A REVIEW OF DIFFERENT TECHNIQUES FOR DETECTING ALZHEIMER’S DISEASE USING ANN, SVM AND DEEP LEARNING

Punitha M
Assistant Professor, Department of Information Science and Engineering, JSS Academy of Technical Education, Bengaluru, Karnataka, India

Shivagobda Patil
Assistant Professor, Department of Computer Science and Engineering, VTU- Visvesvaraya Technological University, Belagavi, Karnataka, India

Dayananda P
Professor & Head, Department of Information Science and Engineering, JSS Academy of Technical Education, Bengaluru, Karnataka, India

ABSTRACT

Alzheimer’s is one of the most commonly found diseases in the elderly population. As per the world health organization report of 2019, more than 50 million suffer from dementia, and the most common form of dementia is Alzheimer’s disease and more than 10 million cases are reported every year. People suffering from neurodegenerative disease primarily pertain to the elderly population. For developing advanced techniques to detect Alzheimer’s disease, analysis of different existing techniques is very important. This paper briefly discusses the various techniques and challenges involved for classification using Machine Learning and Deep Learning concepts such as Artificial Neural Network (ANN), Support Vector Machine(SVM), Ensemble Methods and Deep Learning (DL) and paper will also discuss biomarkers, genetic data and image modality based on CT scan, PET scan, and MRI.

Key words: Alzheimer’s Disease, Support vector machine, Artificial neural network, Deep learning.

1. INTRODUCTION

A German Neurologist in the early 20th century by the name Alois Alzheimer identified Alzheimer’s disease for the first time. 36 million people are getting affected by Alzheimer’s disease. The disease will cause the brain cells to degenerate and in-turn it causes a continuous decline in thinking, behavioural changes in daily routine. Indeed, the person will not be able to act independently in these social activities [6].

By 2030, the rapid increase in the aged population, dementia may affect about 47 million people worldwide. It is very critical to identify the stage-specific biological marker for Alzheimer’s and also the recognition of dependable biomarkers will basically depend on the information in large biological datasets. Alzheimer’s disease can be diagnosed by advanced technological tools for imaging such as Magnetic Resonance Imaging(MRI), Positron Emission Tomography(PET) imaging, and many more [7].

The categories of Alzheimer’s studies are depending on some machine learning technics which falls in the following searching methods: Detection or diagnosis of AD. Even though most will be on studies which works on AD detection, a few more studies are trying to recognize the nature of the disease, for example, with Region of Interest extraction. Cross-sectional or longitudinal. In a cross-sectional study we gather information from a population at a particular point in time; but in a longitudinal study we again and again collect information from the same sample over an increased period of time.

Modality. In variation to single-modality study and the multi-modality study will be using more than one neuroimaging modality i.e., MRI, PET and CT imaging per subject so as to obtain harmonizing information about modality. Automatic or manual. Even though fully automatic systems are choosing for studies, but some studies also involve manual intervention, for mostly to the reduction brain segmentation faults.

Some of the classification techniques involved in classification using Machine learning to diagnose Alzheimer’s Disease are an Ensemble Methods, Deep Learning (DL) using ML, Artificial Neural Network (ANN), and Support Vector Machine (SVM). Machine learning methods have become very successful in classifying the brain images into functional or structural and to distinguish them into ordinary or a peculiar neurodegenerative disorganization. Ensemble learning is a supervised machine learning tasks, and also an umbrella term for processes that merge various inducers to make a resolution. Ensemble methods are machine learning algorithms that build a set of classifiers and then classify new data points by taking a weight of their forecast. The original ensemble process is Bayesian averaging, but new algorithms also include fault-correcting outcome coding, Bagging, and boosting. One of the most effective areas of research in supervised learning has been to study processes for building good ensembles of classifiers. The main analysis is that ensembles are often much more perfect than each single classifiers that make them up.

Deep structured learning is also part of a wide family of machine learning algorithms which are found on ANN constitutes the machine learning. Learning will also supervised, semi-supervised, or unsupervised. Since the computer collects information through practice, there is no need for a human-computer operator formally to specify all of the information needed by the computer.

