Alzheimer's disease Detection and Prediction:A Survey

Chiramel Riya Francis,Prabhjyot Kaur Bamrah and Arlene D'costa
Department of Information technology Engineering
Fr Conceicao Rodrigues College of Engineering,Bandra

Abstract : Alzheimer's disease (AD) is characterized by severe memory loss and cognitive impairment .It is the most common type of dementia, accounting for 60 percent to 80 percent of dementia cases. Even though no cure has been obtained for curing Alzheimer's disease, accurate and early diagnosis of Alzheimer's disease (AD) plays a significant role in patient care, especially at the early stage, because the consciousness of the severity and the progression risks allows the patients to take prevention measures before irreversible brain damages are shaped. Alzheimer's Disease is a disease occurring in different stages causing brain structure changes which can be measured by magnetic resonance imaging(MRI).Detection of Alzheimer's is usually done with these variations in brain anatomical structures and features such as destruction of neurons in parts of the brain evolved in memory such as entorhinal cortex and hippocampus .It later affects areas in cerebral cortex responsible for language ,reasoning and social behavior .These structure changes captured by the MRI can be exploited for further AD detection and prediction using image classification tools ,like Convolutional neural network (CNN) for feature extraction.

IndexTerms **- Alzheimer's disease, prediction, detection, deep learning , machine learning (**key words**)**

I. INTRODUCTION

Alzheimer's disease (AD)is the most common type of dementia, and accounts for a high percentage of dementia cases. It is a chronic neurodegenerative disease that starts slowly and gradually worsens over time. Alzheimer's disease is characterized by progressive memory impairment and decline in other cognitive functions. Also, people suffering from Alzheimer's find it difficult to remember important events or dates. Alzheimer's mainly occurs due to the presence of two proteins beta-amyloid and tau.The beta-amyloid proteins accumulate between the neurons of Alzheimer's Patients and forms clumps.Tau proteins form neurofibrillary tangles that gets collected between neurons in late Alzheimer's patients .Currently in India more than 1.6 million people are suffering from Alzheimer's Disease. As the population ages, prevalence of Alzheimer's is estimated to rise in the coming decades.It is also the seventh leading cause of death in the United States eventually causing heavy financial burden. This disease usually progresses slowly in three stages — mild (early stage), moderate (middle stage), and severe (late stage). Precise diagnosis of Alzheimer's Disease was initially considered difficult due to the confusion of other non-AD dementia syndromes. MCI a prodomal stage of AD has gained attention recently as it is useful for clinical trials and it has been constantly proven that MCI patients are at a high risk of Alzheimer's progression. Neuroimaging technique, such as magnetic resonance imaging (MRI) has been widely used in the assessment of AD, along with many other non-imaging biomarkers .Recently, Convolutional Neural Networks(CNN) are widely used for image classification tasks along with other Machine Learning techniques like KNN,SVM, Decision Trees etc,

II. SURVEY

Machine learning methodologies are extensively used for computer aided diagnosis in medical image formation and classifications of brain disease using CRT images and X-Rays.

2.1 Support Vector Machine

SVM is a binary classifier based on supervised learning .(SVM)[6] are best considered for classification since they are supervised learning models with associated learning algorithms that find the best hyperplane, the one that represents the largest separation, or margin, between the two classes. It is classified in two groups, *Linear SVM and Non Linear SVM.* SVM is highly used for Alzheimer's Prediction and Detection[7] .The most discriminant features for MCI to AD prediction are served as input to a support vector machine (SVM) classifier that groups MCI subjects and Normal Patients.

2.2 Decision tree

Decision trees are used to build classification or regression models in the form of a tree structure.The final result is a tree with *decision nodes* and *leaf nodes*. A decision node has two or multiple branches. Leaf node represents a classification or the final decision. The topmost decision node in a tree which corresponds to the best predictor called **root node .**Some of the examples of decision tree algorithms include ID3 (Iterative Dichotomized 3),CART (Classification and Regression Tree)and C4.5 [8]. These algorithms differ in how attributes are selected as splitting nodes in the tree, along with other differences. Decision tree induction method can be applied to Alzheimer's disease prediction[9] .Entropy or Information Gain is used to determine which attribute to branch on at each level of the tree and construct an optimal tree with fewer nodes and branches. The system is then used to predict the situation of new patients with regards to Alzheimer's disease. The decision tree is traversed from the root node downwards based on the attribute values pertaining to the new patients until a leaf node is reached. The leaf node represents the Alzheimer's disease status prediction.

