

Review

Not peer-reviewed version

---

# Biomarkers of 'Late Onset Alzheimer's Disease', Diagnosis and Role of 'Deep Learning'

---

[Manash Sarma](#)\* and [Subarna Chatterjee](#)\*

Posted Date: 3 September 2024

doi: 10.20944/preprints202409.0154.v1

Keywords: Alzheimer's disease; Biomarker; Deep learning; Convolution neural network (CNN)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Biomarkers of 'Late Onset Alzheimer's Disease', Diagnosis and Role of 'Deep Learning'

Manash Sarma \* and Subarna Chatterjee

Ramaiah University of Applied Sciences, India

\* Correspondence: learnermanash@gmail.com

**Abstract:** Late-onset Alzheimer's disease (LOAD) is a dementia category disease that starts after the age of 65. Progressive problems in thinking, behaviour, learning etc. are its characteristics. Due to its progressive nature, early diagnosis is crucial. A complex pathway mechanism with different risk factors is involved in LOAD progression. A comprehensive study is done on factors causing LOAD, biomarkers used for diagnosis in clinical and preclinical stage of the disease, ML (Machine Learning) techniques to identify disease stage. We reviewed research papers on clinical and preclinical diagnosis of LOAD using ML techniques from biomarker data: MRI, PET image, MMSE, Genome / Gene expression, DNA / RNA binding sites with protein, Epigenetic expression /DNA, methylation, Speech and text. We have listed improvement scope after finding the gaps and scope of using Generative AI and SMOTE (Synthetic Minority Oversampling Technique) for data augmentation. At the end, we have elaborated current use of DL (Deep Learning) and CNN ('Convolution Neural Network') in Alzheimer's disease (AD) diagnosis.

**Keywords:** Alzheimer's disease; biomarker; deep learning; convolution neural network (CNN)

## 1. Introduction

Early diagnosis of Alzheimer's disease can be helpful for precluding the progression. "Accumulation of the protein beta-amyloid outside neurons and twisted strands of the protein tau inside neurons are hallmarks. They are accompanied by the death of neurons and damage to brain tissue. Inflammation and atrophy of brain tissue are other changes" [24]. There is loss of connection between the neurons of the brain. Hippocampus is the likely place where the problem seems to start. But as the neurons die, subsequently other brain parts also get affected. This leads to problems like short-term memory loss in the initial stages. Then happens progressive problems with loss of short-term memory initially. Next, a decline happens in other cognitive faculties followed by behavioural issues [1]. In this paper, our focus is on late onset AD (LOAD). LOAD start after age of 65. While EOAD (Early Onset AD) happens primarily because of genetic factors, multiple factors cause LOAD. 50-75% of people of age 65 years or above are generally prone to Alzheimer. As expectancy of life of people is increasing, AD patient count is also increasing all over the world.

No major cure is established till date [2]. 99.6% of failure is found in AD drugs clinical trials [3]. "Two groups were forced to end AD clinical trials as their drugs failed in preventing progression of AD" in 2018 [4]. Early diagnosis of the stage of the disease is the only viable alternative. This would help minimize progressive deterioration through preventive action. Machine learning (ML) technology is becoming a major support in early diagnosis of disease stage from biomarker datasets. ML is now widely used in disease diagnosis with increasing availability of AD datasets.

An expert's diagnosis from the biomarkers is dependent on the focus on biomarkers from experience. The approach is time consuming and usually human error prone. So, Machine Learning based disease diagnosis model is getting popularity for diagnosis of AD and other diseases. However, there are challenges. A biomarker set generally has missing data and data may be unformatted as well. Selection of the effective biomarkers from huge set of biomarkers is a different challenge. There are other challenges. LOAD has different factors / pathways leading to the diseases. It is extremely

difficult to find a proper approach to make early prediction from certain data for a disease like LOAD where complex mechanism is involved.

Depending on the factors causing LOAD and the stage of the disease, AD can manifest in different biomarkers. The genetic biomarkers, as example, can reflect any genetic cause much before the AD symptoms show up in the individual and useful in preclinical analysis while age, CSF is used in clinical analysis of the disease. In this paper, we have discussed about factors / causes of LOAD, biomarkers used for clinical and preclinical diagnosis of AD, current research on diagnosis of AD stages using ML techniques with different biomarkers data and utility of deep learning techniques for LOAD diagnosis. Primary aim of this paper is to investigate the factors of LOAD, symptoms and the associated biomarkers at clinical and preclinical stage of the disease, study of existing literature on usage ML techniques, critical analysis, discuss the prospects of DL techniques in AD diagnosis and help upcoming researchers with an introduction on the subject from different aspects.

