Machine Learning Algorithms for Classification of Various Stages of Alzheimer's Disease: A review

Sameer Mahajan[1], Gayatri Bangar[2], Nahush Kulkarni[3]

[1,3]Student, Department of Computer Engineering, TEC, University of Mumbai, Mumbai, India
[2]Student, Department of Information Technology, TEC, University of Mumbai, Mumbai, India

---

Abstract: *Alzheimer's disease (AD) is an incurable chronic disorder that leads to cerebral atrophy resulting in memory impairment and cognitive dysfunction. Machine learning techniques are extensively used in various medical fields. There has been considerable research for the classification of Alzheimer's disease. In this paper, we have reviewed various papers by different researchers that use different machine learning techniques, in order, to provide a better understanding of the work that has been done in the field of Alzheimer's disease. The sole purpose of doing so is to evaluate more effective and coherent learning techniques for the classification of Alzheimer's Disease. Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Deep Learning (DL) are the main machine learning techniques that are scrutinized.*

Index Terms: Alzheimer's Disease, Support Vector Machine, Artificial Neural Network, Deep Learning, Machine Learning, Classification of Alzheimer's Stages.

1. INTRODUCTION

Alzheimer's disease (AD), a regular dementia type, is a progressive neurodegenerative disease that results in gradual but continuous deterioration of memory and causes impairment in intellectual capabilities and other mental functions. Structural changes in the brain arise due to Alzheimer's disease. Symptoms develop gradually and slowly and worsen with passing time. The patient, initially, develops a mild cognitive impairment (MCI) and gradually progresses to AD. MCI is an intermediate stage of AD. Although, not all MCI patients get converted into AD [1]. AD is currently incurable but the disease can be prevented from progressing if it is detected at an early stage [2]. 26.6 million people suffered from AD in 2006. By 2050, AD is predicted to affect 1 in 85 people globally, roughly 43% of ubiquitous cases need a high level of care[3]. Hence, the main aim of this research is to predict the conversion of MCI to AD.

In the past decade, machine learning techniques have been widely used in the diagnosis of AD[4,5,6]. Support vector machines (SVM), artificial neural networks (ANN), and deep learning (DL) are the most substantially used classification techniques. The main difference between SVM and ANN is the nature of the optimization problem. The scope of solution in SVM is global [7] and local in the case of ANN. The feature extraction step is most important in both SVM and ANN. Whereas, the feature extraction step in deep learning is integrated into the learning model itself. Deep learning is found to be useful for image data in large datasets[8]. The Alzheimer's Disease Neuroimaging Initiative (ADNI) [9](adni.loni.usc.edu), Australian Imaging, Biomarker & Lifestyle Study of Ageing(AIBL)(aibl.csiro.au), Open Access Series of Imaging Studies(OASIS)(www.oasis-brains.org) are most widely used databases. Magnetic resonance imaging (MRI) images of AD, MCI, cognitive normal healthy (CN) are pooled from ADNI datasets[10]. The diagnostic accuracy of AD is greatly improved by using high-resolution magnetic resonance (MR)images.

For the classification of AD patients from CN, MR images play an important role [11] as patients with AD show patterns of grey matter(GM) atrophy. The use of Voxel-based Morphometry (VBM) is proposed by many researchers for measuring the spatial GM atrophy distribution in the brains of MCI and AD patients [12,13,14]. VBM uses Statistical Parametric Mapping (SPM) for structural MRI data analysis [15,16]. SPM is developed by the Wellcome Centre for Human Neuroimaging for public use. Many researchers use FreeSurfer which is another popular open-source software that is developed for volume-based morphometry.

In this paper, we present a detailed and separate analysis for three main machine learning techniques namely SVM, ANN, DL on the diagnosis of AD. Section 2.1, 2.2, 2.3 provides analysis on the classification of AD using SVM, ANN, DL algorithms respectively. A brief comparison of SVM, ANN, and DL and their merits and demerits are also described in section 3.