The Artificial Neural Networks (ANN) used for diagnosis of AD which uses the Multilayer Perceptron qualified with the Backpropagation algorithm (BP), Learning Vector Quantization (LVQ), Probabilistic Neural Networks (PNN), and Radial Basis Networks (RBF). Backpropagation algorithm is the process which helps to calculate the gradient of a loss function with respects to all the weights in the network. Radial basis function networks are well-known from different networks due to their faster learning speed and universal approximation.
Probabilistic Neural Network (PNN) is a special type of neural network that uses a kernel-based approximation to form an estimate of the probability density function of group in a classification problem. Learning Vector Quantization (LVQ) provides a process for qualifying competing layers in supervised mode. The system is comprised of an unsupervised qualifying competing layer that operates a separation of the intake space. The supervised trained outcome layer provides the labelling of the intake data according to its belonging to a crisp clustering or soft clustering.

The support vector machines (SVM) are immensely capable tools for supervised learning. However, SVM is for clarifying the two-class pattern recognition problem. An SVM is the most commonly used multivariate process in MRI based dementia classification and the classifier has also been extensively and strongly applied in study of using data.

2. METHODS USED FOR AD DETECTION
2.1. Deep Learning for AD Detection
AD detection with the help of some machine learning techniques and also how to provide a better understanding of AD with deep learning and can be used in both supervised and unsupervised learning modes. Basically deep learning methods are combined with the 3 main steps: The first step is to extract the feature, In the second step the dimension reduction of features are being done, and finally, the third step classification of the features. But, the author combines all these steps when using deep learning techniques on features. The brain samples can be processed and will be helpful for below methods: voxel-based, slice-based, Region of Interest -based, and patch-based. When compared with other techniques the classification accuracy is more in deep models than in general machine learning techniques [5].

The data handling methods for AD detection are:

**Sliced-based methods:** - In sliced based architecture it assumes that some region of interest can be decreased to 2D image, and also it reduces the no. of hyper-parameters. The slice-based methods take the middle part of the brain and it will be ignored at rest. The most widely used view is the Axial projection. Slice-based axial scans uses Grey Matter volumes such that slices from the begin and stop, which contain no data, will be removed. From each MRI scan an entropy-based arranged in order to procedure was used to select 32 most important slices from the axial plane. In slice-based process, the image entropy of each slice was calculated from the histogram, which conveyed a estimate of difference in every slice, and the slices with the tallest entropy values were think about the most informative. The advantages of the process is that it will keep away from challenging with millions of variables throughout training and results in simplified networks. The disadvantages of the method is that there will be loss in spatial dependencies of adjacent slices [1].

**Voxel-based methods:** - The Voxel-based method uses straightforward analysis technique which uses the intensity values from the brain image modality or on tissue parts utilizing both grey matter / white matter in the MRI. Voxel-based procedure utilizes spatial co-alignment; the individual images of brain are standard 3D space. The full brain image analysis is performed neither in single-modality or multi-modality mode [9]. On MRI images tissue segmentation i.e., the extraction of grey matter is performed before applying a deep learning model. The advantages of the method is it can obtain three dimensional data of a brain scan. The disadvantages of the method is that it as high feature dimensionality and high computation load and also it passes over the local data of the neuroimaging modalities as it treats each voxel independently [1].

**ROI-based methods:** - In this method it is basically disturbed with the whole brain i.e., Region of Interest methods gives a spotlight on some parts of the brain i.e., is the damaged in the early
stages of AD. ROI usually requires pervious information of the unexpected regions and a brain outline. A deep learning model to excerpt descriptive ROIs from Positron Emission Tomography scans selected Region of Interest from Grey Matters segmentation of Magnetic Resonance Imaging, and then calculated a weighted connectivity matrix of ROIs, which represents the connection advantage between region pairs, to produce a final brain network. On behalf of directly learning the topological features from complicated brain networks, the method learns the corresponding eigenvalues of the matrix, giving a compact and complete feature representation. The advantages of ROI-based is that it can Easily be interpretable, has a low feature dimension and fewer features can reflect the entire brain. The disadvantage has limited available information about the brain regions involved in AD and also avoid detailed irregularities [1].