2.3 K Nearest Neighbors

K nearest neighbors is a simple **algorithm** storing all available cases and classifies new cases based on a similarity measure by calculating the Euclidean distance to measure the difference between two instances. KNN is widely used in classification. Early diagnosis of Alzheimer's can be done with the help of KNN[10].KNN is highly sensitive to class imbalance.It significantly enhances performance by achieving better balance between discrimination of MCI and AD instances.

2.4 Naïve Bayes

Naive Bayes classifier is a machine learning model that's used for binary and multi-class classification tasks. The essence of Naive Bayes classifier is on the Bayes theorem. This classifier calculates the set of probabilities by counting the frequency and combination of values in a given data set and assumes that all variables contributing towards classification are mutually independent[23]. Equation 1 is the probability of a document 'd' with vector 'x= {x1, x2…xn}' belongs to hypotheses 'h' is given by,

$$P(h1|xi) = \frac{P(xi|h1)P(h1)}{P(xi|h1)P(h1)+P(xi|h2)P(h2)}) \quad (1)$$

P (h1|xi) is the posterior probability and P(h1) is the prior probability associated with hypothesis h1.Equation 2 shows the posterior probability for n different hypotheses, [23].

$$P(h1|xi) = \frac{P(xi|h1)P(h1)}{P(X1)} \quad (2)$$

$$P(xi) = \sum_{j=1}^{n} P(xi|hj)P(hj) \quad (3)$$

The naïve bayes classifier is extensively used for classification and has been for classification of Alzheimer's Disease from MRI images[24]. It was proved that Naive Bayes had the second best build time at 1.6 seconds but a detection rate of 93.52%.

III. Data Set

The neuroimaging data used in Alzheimer's Disease Detection and Prediction is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. This database was launched in 2004 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations as a five-year public partnership. The primary purpose of the ADNI project was to study the effects of combining multiple biomarkers, such as MRI, PET, and CSF data accompanied with neuropsychological assessments, to predict the progression of MCI and early AD. The ADNI study has four phases: ADNI1, ADNI GO, ADNI2 and ADNI3 [11]. ADNI phase 1 was launched in October 2004 and analyzed thousands of brain scans, genetic profiles, and blood and cerebrospinal fluid biomarkers. Phase 2 focused on examining how brain imaging and biomarkers are used in order to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. ADNI Go was formed with the purpose of examining biomarkers at an earlier stage of Alzheimer progression. ADNI3 started in 2016 with an objective of optimizing, standardizing existing clinical trials.

IV. Deep Learning

4.1 CNN

In deep learning , a convolutional neural network is a class of deep neural networks which are applied on images .A convolutional neural network consists of an input and output layer with variable number of hidden layers.The hidden layers further comprise of convolutional layers , RELU layer i.e. activation function, pooling layers, fully connected layers and normalization layers. The role of the Convolutional neural network is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.[1]CNN comprises of the following layers

4.1.1 Organization of the Input Data to the CNN

While using CNN, the input data need to be separated and organized as a number of feature maps to be fed into the CNN. This is a term borrowed from image-processing applications, in which it is mandatory to organize the input as a two-dimensional(2-D) array, for the pixel values at the horizontal and vertical coordinate indices. For color images, RGB(red, green, blue) values can be viewed as three different 2D feature maps. CNNs run a small window over the input image at both training and testing time, so that the weights of the neural network that operate on the data can learn from various features of the input data regardless of their absolute position within the input.Weight sharing, or full weight sharing refers to the decision to using the same weights at every position of the window.CNNs are also often said to be instantaneous because the individual units that are computed at a particular position of the window depend upon features at the current region of the image that the window currently looks upon.

There exists different alternatives for organizing different features into maps.To extract features from CNN model first we need to train the CNN network with last sigmoid/logistic dense layer(here dimension or w.r.t. target variable. The objective of the training network is to identify the correct weights for the network by multiple forward and backward iterations, which eventually try to minimize binary cross entropy(misclassification cost). Once the input feature maps are formed, the convolution and pooling layers apply their respective operations to generate the activation functions of the units in those layers, in sequence. Similar to those of the input layer, the units of the convolution and pooling layers can also be organized into maps. According to the CNN terminology, a pair of convolution and pooling layers in succession is usually referred to as one CNN "layer." A deep CNN thus consists of two or more of these pairs following each other. To avoid confusion, we will refer to convolution and pooling layers as convolution and pooling plies, respectively.