2. CLASSIFICATION ALGORITHMS USED IN ALZHEIMER'S DISEASE

2.1 Support Vector Machine (SVM)

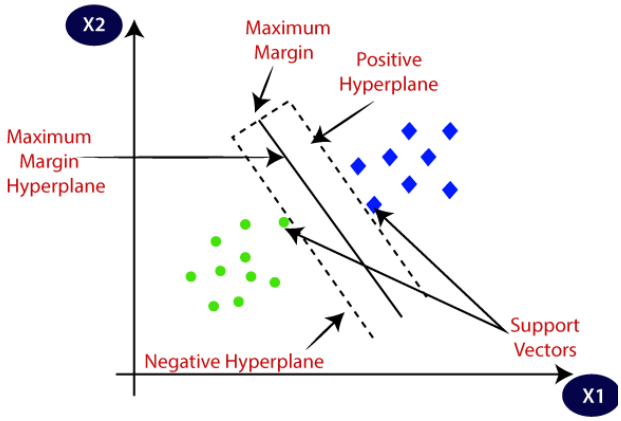


Fig -1: Support Vector Machine

Support Vector Machine is a binary, machine learning classification algorithm. It is suited for extreme cases and draws a decision boundary also known as a hyperplane. Zhu et al. [17] proposed early detection of AD using a spatial-temporal solution by recognizing the aberrant structure from a longitudinal MRI sequence. They initially compared the classification performance of the Temporally structured SVM (TS-SVM) early detection method and the standard SVM based method. Later feature selection was implemented to both TS-SVM early detection method and the standard SVM based method. Hence, the classification performance of all the methods namely, SVM, SVM+FS, TS-SVM, and TS-SVM+FS respectively were compared. The classification accuracy of TS-SVM was found to be 10% more than standard SVM, this indicates the advantage of temporal and monotony constraints. Feature selection is essential in improving accuracy in detecting AD accurately, SVM+FS and TS-SVM+FS are estimated to obtain 3.8% and 2.9% increment over SVM and TS-SVM respectively. Using TS-SVM instead of standard SVM has two major benefits in early determination and high accuracy in detection of AD: (1) Temporal consistency, (2) Early detection. Therefore, TS-SVM was implemented to predict AD at an early stage and classify longitudinal MRI of MCI converters and non-converters. 151 subjects were selected out of which 70 were MCI-C subjects and 81 MCI-NC subjects and after 10-fold cross-validation, an accuracy of 82.5% was achieved. In order to obtain high accuracy in classification, multiple classification algorithms were carried out in Matlab machine learning toolbox by Sheng et al. [18]. Classifiers with different kernel functions were trained using different features. Proper feature selection has two main ways, filter, and wrapper feature selection. The criterion for filter feature selection is independent of the particular classification algorithm, generally, estimation of statistical properties are performed using them. The 6 most common criteria are the Kruskal-Wallis test, Chi-square score, multivariate minimal redundancy maximal relevance (MRMR), Fisher score, Gini score, and relief feature score. After performing these 6 filter selections in Matlab it was concluded that the relief feature score performed the best. To reduce the number of features wrapper feature selection was implemented. More than three thousand features were processed using filter and wrapper feature selection and finally, thirty features were selected. Support vector machine (SVM: Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, Coarse Gaussian), Tree (Fine, Medium, Coarse), Discriminant (Linear, Quadratic, Logistic), Ensemble (Boosted Trees, Bagged Trees, Subspace Discriminant, Subspace KNN, RUS Boosted Trees) and KNN (Fine, Medium, Coarse, Cosine, Cubic, Weighted) were the tested classifiers. Out of all of the above classifiers, the SVM classifier yielded very accurate results. They used field mapping technique and functional MRI dataset of 96 subjects (24 CN, 24 EMCI, 24 LMCI, 24 AD) to achieve classification accuracies of 93.8% for EMCI vs. CN, 95.8% for LMCI vs. CN, 95.8% for AD vs. CN, and 91.7% for LMCI vs. AD, respectively using 5 fold cross-validation. High dimensional multi-modality imaging and genetic data were integrated by Peng et al. [19] including MRI sequence, Positron Emission Tomography (PET), and Single Nucleotide Polymorphism (SNP) genome data of 189 subjects (47 CN, 93 MCI, 49 AD). Each feature was assigned its own feature mapping. Using multiple kernel SVM accuracies of 96.1% for CN vs. AD, 80.3% for CN vs. MCI, 76.9% for MCI vs AD were observed using 10 fold cross-validation. Fludeoxyglucose (18F-FDG), Positron Emission Tomography (PET) data, and segmented Magnetic Resonance images (MRI) multi-modality datasets with Sparse Inverse Covariance Estimation (SICE) methods were used by Ortiz et al. [20]. SICE is used to analyze the undirected graphs to derive the functional and structural connectivity patterns among CN, MCI, AD. Reliable estimation of the inverse covariance is reached using SICE even when the sample size is close or less than the number of brain regions. Linear SVM was employed for 249 subjects (68 CN, 111 MCI, 70 AD) and accuracy of 92% was calculated for CN vs AD, 86% for CN vs MCI, 84% for MCI vs AD. 10-fold cross-validation was used. Zhang et al.[21], for the classification of CN, MCI, and AD patients, divided the MCI group of patients into early MCI (EMCI) and late MCI (LMCI). The division of the MCI group was done in order to test the classification capability of classifiers. The Diffusion Tensor Imaging (DTI) images were used as it estimates the location, direction, and anisotropy of White Matter (WM) regions of the brain. The AD patients show patterns of Gray matter (GM) and White Matter (WM) atrophy. WM atrophy is considered to be secondary to GM atrophy. Fractional Anisotropy (FA) and Mean Diffusivity (MD) is the most frequently used DTI metrics [42], FA is an estimation of the degree of water diffusion directionality in the tissue and MD is an estimation of gross voxel diffusion, Axial Diffusivity (DA) and Radial Diffusivity (RD) designate the diffusion coefficient that is parallel and vertical with the direction of WM region respectively. Local Diffusion Homogeneity (LDH) uses the Spearman's rank correlation coefficient (LDHs) and Kendall's coefficient concordance (LDHk) are used to quantify the comprehensive