**Patch-based methods:** - In Patch-based method it is basically a patch is defined as cube of three Dimensional. In the brain extracting features from small image patches can also capture disease-related patterns. One of the main goal in this method is to choose both patch-level and image-level features and then will select the most informative image patches for capturing. In the whole brain MRI, from each local region is extracted uniformly partitioned into different small size and several 3D patches. These patches from more clusters then each local region were grouped with the K-means clustering method before the classification. The advantages of the method is that the sensitive to small changes it does not require Region of Interest identification. The main disadvantages of this method is that it has challenges to select the most informative image patches [1].

The fundamental concepts of deep learning techniques are split into two methods: Unsupervised deep learning and Supervised deep learning methods. The first is unsupervised learning these will try to have a task-specific representation from imaging data. By using an unsupervised deep neural network, we can extract high level features from machine learning to deep learning will be first taken. It is also used as an ordered manner to preferred features from any one of the scans MRI, PET biomarkers which interactively removed uninformative features. Unsupervised features learning methods are detached into Auto Encoder and Restricted Boltzmann machine.

**Auto-Encoder:** The auto-encoder also consists of two modules in auto-encoder they are an encoder and a decoder. The rephrasing error between the intake and outcome values of the network can be obtained by compressing representation from input data will be minimized. The encoder will map the information from the intake space to the depiction space to keep the most absolute features. The decoder plans it back into the intake space from the representation space, it reconstructing the data. The features learned in the middle layer of an auto encoder can be removed and used as the pre-training phase for features removal and we can also reduce the dimension in an unsupervised learning. The advantages of Auto Encoder are that it can depict highly nonlinear and multiplex patterns, these method is good initialization for convolutional neural network, it also a good dimension reduction easy to implement. The main disadvantages are to learn to arrest as much data as feasible than as much relevant data [1].

**Restricted Boltzmann machine:** RBM is a visible layer and a hidden layer. The hidden layer will be single-layer one directional graphical model. The Restricted Boltzmann machine adopts symmetric links between visible and hidden layer but doesn’t have connection among units of the same layer. It can generate intake data from the hidden layer [10]. To change these intake information features of an MRI into an embedded low-dimensional space by detecting of the Regions of interest. A Deep Belief Network is a Restricted Boltzmann Machine can be stacked to constrict a deep architecture. The Deep Belief Network model as many advantages but the computational exorbitant to create a elaborated initialization process. It also contains a one directional connection at the top two layer and a directional connection at the below layer. Some
of the advantages are it can learn very good generative model, the method is able to generates patterns if there are removed information. The disadvantages of restricted Boltzmann machine is that computationally high-rated in the training process[1].

The Supervised deep learning methods has higher popularity which has feature extraction and classification are combined together in a single model but the unsupervised method consists a feature extractor. Supervised features learning methods are separated into DNN, DPN, CNN and RNN.

**Deep neural network:** The deep neural network incorporates more piled layers with the same organization as established into Multi-Layer Perception network. It can also discover complex relations of intake information patterns and also provides a good understanding of the information. Deep neural network is an absolutely supervised deep learning which finds extremely abstract patterns and correlations. The deep feature learning and fusion framework as three main stages for MRI, PET and Genetic data, here the first stage involving the learning of representations of each modality, the last stage as the learning of joint features and in each stage a Deep neural network is used and each Deep neural network consist of several fully-connected hidden layers and an outcome layer. Deep neural network is mostly used in supervised manner; an unsupervised method has three hidden layers to extract high level feature with a classifier. The strengths of the deep neural network are that a good for vector-based problems can be handled datasets with a large number of samples and it can also detect complex nonlinear relationships. Limitations of deep neural network which has a slow training process and not optimal for image and also has generalization issues[1].

**Deep polynomial network:** It is one more supervised deep learning method; it is similar or even good at performance when compared to Deep Boltzmann Network and stacked Auto-Encoder. It can be future enhanced representation performance by constructing a stacked and much deeper configuration. In a multi-modal stacked deep polynomial network which consists of two stages for feature extraction from multi-modal imaging data. In two multi-modal stacked deep polynomial network the first one is used to learn first-level attributes from MRI & PET, from here attributes are the fed to one more multi-modal stacked deep polynomial network to fuse multi model imaging data. The final result learned first-level attributes which contain both the inherent properties of each modality. Some of the strength of these method is that it can detect complex attribute representation from small samples. The limitations of deep polynomial network it has limited achievement due to the simple concatenation of the learned hierarchical features from different layers [1].