4.1.2 Convolution Layer

Every input feature map (assume is the total number), oO_i(i=1,…..I) is connected to many feature maps(assume in the total number), Q_i(i=1,…..J)in the convolution ply based on a number of local weight matrices (in total) $(I \times J)w_{i,j}(i==1,\ldots..,J)$.The mapping can be represented as the well-known convolution operation in signal processing. Assuming input feature maps are all one dimensional,each unit of one feature map in the convolution ply can be computed as:

$$q_{j,m}=(\sum_{i=1}^{I}\sum_{n=1}^{F} oi,n+m-1wi,j,n \ + w_{0,j}) \ , \ (j=1,\ldots..,J \qquad (1)$$

Where $o_{i,m}$is the *i-th* unit of the *j-th* input featuremap ,$q_{i,m}$is the i^{th} unit of the j^{th}feature map$Q_i$in the convolution player,$w_{i,j}$ is the i^{th} element of the weight vector,$w_{i,j}$ which connects the th input feature map to the th feature map of the convolution layer is called the filter size, which determines the number of frequency bands in each input feature map that each unit in the convolution ply receives as input.*F*is called the filter size, which determines the number of frequency bands in each input feature map that each unit in the convolution ply receives as input.[1]Equation (2) can be written in a more concise matrix form using the convolution operator as:

$$Q_j = \sigma(\sum_{i=0}^{n} 1 * o_i * \ w_{i,j}) \ (j = 1,\ldots..,J) \qquad (2)$$

Where O_i represents thei^{th}input feature map and $w_{i,j}$represents each local weight matrix,[1] flipped to refer to the convolution operation's definition. BothO_i and $w_{i,j}$are vectors if one dimensional feature maps are used, and are matrices if two dimensional feature maps are used .The number of feature maps in the convolution layer directly determines the number of local weight matrices that are used in the above convolutional mapping .

A convolution layers differs from a standard, fully connected hidden layer in two important aspects, however. First, each convolutional unit receives input only from a local area of the input.This means that each unit represents some features of a local region of the input. Second, the units of the convolution ply can themselves be organized into a number of feature maps, where all units in the same feature map share the same weights but receive input from different locations of the lower layer.[3]

4.1.3 Learning Weights in the CNN

All weights in the convolution layer can be further trained using the error back-propagation algorithm or gradient descent algorithm [1].

4.2 RNN

In a RNN, the information cycles circulates through a loop. When it makes a decision, it takes into consideration the current input and also what it has learned from the inputs it received previously.RNN have a **"memory"** which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.Recurrent networks are those networks who take input from not only from the current data but also the data that has been perceived previously in time.[2]

RNN consists of individual nodes which have mathematical operations that operate on each node respectively.Information circles continuously as a loop recurring repeatedly hence the network is called recurrent neural network.[4]The decision taken at $(t-1)^{th}$ minute affects the decision taken by the node at the $(t)^{th}$ minute.Recurrent networks are furthermore added with memory.Sequential information is saved in the hidden layers of the recurrent networks which cascades forward with the time and can help in processing of the corresponding example.It thus obtains correlation between events separated by time that is moments and this co-relation is called "long term dependencies".RNN are a way of sharing weights over with time in the form of functions.Thus , a mathematical formula for carrying the memory carrying function forward is as follows:

$$h_t = \emptyset(Wx_t + Uh_{t-1}), \qquad (5)$$

At time t,the hidden time is h_t.It is defined as the function of the input at the time step x_t , changing the weight w according to the time step .The data obtained at the time step t-1-1 can be further utilized to update the following corresponding weight at the current time step t which is further multiplied by hidden state matrix U,known as transition matrix or also known as Markov chain.Weights present in any neural network act as filters for determining the importance of data for the hidden as well as the the input layer.During this weight changing process, errors are generated . These errors are then sent via back propagation to update the weights producing errors and this process should be continued till the errors are removed or reduced to negligible amount.

4.2.1 Back propagation through time(bptt)

Backpropagation or Gradient Descent can be used for removal of errors among weights in the hidden as well as input layer[2].The algorithm of backpropagation works by obtaining partial derivatives$-\partial E/\partial w$,or the relationship between the rate of change of weights .Gradient descent is then used for adjusting the weights which is further used for minimizing error.

Backpropagation through time form an integral part of recurrent networks[2] .Time, is thus used for linking one time step to another with a series of calculations which is essential for backpropagation to work.

Neural networks are just a group of composite functions such as *f(g(h(x)))*nested together supported by time series functions sufficing chain rule[4].