consistency of the diffusivity series. These diffusion metrics distinguish microstructural WM properties. The preprocessed DTI data gave an idea of WM diffusion metrics. SVM and Logistic Regression (LR) algorithms are then applied to classify the four groups (CN, EMCI, LMCI, AD). The evaluation of classification performance is done using leave-one-out cross-validation (LOOCV) even when the sample size is small. LOOCV yields a good estimation in the generalization of the classifiers. Recursive Feature Elimination (RFE) method was used along with SVM to acquire the supreme feature dimension. DTI scans of 213 subjects (51 CN, 75 eMCI, 39 lMCI, and 48 AD) from ADNI dataset and achieved accuracies of 89.9% for CN vs AD, 88.1% for CN vs eMCI, 100% for CN vs lMCI, 92.98% for eMCI vs lMCI, 84.55% for eMCI vs AD, 97.7% for lMCI vs AD.

2.2 Artificial Neural Networks (ANN)

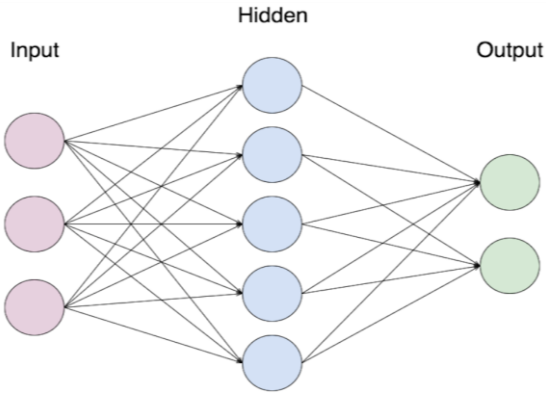
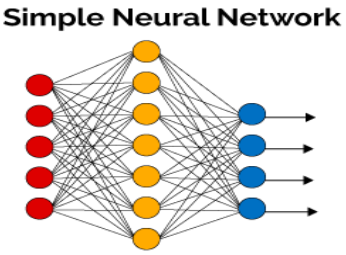


Fig -2: Artificial Neural Network

An artificial neural network is a computing system that imitates the working of the human brain i.e it mimics the ways the human brain analyzes and processes information. ANNs have input, hidden, output layers and they have the capability of self-learning, empowering them to produce better results. Gorji et al. [22] proposed efficient pseudo Zernike moments (PZMs) for the diagnosis of MCI, AD from CN. PZM is used to extract any discriminatory information from structured MRI (T1). Pattern Recognition Neural Network (PRNN), a feed-forward network, and Learning Vector Quantization neural network (LVQNN), a supervised neural network were used to classify 500 MRIs dataset (148 CN, 172 MCI, 180 AD). Accuracies of 97.27% for CN vs AD, 94.88% for MCI vs AD, 95.59% for CN vs MCI were observed using 10-fold cross-validation. Multiple Kernel Learning (MKL) is a promising group of machine learning techniques for intricate data mining tasks by integrating with multiple diverse kernel functions.[23]. MKBoost, an MKL technique that boosts the multiple kernel learning classifier for classification problems. The basic principle of MK boost is to employ a boosting method in order to learn a set of multiple base kernel classifiers, each is learnt from a single kernel. Liu et al. [24] used SVM with MKBoost to classify 710 subjects (200 AD, 120 MCIc, 160 MCInc, and 230 NC). With 10-fold validation, the accuracies observed were 95.37% for CN vs AD, 90.41% for MCI vs AD, 86.56% CN vs MCI, 73.95% for MCIc vs MCInc. Multiple different but related problems can be solved using Transfer Learning which makes use of the knowledge gained by solving one problem and then applying that knowledge in some other problem. Conventional Machine Learning algorithms use samples from a single domain and it is observed that their performance is affected when the samples are less. Correlation information from multiple domains renders crucial observations. Cheng et al.[25] in their paper proposed a Multi-Domain Transfer Learning (MDTL) framework which consists of a multi-domain transfer feature selection (MDTFS) model and a multi-domain transfer classification (MDTC) model that can aid in early detection of AD. They also compared standard SVM with least absolute shrinkage and selection operator (LASSO) [31], Multitask Feature Selection (MTFS) [32], and Manifold regularized MTFS (M2TFS) [33] and their proposed MDTFS, they found that MTFS and M2TFS and MDTFS achieves better classification performance than Lasso method because all other methods except Lasso are multi-auxiliary domains. MDTFS yields better results than MTFS and M2TFS. Using 10 fold validation and MRI of a total of 807 subjects (186 AD, 167 pMCI (progressive MCI), 228 sMCI (stable MCI), and 226 NC), accuracies of 94.7% were achieved for classification of CN vs AD, 81.5% for CN vs MCI, 73.8% for progressive MCI (pMCI) vs stable MCI (sMCI). To curtail the negative effects of unrelated source domains Cheng et al.[26] furthermore refined their methods by developing a robust multi-label transfer feature learning(rMLTFL). The framework consists of three parts, (1) image pre-processing and feature extraction, (2) rMLTFL (3) classification using baseline SVM. rMLTFL can simultaneously detect and capture common features from a multi-source domain, it also aids in identifying unrelated domains. 406 subjects(112 CN, 86 pMCI, 106 sMCI, 102 AD) were classified to achieve accuracies of 95.2% for CN vs AD, 82.4% for CN vs MCI, 76.3% for pMCI vs sMCI, 76.7% for MCI vs AD. Thus a significant increase in accuracy was observed using rMLTFL.

