tumor detection is comparatively more accurate. Below figures shows DBN based detection algorithm detects the affected region accurately.

Figure 13: Tumor detected using DBN in an MRI image

Figure 14: Tumor detected using DBN in dicom MRI image

5. Conclusion

Convolutional neural networks (CNNs) are widely used in pattern- and image-recognition problems as they have a number of advantages compared to other techniques. A CNN is a special case of the neural network. CNN may includes one or more convolutional layers consist of numerous neurons, often with a sub sampling layer, which are then followed by one or more fully connected layers as in a standard neural network.

A deep belief network is a generative graphic model, or even a network of neurons deep in machine learning type, consist of several layers of latent variables, connections within the layers, but not within the units, between each layer. A DBN can learn to probabilistically reconstruct its inputs when trained on a set of examples in an unsupervised way. The layers then act as feature detectors on inputs. After this learning step, to perform classification, a DBN can be further trained in a supervised way.

The DBN based brain tumor detection algorithm is successfully implemented and applied on dicom images collected from hospitals. Since DBN is a type of deep learning method brain tumor detection is comparatively more accurate.

References

[1] Bauer, Stefan, "A survey of MRI-based medical image analysis for brain tumor studies." Physics in medicine and biology 58.13 (2013): R97.

[2] D. N. Louis et al., "The 2007 who classification of tumors of the central nervous system," Acta neuropathologica, vol. 114, no. 2, pp. 97–109, 2007.

[3] Van Meir, Erwin G., "Exciting new advances in neyro-oncology: the advenue to a cure for malignant glioma." CA: a cancer journal for clinicians 60.3(2010): 166-193.

[4] Sahoo, Laxmipriya, "Alternate machine validation of early brain tumor detection."Information Communication and Embedded Systems (ICICES) 2016 International Conference on IEEE 2016.

[5] Singh, Amritpal. et al "Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM." Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on. IEEE, 2015.

[6] Mr. Ajaj Khan, Ms. Nikhat Ali Syed , et al. "Image Processing Techniques for Automatic Detection of Tumor in Human Brain Using SVM." International Journal of Advanced Research in Computer and Communication Engineering , vol. 4, Issue 4, April 2015.

[7] Kamavisdar, Pooja, Sonam Saluja, and Sonu Agrawal. "A survey on image classification approaches and techniques." International Journal of Advanced Research in Computer and Communication Engineering 2.1 (2013): 1005-1009.

[8] Pereira, Sérgio, et al. "Brain tumor segmentation using convolutional neural networks in MRI images." IEEE transactions on medical imaging 35.5 (2016): 1240-1251.

[9] Arriaga-Gomez, Miguel F., et al. "A comparative survey on supervised classifiers for face recognition." Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE, 2014.

[10] Ramteke, R. J., and Y. Khachane Monali. "Automatic medical image classification and abnormality detection using k-nearest neighbour." International Journal of Advanced Computer Research 2.4 (2012): 190-196.

[11] Kamavisdar, Pooja, Sonam Saluja, and Sonu Agrawal. "A survey on image classification approaches and techniques." International Journal of Advanced Research in Computer and Communication Engineering 2.1 (2013): 1005-1009.

[12] Dhaware, Chaitali, and Mrs KH Wanjale. "Survey On Image Classification Methods In Image Processing."

[13] Das, Deepjoy, and Alok Chakrabarty. "Emotion recognition from face dataset using deep neural nets." innovations in Intelligent Systems and Applications (INISTA), 2016 International Symposium on. IEEE, 2016.

[14] Kim, J. W. Classification with deep belief networks.

[15] Elleuch, M., Tagougui N., and Kherallah, M. (2015). Arabic handwritten characters recognition using deep belief neural networks. *In Systems, Signals & Devices (SSD),* 2015 *12th International Multi-Conference on*, pages 1–5. IEEE.

Author Profile

Thahseen P. received her B.Tech. degree in computer science and engineering from the MEA Engineering College, Kerala, in 2015. Right now she is pursuing her M. Tech degree in computer science at MEA Engineering College, Kerala from 2015 to 2017. Her research interests lie in Image Processing.

Anish Kumar B received his B.Tech. degree in computer science and engineering from the college of Engineering, Karunagappally, Kerala, in 2006. He had completed his M. Tech degree in computer science at college of Engineering, Chengannur, Kerala from 2014. His research interests lie in Image Processing.