V. COMPARATIVE STUDY

Author	Dataset	Algorithms	Result
BaiyingLei,Peng Yan, Tianfu Wang, Siping Chen, and Dong Ni [11]	ADNI dataset [16]	The algorithms used are Lasso , DLSR and M3T used for regularization further enhancing feature selection.	Feature selection algorithms yielded accuracy of 82.03%.
Jing Wan, Zhilin Zhang, Bhaskar D. Rao,Shiaofen Fang, Jingwen Yan, Andrew J. Saykin, Li Shen[12]	ADNI dataset[16]	The algorithms used in this paper are CORNLIN,T-MSBL-FP Elastic Net Regression and Ridge regression. These models the non-linear function of data to be predicted by extending the prediction matrix with block structures.	This model yields better results than tradional sparse learning algorithm.
Siqi Liu, Sidong Liu,WeidongCai,HangyuChe,SoniaPujol, Ron Kikinis, DaganFeng,Michael J.Fulham, [13]	ADNI dataset[16]	The algorithms used in this paper are SAE,Softmax Logistic Regression,Zero mask training strategy which has the ability of fusing multi-modal neuroimaging features and can thus use less or non-labeled data.	Binary classification achieved classification accuracy of 91% and specificity of 91.04% for ND and AD.
LiqiangNie,Luming Zhang, Lei Meng, Xuemeng Song, Xiaojun Chang, and Xuelong Li[14]	ADNI dataset [16]	The algorithms used in this paper are multisource multitask Learning and Gradient Descent .This paper studied the modeling of chronic diseases by studying imaging and nonimaging sources at a given time.	The MF technique was used to solve the data solving problem and it was proved that the model is a linear model.
Sidra Minhas, Aasia Khanum2, Farhan Riaz1, Shoab A. Khan1, Atif Alvi2, [15]	ADNI dataset[16]	The algorithms used in this paper are CSF bio-specimens ,PET, Aβ imaging .	For AD prediction the accuracy obtained is 88.93% with the algorithms in the given paper.
Ciprian D. Billones, Jr., Olivia Jan Louville D. Demetria, David Earl D. Hostallero, and Prospero C. Naval, Jr.	ADNI website[16]	The algorithms used in this paper is CNN which includes 3- way classification.	The accuracy obtained for this paper with CNN is 91.85%.
H. M. Tarek Ullah, Zishan Ahmed Onik,Riashat Islam, Dr. Dip Nandi[18]	ADNI website[16]	The algorithms used in this paper are CNN for image classification and SVM for feature extraction.	The accuracy obtained using CNN and SVM is 92%.
Arezou Moussavi-Khalkhali and Mo Jamshidi, Subhashie Wijemanne[19]	ADNI website[16]	CNN is used for classification of images and Autoencoders for feature fusion.	The accuracy obtained is 95% using these following algorithms.

Ehsan Hosseini-Asl1, Robert Keynton, Ayman El-Baz[20]	ADNI website[16] with 70 AD ,MCI and NC patients respectively.	Feature extraction is done using 3D Convolutional neural network.AD classification is futher done using 3D adaptive CNN.	The accuracy obtained by using 3D A_CNN is 90.38.
Saman Sarraf, Ghassem Tofighi[21]	fMRI data from ADNI website[16]	The algorithms used in this paper are conv-nets which were fed raw-fmri data for feature extraction and further classified using SVM.	The accuracy obtained by ConVnet classification is 96.86%.
Shui-Hua Wang, Preetha Phillips Yuxiu Sui4 & Bin Liu & Ming Yang & Hong Cheng7	MRI data from ADNI website.[16]	Algorithms used in the paper are CNN for image classification and softmax classifier for classification for prediction of AD.	The accuracy obtained is 97.05%.

VI .CONCLUSION

Thus it can be concluded from the survey that machine learning algorithms yield lesser accuracy as compared to the deep learning algorithms for feature extraction and image classification for Alzheimer's prediction. The dataset for Alzheimer's include MRI images providing data which is unstructured and hence is difficult to analyze by the machine learning algorithms. Deep learning algorithms do not need labeled data for classification and excel at quick learning without guidelines. Machine learning algorithms are not successful in adaptive learning. Although deep learning algorithms need a large amount of data for training their neural network they provide quick optimized and highly accurate results as compared to machine learning algorithms which are trained quickly hereby providing results with lower accuracy.

REFERENCES

[1] Islam, J. and Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics*, 5(2).