2.3 Deep Learning (DL)



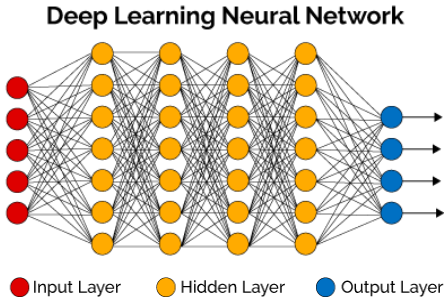


Fig. 3. *Deep Learning Network*

Deep learning is an artificial intelligence technique and a subset of machine learning that imitates the structure and function of the brain using artificial neural networks. Suk et al.[27] proposed a deep learning-based latent feature representation with a stacked auto-encoder (SAE). Deep learning is based on the concept that indicates the effectiveness of deep architectures over shallow architectures with respect to computational elements and parameters that are required to delineate unknown functions [34]. One of the most important features of deep learning is that the low-layer characterizes low-level features and the high-layer extracts those low-level features. The latent representations from the original neuroimaging and biological features are uncovered using stacked auto-encoder (SAE). The data used by them was extracted from ADNI dataset, neuroimaging data i.e the baseline MRI and 18-fluoro-deoxyglucose (FDG) PET image data and biological data i.e CSF data were acquired from 51 AD patients, 99 MCI patients among which 43 were MCI converters i.e. those patients who progressed to AD and 56 MCI non-converters i.e those patients whose progression to AD did not happen in 18 months, and 52 CN subjects. The deep architecture, in the case of neuroimaging and biological data, can be employed effectively to unfold latent or hidden representation, immanent in the low-level features from modalities, and eventually to increase the accuracy of classification. They used GM for classification as it is highly related to AD and MCI patients in comparison to WM and CSF [35]. The GM tissue volume from MRI and the mean intensity from FDG-PET were used as features for each region of interest (ROI). 93 features were obtained from MR images and the same dimensional features from FDG-PET images. They concluded and corroborated the effectiveness of the proposed method, for the classification of AD vs CN, MCI vs CN, AD vs MCI and MCI-C vs MCI-NC the accuracy was found to be 98.8%, 90.7%, 83.7%, and 83.3% respectively, using 10 fold validation. For the prediction of AD Hosseini et al. [28] proposed a 3D convolutional neural network (3D-CNN) that is pre-trained by 3D Convolutional Autoencoder (3D-CAE) in order to discover generic selective AD features in the lower layers. A few co-oriented scalar feature maps for input 3D image sets with scalar or vector voxel-wise signals are extracted using conventional unsupervised autoencoder by integrating data encoding and decoding. The autoencoder training is implemented using back-propagation and limits the feature space properties to minimize the reconstruction error, this is done for the extraction of features that encapsulate distinctive patterns of variations of input data. The pre-trained 3D-CAE updation and fine-tuning of the Deeply Supervised Adaptive 3D-CNN (DSA-3D-CNN) was performed using the Adadelta gradient descent [36]. They used structural MRI (sMRI) images of 210 subjects(70 AD,70 CN, 70 MCI) from the ADNI database for classification of 5 groups, 4 binary ones (AD vs NC, MCI vs NC, AD vs MCI, AD+MCI vs NC) and 1 ternary one (AD vs MCI vs NC). The DSA-3D-CNN was pre-trained on 30 CADDementia subjects as a source domain; the features then extracted were used for the identification of AD biomarkers in lower layers of the 3D CNN network. For classification of the group's AD vs NC, MCI vs NC, AD vs MCI, AD+MCI vs NC, and AD vs MCI vs NC the accuracy was found to be 99.3%, 94.2%, 100%, 95.7%, and 94.8% respectively. They concluded and validated the efficiency of the proposed system by differentiating AD, MCI and NC. 7 classification metrics (accuracy (ACC); specificity (SPE); sensitivity (SEN); positive predictive value (PPV), or precision; balanced accuracy (BAC); negative predictive value (NPV), and F1-score) were estimated using 10-fold cross-validation. Zheng et al.[29] proposed a stacked Deep polynomial network (S-DPN), DPN algorithm is implemented to efficiently discover feature representation from smaller samples, whereas S-DPN is supposed to enhance the feature representation. They used neuroimaging data i.e MR images and PET images from the ADNI database. As multi-modality data was used, they proposed a multi-modality S-DPN (MM-S-DPN) algorithm for fusing multi-modality neuroimaging data and learning more differentiating and powerful feature representation for classification of AD. For S-DPN, numerous 2-level basic DPNs are stacked consecutively to form a deep hierarchy; the output of the previous level of a basic DPN is fed as input to the next level of a basic DPN and every DPN is built with a 3-layer network. A two-stage S-DPN is used in the MM-S-DPN algorithm. In the first stage, two S-DPN are used to extract the high-level feature representation from MRI and PET respectively; these extracted high-level features are then fed to another S-DPN, in the second stage, in order to relate the modalities as these modalities have their unique attributes of AD. They used the same neuroimaging data from 51 AD patients and 52 CN. Segmentation of sMRI into grey matter, white matter, cerebrospinal fluid, skull-stripping, Anterior Commissure (AC) Posterior Commissure (PC) correction, and removal of the cerebellum was carried out. Each MR image was subdivided into 93 regions of interests (ROIs) using atlas warping. Furthermore, the grey matter volumes were considered as a feature for each of the 93 ROIs. Each PET image is aligned with a rigid registration with the corresponding MR image of the same person. For PET images, the average intensities of the 93 ROIs were considered as a feature. As a result, 93 features were extracted from each modality in total. The classification was carried out using 2 classifiers namely: (1) Linear SVM classifier and (2) embedded classifier. These two classifiers were used to scrutinize the accuracy of the MM-S-DPN