Convolutional neural network: The convolutional neural network is a type of deep neural network which is inspired by visual cortex of the brain. The main benefit and idea behind the convolutional neural network is that to merge feature extraction and classification. The majority successful deep learning model for neuroimaging analysis and has been designed for better utilization of spatial information by using two-dimensional or three-dimensional neuroimages as the intake data and extract attributes by using several convolutional layers, which results in more abstract attributes. Convolutional neural network architecture is made up of several softmax layer, activation layers, convolutional layers, pooling layers and fully-connected layers. The convolutional layer is followed by a pooling layer, which down-samples the input data feature map by replacing each non-overlapping block with its max or average. The functions of the softmax layer is the highest values in a vector which significantly below the max. The convolutional neural networks are specially designed for recognizing patterns in two-dimensional images. The distinct deep architectures using two-dimensional convolutional neural networks are using usual deep models with a couple of fundamental convolutional layer paired with pooling layers and then by fully-connected layers & a softmax layer.
The three-dimensional convolutional neural network is used because imaging provides a three-dimensional image and there is a spatial relationship among the different image data, this process is most popular one. The complexity in detecting AD must take the full image or some Region of Interest’s as the intake. It also may be essential that training a more number of parameters on a small dataset, this will result in does not fit the data enough. Cascaded convolutional neural network has been constructed to learn attributes from different brain scans. Multiple deep three-dimensional convolutional neural networks with four convolutional layers are built on different local image data patches to transform the local brain image into more compact first-level attributes. Let’s now see the first-level two-dimensional convolutional neural network with two convolutional layers is cascaded to merge the high-level attributes and produce the multi-modal correlation attributes of the image data patches. These extracted attributes are then merged by a fully-connected layer succeeded by a softmax layer. The strength of the two-dimensional convolutional neural network have good achievement in local attribute extraction in images data, the datasets are easy to be trained, in three-dimensional convolutional neural network it is good to perform in local attribute extraction in images and it can also capture three-dimensional data from the three-dimensional volumes of a brain scan. Some of limitations of both two-dimensional and three-dimensional are two-dimensional convolutional neural network cannot encode the spatial data of the three-dimensional images across the third dimension and the in three-dimensional convolutional neural network is computationally expensive in the training process [1][2][4].

**Recurrent neural network:** Recurrent neural network is used in video applications, which includes a ‘memory to model’ temporal dependency. In recurrent neural network method, the previously data is implicitly stored in hidden units known as state-vectors. These state-vector, the outcome of the sequential intake to be calculated and considered as the intake data as well as past intake data. This process is not same as deep neural network and convolutional neural network in terms of the number of layers and as some limitation of memorizing long-term intake data. A recurrent neural network uses a large dataset. Long short-term memory is the exchanged the plain perceptron hidden data units with more complicated units. The Gated recurrent unit function as memory cells which helps in overcoming the memory limitations. Recurrent neural networks are more likely to be used on increasingly applied images. The strengths of recurrent neural network are good for sequential 2D images and also good for longitudinal studies. The limitations of the method are it has issues related to the training process due to vanishing gradients [1][2].

**2.2. Support Vector Machine for AD detection**

The support vector machine is one of the stable and widely used technique for classification problem. By making use of the Structural Risk Minimization(SRM) principle, SVM allows sound generalization performance and to classify the data points, it also uses the maximum margin principle [8]. In support vector machine the important part is feature selection in the classification of data. In this method for AD detection the different attributes are combined to setup the attributes vector for future work; which is also found that it includes demographic and genetic data with MRI/PET brain images which improves classification accuracy. To detect subtle changes in Grey, matter a refined parcellation method is being used. Classification of various form of MCI patients is easier with the SVM-based analysis of white matter DTI attributes [5].