[2] Nguyen, M., Sun, N., Alexander, D., Feng, J. and Yeo, B. (2018). Modeling Alzheimer's disease progression using deep recurrent neural networks. *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*.

[3] Farooq, A., Anwar, S., Awais, M. and Rehman, S. (2017). A deep CNN based multi-class classification of Alzheimer's disease using MRI. *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*.

[4] Faust, O., Hagiwara, Y., Hong, T., Lih, O. and Acharya, U. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161, pp.1-13.

[5] Schuster, M. and Paliwal, K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), pp.2673-2681.

[6] V.N. Vapnik. Statistical Learning Theory. John Wiley and Sons, Inc., New York, USA, 1998

[7] Dominguez, A., Ramirez, J., Gorriz, J., Segovia, F., Salas-Gonzalez, D., Martinez-Murcia, F. and Illan, I. (2016). Statistical feature selection and classification models for Alzheimer's disease progression assessment. *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*.

[8] Barros, Rodrigo C., Basgalupp, M. P., Carvalho, A. C. P. L. F., Freitas, Alex A., A Survey of Evolutionary Algorithms for Decision-Tree Induction. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 42, n. 3, p. 291-312, May 2012.

[9] AL-Dlaeen, D. and Alashqur, A. (2014). Using decision tree classification to assist in the prediction of Alzheimer's disease. *2014 6th International Conference on Computer Science and Information Technology (CSIT)*.

[10] Al-Badarneh, A., Najadat, H. and Alraziqi, A. (2012). A Classifier to Detect Tumor Disease in MRI Brain Images. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

[11] Lei, B., Yang, P., Wang, T., Chen, S. and Ni, D. (2017). Relational-Regularized Discriminative Sparse Learning for Alzheimer's Disease Diagnosis. *IEEE Transactions on Cybernetics*, 47(4), pp.1102-1113.

[12] Wan, J., Zhang, Z., Fang, S., Risacher, S., Saykin, A. and Shen, L. (2013). Sparse Bayesian learning for identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease. *Alzheimer's & Dementia*, 9(4), pp.P62-P63.

[13] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M. and ADNI (2015). Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. *IEEE Transactions on Biomedical Engineering*, 62(4), pp.1132-1140.

[14] Nie, L., Zhang, L., Meng, L., Song, X., Chang, X. and Li, X. (2017). Modeling Disease Progression via Multisource Multitask Learners: A Case Study With Alzheimer's Disease. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7), pp.1508-1519.

[15] Minhas, S., Khanum, A., Riaz, F., Khan, S. and Alvi, A. (2018). Predicting Progression From Mild Cognitive Impairment to Alzheimer's Disease Using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers. *IEEE Journal of Biomedical and Health Informatics*, 22(3), pp.818-825.

[16] Adni.loni.usc.edu. (2019). *ADNI | Data Types*. [online] Available at: http://adni.loni.usc.edu/data-samples/data-types/ [Accessed 24 Apr. 2019].

[17] Kim, S. and Choi, M. (2012). Neuropsychiatric symptoms in mild Alzheimer's disease and mild cognitive impairment. *Alzheimer's & Dementia*, 8(4), p.P367.

[18] Ullah, H., Onik, Z., Islam, R. and Nandi, D. (2018). Alzheimer's Disease and Dementia Detection from 3D Brain MRI Data Using Deep Convolutional Neural Networks. *2018 3rd International Conference for Convergence in Technology (I2CT)*.

[19] Moussavi-Khalkhali, A., Jamshidi, M. and Wijemanne, S. (2016). Feature Fusion for Denoising and Sparse Autoencoders: Application to Neuroimaging Data. *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*.

[20] Hosseini-Asl, E., Keynton, R. and El-Baz, A. (2016). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. *2016 IEEE International Conference on Image Processing (ICIP)*.

[21] Sarraf, S. and Tofighi, G. (2016). Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. *2016 Future Technologies Conference (FTC)*.

[22] Wang, S., Phillips, P., Sui, Y., Liu, B., Yang, M. and Cheng, H. (2018). Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling. *Journal of Medical Systems*, 42(5).

[23] K.P.Soman, Shyam Diwakar, V. Ajay "Insight into data mining theory and practice", New Delhi PHI learning Pvt. Ltd.

[24] Automatic Detection and Classification of Alzheimer's Disease from MRI using TANNN. (2016). [ebook] Available at: https://pdfs.semanticscholar.org/7413/739f99c7e9225b14e8db68bf60101d56c77a.pdf [Accessed 25 Apr. 2019].