algorithm. The classification result of the MM-S-DPN algorithm using the SVM classifier yielded 97.27% accuracy because it coalesces MRI and PET data successfully. AD vs CN classification using MM-S-DPN with embedded classifier resulted in an accuracy of 97.27%. Zheng et al. [30] constructed a multi feature-based network (MFN) by employing sparse linear regression (LASSO) that was executed on six types of morphological features to foster structure-based auto diagnosis. They used baseline MR images of 528 subjects, 165 CN, 142 AD patients, 221 MCI patients among which 95 were stable MCI (sMCI) and 126 were progressive MCI (pMCI), from the ADNI database. These images were preprocessed using FreeSurfer, non-brain tissue removal, gray matter (GM) segmentation, reconstruction of GM/white matter boundaries, motion correction, and coordinate transformation were included while preprocessing [37,38,39]. The cross-validation process was employed to evaluate the morphological features, network connections, network properties, and their combinations. In each interaction, the nested feature selection, parameter optimization, and classifier training were employed to execute cross-validation. They applied a two-step feature selection strategy to find the preeminent subset and to minimize the overfitting risk. The feature selection strategy used were minimum redundancy and maximum relevance and SVM-based recursive feature elimination (SVM-RFE), the minimum redundancy and maximum relevance select features that are highly distinguishable and have low redundancy [40] whereas SVM-RFE is an iterative backward feature elimination strategy, it computes the ranking weights for all the features and eliminates the features that have the lowest weight [41]. As MFN is sparse, 75% of less discriminative or non-discriminative features were eliminated and the rest of them were further evaluated using SVM-RFE. The classification performance of MFN resulted in an accuracy of 96.42% for CN vs AD, 96.37% for CN vs MCI, 70.52% for MCI vs AD, and 65.61% for sMCI vs pMCI. The classification performance increased when then the connection was linked with morphological features and network properties, the accuracies of 98.70%, 97.93%, 73.82%, and 67.42% for CN vs AD, CN vs MCI, MCI vs AD and sMCI vs pMCI respectively.

3. DISCUSSION

SVM, ANN, DL algorithms for the classification of AD were discussed. Table 1 provides a summary of all the papers we have reviewed. The classification was majorly done for CN vs AD, CN vs MCI, MCI vs AD, CN vs eMCI, CN vs lMCI, etc. The main purpose of this comparison is to compare various algorithms. Depending on the dataset, modality, and feature extraction methods used the accuracy of classification varies. For cross-validation, mostly 10 fold validation was used. Each algorithm has its own merits and demerits as discussed in Table 2. SVM works efficiently when the dataset is small, ANN works efficiently when the dataset is large and also small, Deep learning works well when the dataset is large.SVM works well with both linear and non-linear data whereas ANN and DL work well with non-linear data. ANN and DL algorithms have complex implementations in comparison to SVM. SVM requires a long training period in comparison to ANN and DL.