In classification of data, the attribute choosing is a part which plays an important role. The feature vector is constructed by combining different attributes. It is opted that Non-negative matrix factorization (NMF) based attributes with SVM perform superiorly than PCA with SVM to classify CN vs AD. The effects of hardware no uniformity on classification accuracy of SVM.
They also found high confidence level of classification performance for large samples. The attribute choosing techniques for MRI scans may lead to low classification accuracy due to addition of hyper-parameters. Volume based attributes to be more beneficial than voxel based morphometry (VBM). Morphological attributes of brain image regions are being utilized. Variety of kernels have been utilized with SVM for classification of Alzheimers. The linear kernel is the most widely used one with SVM by various researchers as kernel parameter is not required to be tuned. It is also noticed that researchers use multiple kernels with SVM. Using linear SVM to categorize pathologically verified cases of Alzheimer with CN, and Proposed that SVM can help in the analysis of Alzheimer. In linear kernel also gives good classification results for high dimensional data when opposed to polynomial. However, polynomial kernel is also utilized. Usage of the polynomial kernel is beneficial in the classification of CN vs Alzheimer using the PCA attribute.

Image modality is used as a source for analyses of Alzheimer. Prediction of AD is done by using one of the classification algorithms in Machine learning called support vector machine (SVM). Most of the recent work is using image modality especially MRI and MRI is based on neuroimaging, Accuracy of classification is achieved good. Scans w.r.t MRI is costly for normal person and time required is significant. Extraction of molecule was very difficult task. In order to reduce features earlier than classification, Principal component analysis is used. Multimodel data have been engaged by multiple biomarkers for identifying AD. Author’s[1] explains about Prediction of AD using SVM based on gene-coding protein sequence. The sequence information is described by frequent occurrence of amino acids, protein sequence is used to predict AD. 400dimension vector is used to represent sequence and two amino acids are used to represent each dimension. Accuracy of 85.7 % is achieved using SVM. A dataset with non-AD and AD samples are formed. This dataset could also be used for further prediction studies of AD.

2.3. Artificial Neural Network (ANN) in AD Detection

There are four categories model used in Artificial Neural Networks (ANN): Backpropagation (BP), Radial Basis Networks (RBF), Learning Vector Quantization Networks (LVQ) and Probabilistic Neural Networks (PNN) to perform grouping of patients of mild Alzheimer’s disease vs. managed subjects. In Backpropagation, the activation function used by the "nodes" is differentiable with its plagiarize being a simple function of itself. It makes use of squared fatal gradient descent rule for updating weights in nodes of one-layer perceptron, generalized to feed-forward networks. ANN utilizes radial basis function as activation functions and also consists of 2-layers network that is hidden in each implements as a radial activated function [5]. A Probabilistic Neural network(PNN) is a type of neural network that uses a kernel-based approximation on approximate the probability density function of classes in a classification problem. The optimal sigma value of the spread of the RBF functions is selected to tune the PNN network [3].
3. COMPARITIVE ANALYSIS TABLE

<table>
<thead>
<tr>
<th>TECHNIQUES</th>
<th>FEATURES</th>
<th>IMAGE MODALITY</th>
<th>ACCURACY</th>
<th>ADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Neural Network (ANN)</td>
<td>Voxel-based Morphometry</td>
<td>MRI</td>
<td>83%</td>
<td>Simple neural network, High Fault tolerance , Input is stored in open neural network instead of database, they can perform multiple task in parallel without effecting the system performance.</td>
</tr>
<tr>
<td>Support Vector Machine for AD detection</td>
<td>Support vector, Hyperplane, Margin</td>
<td>Text, MRI, CT Scan,</td>
<td>89%</td>
<td>SVM is more effective in high dimensional, relatively memory efficient, SVM don’t suffer condition of over fitting, Faster prediction along with better accuracy.</td>
</tr>
<tr>
<td>Deep Learning for AD Detection</td>
<td>Sliced-based, Voxel-based, ROI-based and Patch-based</td>
<td>CT Scan, MRI, PET, fMRI and Genetic data</td>
<td>94%</td>
<td>High Quality Prediction, Works well with unstructured data, Parallel Computing, Eliminates the need of data library, delivers high results</td>
</tr>
</tbody>
</table>

4. CONCLUSION

In this paper analysis of different techniques Alzheimers disease detection using Artificial neural network, Support vector machine and deep learning. ANN has an advantages like high fault tolerance, input is also stored in neural network and also perform multiple task in parallel. SVM is easier to understand compared to deep learning and deep learning handles variety of unstructured data and accuracy achieved is more compared to ANN and SVM. Training and feature extraction is easy compared to ANN and SVM. More number researchers are more concerned about features extraction.

REFERENCES