Table -1: Overview of all algorithms and their accuracies

Algorithm	Papers	Target	Dataset	Accuracy
SVM	Zhu et al. [17]	MCIc vs MCInc	151(70 MCIc, 81 MCInc)	82.50%
	Sheng et al. [18]	eMCI vs CN	96(24 CN, 24 EMCI, 24 LMCI, 24 AD)	93.80%
		lMCI vs CN		95.80%
		AD vs CN		95.80%
		lMCI vs AD		91.70%
	Peng et al. [19]	CN vs AD	189(47 CN, 93 MCI, 49 AD)	96.10%
		CN vs MCI		80.30%
		MCI vs AD		76.90%
	Ortiz et al. [20]	CN vs AD	249 (68 CN, 111 MCI, 70 AD)	92%
		CN vs MCI		86%
		MCI vs AD		84%
	Zhang et al. [21]	CN vs AD	213 (51 CN, 75 eMCI, 39 lMCI, and 48 AD)	89.90%
		CN vs eMCI		88.10%
		CN vs lMCI		100%
		eMCI vs lMCI		92.98%
		eMCI vs AD		84.55%
		lMCI vs AD		97.70%
ANN	Gorji et al. [22]	CN vs AD	500 (148 CN, 172 MCI, 180 AD).	97.27%
		MCI vs AD		94.88%
		CN vs MCI		95.59%
	Liu et al. [24]	CN vs AD	710 (200 AD, 120 MCIc, 160 MCInc, and 230 NC).	95.37%
		MCI vs AD		90.41%
		CN vs MCI		86.56%
		MCIc vs MCInc		73.95%
	Cheng et	CN vs AD	807 (186	94.70%

	al. [25]	CN vs MCI	AD, 167 pMCI, 228 sMCI, and 226 NC)	81.50%
		pMCI vs sMCI		73.80%
	Cheng et al.[26]	CN vs AD	406 (112 CN, 86 pMCI, 106 sMCI, 102 AD)	95.20%
		CN vs MCI		82.40%
		pMCI vs sMCI		76.30%
		MCI vs AD		76.70%
DL	Suk et al. [27]	AD vs CN	202 (51 AD, 99 MCI, 43 MCIc, 56 MCInc, 52 CN)	98.80%
		MCI vs CN		90.70%
		AD vs MCI		83.70%
		MCIc vs MCInc		83.30%
	Hosseini et al. [28]	AD vs CN	210 (70 AD,70 CN, 70 MCI)	99.30%
		MCI vs CN		94.20%
		AD vs MCI		100%
		AD + MCI vs CN		95.70%
		AD vs MCI vs CN		94.80%
	Zheng et al. [29]	AD vs CN	103 (51 AD, 52 CN)	97.27%
	Zheng et al. [30]	CN vs AD	528 (165 CN, 142 AD, 95 sMCI, 126 pMCI)	98.70%
		CN vs MCI		97.93%
		MCI vs AD		73.82%
		sMCI vs pMCI		67.42%

Table -2: Merits and Demerits of algorithms

Algorithm	Merits	Demerits
SVM	• The accuracy obtained is high. • Non-Linear data can be efficiently handled using Kernel tricks. • SVM is stable.	• Large datasets take a long time. • Feature scaling is required • Large datasets are not scaled • A lot of memory is used.
ANN	• Scales well to larger and smaller datasets. • Less training is required. • Paved a way for Deep Learning.	• Interpreting the model is difficult. • They are computationally rigorous to train.
DL	• Feature engineering is not required • Unstructured data is efficiently used. • High-quality results.	• Very expensive to implement. • Expensive GPUs are required

4. CONCLUSION

Machine Learning has gained a lot of popularity and a commensurate amount of developments have paved a path of machine learning to be a major part of Health Care, detection and diagnosis of many diseases, developing new procedures aiding health care. This paper is an attempt to compare three machine learning algorithms: Support Vector Machine (SVM), Artificial Neural Networks (ANN), Deep Learning (DL), all these algorithms were used to classify various stages of Alzheimer's Disease. It is observed that more research can be done in the classification of eMCI vs lMCI and sMCI vs pMCI. It is observed that abundant research is available in the classification of CN vs AD and CN vs MCI. We aim to compare more machine learning algorithms for the detection and diagnosis of Alzheimer's Disease in our future work.

5. REFERENCES

[1] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging", Neurobiology of Aging, vol. 29, no. 4, pp. 514-523, 2008. Available: 10.1016/j.neurobiolaging.2006.11.010.

[2] P. Lodha, A. Talele and K. Degaonkar, "Diagnosis of Alzheimer's Disease Using Machine Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4, DOI: 10.1109/ICCUBEA.2018.8697386.

[3] R. Brookmeyer, E. Johnson, K. Ziegler-Graham and H. Arrighi, "Forecasting the global burden of Alzheimer's disease", Alzheimer's & Dementia, vol. 3, no. 3, pp. 186-191, 2007. Available: 10.1016/j.jalz.2007.04.381.
[4] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects", NeuroImage, vol. 104, pp. 398-412, 2015. Available: 10.1016/j.neuroimage.2014.10.002.
[5] E. Pellegrini, L. Ballerini, M. C. Valdes Hernandez, and F. Chappell, "Machine learning of neuroimaging for the assisted diagnosis of cognitive impairment and dementia: A systematic review", pp. 519 - 535, 2018. Available: https://doi.org/10.1016/j.dadm.2018.07.004.
[6] M. Termenon, M. Graña, A. Besga, J. Echeveste, and A. Gonzalez-Pinto, "Lattice independent component analysis feature selection on diffusion-weighted imaging for Alzheimer's disease classification", Neurocomputing, vol. 114, pp. 132-141, 2013. Available: 10.1016/j.neucom.2012.08.044.
[7] H. Bisgin et al., "Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles", Scientific Reports, vol. 8, no. 1, 2018. Available: 10.1038/s41598-018-24926-7.
[8] L. Mesrob, B. Magnin, O. Colliot, and M. Sarazin, "Identification of Atrophy Patterns in Alzheimer's Disease Based on SVM Feature Selection and Anatomical Parcellation", Medical Imaging and Augmented Reality, 2008.
[9] D. Veitch, M. Weiner, P. Aisen and L. Beckett, "Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative", Alzheimer's Dement, pp. 106-152, 2019. Available: 10.1016/j.jalz.2018.08.005.
[10] A. Abdulkadir, B. Mortamet, P. Vemuri, C. Jack, G. Krueger, and S. Klöppel, "Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier", NeuroImage, vol. 58, no. 3, pp. 785-792, 2011. Available: 10.1016/j.neuroimage.2011.06.029.
[11] Y. Zhang and S. Wang, "Detection of Alzheimer's disease by displacement field and machine learning", 2015. Available: 10.7717/peerj.1251.
[12] X. Guo et al., "Voxel-based assessment of gray and white matter volumes in Alzheimer's disease", Neuroscience Letters, vol. 468, no. 2, pp. 146-150, 2010. Available: 10.1016/j.neulet.2009.10.086.
[13] S. Kim et al., "Voxel-based morphometric study of brain volume changes in patients with Alzheimer's disease assessed according to the Clinical Dementia Rating score", Journal of Clinical Neuroscience, vol. 18, no. 7, pp. 916-921, 2011. Available: 10.1016/j.jocn.2010.12.019.
[14] A. Hämäläinen et al., "Voxel-based morphometry to detect brain atrophy in progressive mild cognitive impairment", NeuroImage, vol. 37, no. 4, pp. 1122-1131, 2007. Available: 10.1016/j.neuroimage.2007.06.016.
[15] C. Good, I. Johnsrude, J. Ashburner, R. Henson, K. Friston, and R. Frackowiak, "Cerebral Asymmetry and the Effects of Sex and Handedness on Brain Structure: A Voxel-Based Morphometric Analysis of 465 Normal Adult Human Brains", NeuroImage, vol. 14, no. 3, pp. 685-700, 2001. Available: 10.1006/nimg.2001.0857.
[16] J. Ashburner and K. Friston, "Voxel-Based Morphometry—The Methods", NeuroImage, vol. 11, no. 6, pp. 805-821, 2000. Available: 10.1006/nimg.2000.0582.
[17] Y. Zhu and X. Zhu, "Early Diagnosis of Alzheimer's Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine", International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 264-272, 2016. Available: 10.1007/978-3-319-46720-7_31.
[18] Sheng et al., "A novel joint HCPMMP method for automatically classifying Alzheimer's and different stage MCI patients", Behavioural Brain Research, vol. 365, pp. 210-221, 2019. Available: 10.1016/j.bbr.2019.03.004.
[19] Peng, X. Zhu, Y. Wang, L. An, and D. Shen, "Structured sparsity regularized multiple kernel learning for Alzheimer's disease diagnosis", Pattern Recognition, vol. 88, pp. 370-382, 2019. Available: 10.1016/j.patcog.2018.11.027.
[20] A. Ortiz, J. Munilla, I. Álvarez-Illán, J. Górriz, and J. Ramírez, "Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis", Frontiers in Computational Neuroscience, vol. 9, 2015. Available: 10.3389/fncom.2015.00132.
[21] Y. Zhang and S. Liu, "Individual identification using multi-metric DTI in Alzheimer's disease and mild cognitive impairment", Chinese Physics B, vol. 27, no. 8, p. 088702, 2018. Available: 10.1088/1674-1056/27/8/088702.
[22] H. Gorji and J. Haddadnia, "A novel method for early diagnosis of Alzheimer's disease based on pseudo-Zernike moment from structural MRI", Neuroscience, vol. 305, pp. 361-371, 2015. Available: 10.1016/j.neuroscience.2015.08.013.
[23] H. Xia and S. Hoi, "MKBoost: A Framework of Multiple Kernel Boosting", IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 7, pp. 1574-1586, 2013. Available: 10.1109/tkde.2012.89.
[24] J. Liu, M. Li, W. Lan, F. Wu, Y. Pan and J. Wang, "Classification of Alzheimer's Disease Using Whole Brain Hierarchical Network", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 2, pp. 624-632, 2018. Available: 10.1109/tcbb.2016.2635144.
[25] B. Cheng, M. Liu, D. Shen, Z. Li and D. Zhang, "Multi-Domain Transfer Learning for Early Diagnosis of Alzheimer's Disease", Neuroinformatics, vol. 15, no.

2, pp. 115-132, 2016. Available: 10.1007/s12021-016-9318-5.
[26] B. Cheng, M. Liu, D. Zhang, and D. Shen, "Robust multi-label transfer feature learning for early diagnosis of Alzheimer's disease", Brain Imaging and Behavior, vol. 13, no. 1, pp. 138-153, 2018. Available: 10.1007/s11682-018-9846-8.
[27] H. Suk, S. Lee and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis", Brain Structure and Function, vol. 220, no. 2, pp. 841-859, 2013. Available: 10.1007/s00429-013-0687-3.
[28] E. Hosseini and M. Ghazal, "Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network", Frontiers in Bioscience, pp. 584-596, 2018.
[29] X. Zheng, J. Shi, Y. Li, X. Liu and Q. Zhang, "Multi-modality stacked deep polynomial network-based feature learning for Alzheimer's disease diagnosis," 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, 2016, pp. 851-854, DOI: 10.1109/ISBI.2016.7493399.
[30] W. Zheng, Z. Yao, Y. Xie, J. Fan, and B. Hu, "Identification of Alzheimer's Disease and Mild Cognitive Impairment Using Networks Constructed Based on Multiple Morphological Brain Features", Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, vol. 3, no. 10, pp. 887-897, 2018. Available: 10.1016/j.bpsc.2018.06.004.
[31] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso", Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267-288, 1996. Available: 10.1111/j.2517-6161.1996.tb02080.x.
[32] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease", NeuroImage, vol. 59, no. 2, pp. 895-907, 2012. Available: 10.1016/j.neuroimage.2011.09.069.
[33] B. Jie, D. Zhang, B. Cheng and D. Shen, "Manifold regularized multitask feature learning for multimodality disease classification", Human Brain Mapping, vol. 36, no. 2, pp. 489-507, 2014. Available: 10.1002/hbm.22642.
[34] Y. Bengio, P. Lamblin and D. Popovici, "Greedy layer-wise training of deep networks", Proceedings of the 19th International Conference on Neural Information Processing Systems, pp. 153–160, 2006.
[35] M. Liu, D. Zhang and D. Shen, "Ensemble sparse classification of Alzheimer's disease", NeuroImage, vol. 60, no. 2, pp. 1106-1116, 2012. Available: 10.1016/j.neuroimage.2012.01.055.
[36] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. Vaughan, "A theory of learning from different domains", Machine Learning, vol. 79, no. 1-2, pp. 151-175, 2009. Available: 10.1007/s10994-009-5152-4.
[37] B. Fischl, A. Dale, M. Sereno, R. Tootell, and B. Rosen, "A Coordinate System for the Cortical Surface", NeuroImage, vol. 7, no. 4, p. S740, 1998. Available: 10.1016/s1053-8119(18)31573-8.
[38] A. Dale, B. Fischl, and M. Sereno, "Cortical Surface-Based Analysis", NeuroImage, vol. 9, no. 2, pp. 179-194, 1999. Available: 10.1006/nimg.1998.0395.
[39] B. Fischl, A. Liu, and A. M. Dale, "Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex," in IEEE Transactions on Medical Imaging, vol. 20, no. 1, pp. 70-80, Jan 2001, DOI: 10.1109/42.906426.
[40] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, 2005. Available: 10.1109/tpami.2005.159.
[41] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines", Machine Learning, pp. 389–422, 2002. Available: 10.1023/A:1012487302797.
[42] S. Mori and J. Zhang, "Principles of Diffusion Tensor Imaging and Its Applications to Basic Neuroscience Research", Neuron, vol. 51, no. 5, pp. 527-539, 2006. Available: 10.1016/j.neuron.2006.08.012.