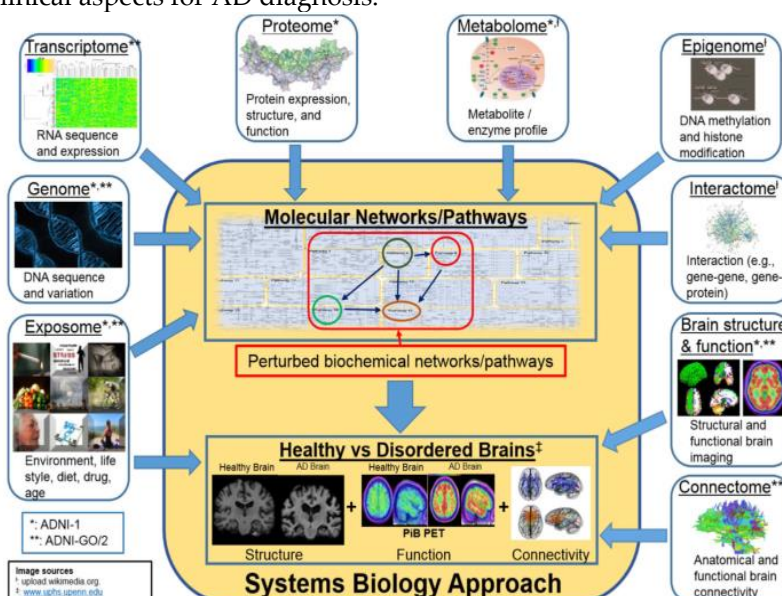
## 2. Disease Cause and Biomarkers

### 2.1. Factors Leading to LOAD and Diagnosis

A decline of cognitive function and memory primarily define AD from clinical aspect. "Histological features of AD are senile plaques, made up of accumulations of  $\beta$ -amyloid ( $A\beta$ ) peptide, and neurofibrillary tangles (NFT), which are fibrillar deposits of hyperphosphorylated Tau protein (p-Tau)" [19]. So, AD diagnosis "gold standard" is amyloid plaques and neurofibrillary tangles presence in brain. However, AD starts decades before such clinical observations.

In Figure 1, the 'System Biology Approach' block has two sub-blocks – molecular networks / pathways and healthy vs disordered brains. First one mentions how different internal and external factors can impact molecular networks / pathways that can lead to healthy or disordered brains. Both can be of help in diagnosis of AD at an early stage. The first one is about pre-clinical diagnosis where genomic or proteomic analysis can predict AD risk at a much early stage. Studies have found that the risk factors of LOAD (Late Onset AD) are age, environmental and genetics. Currently, some research is going on to find the nature of genes that may cause LOAD. However, only apolipoprotein E-4 (APOE4) is proven to carry risk of LOAD. Identifying other genes would help identify patients at risk of LOAD.

The second one is about brain structure, connectivity and functionality which are the impacts of molecular networks / pathways, the first part. The brain structure, connectivity and functionality ultimately impact the behaviour of an individual. The brain structure, connectivity, functionality, behaviours are considered as input for clinical diagnosis of the disease. This paper considers both preclinical and clinical aspects for AD diagnosis.

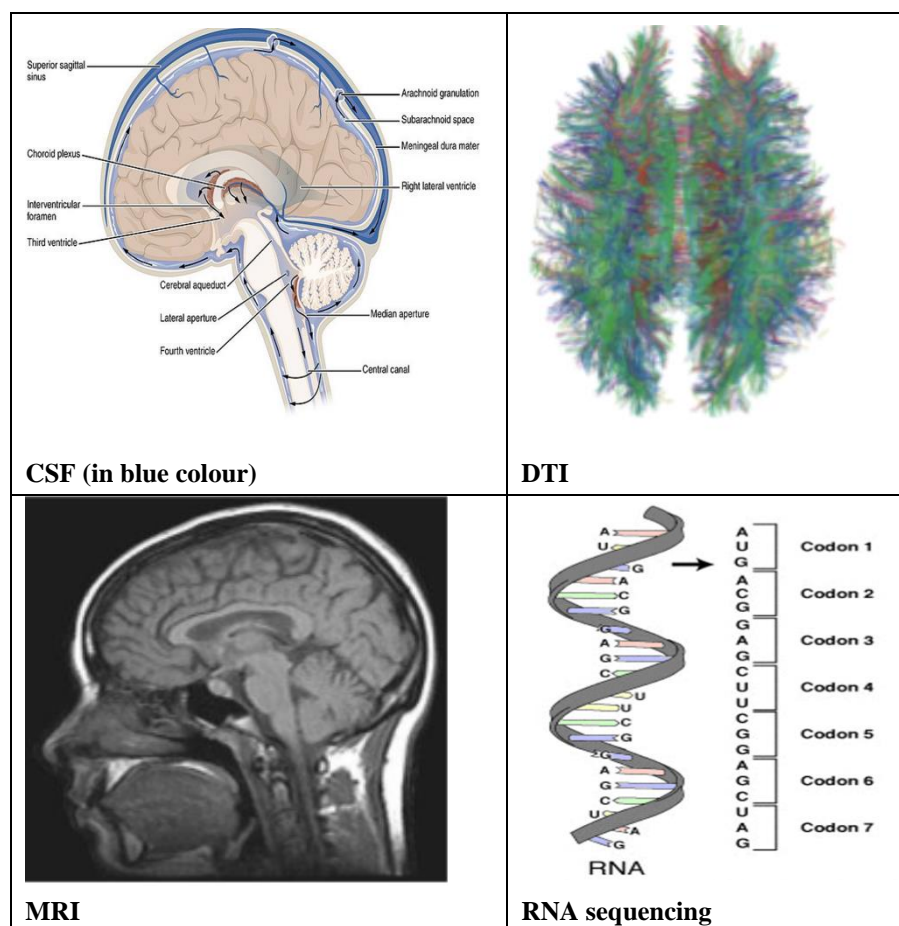


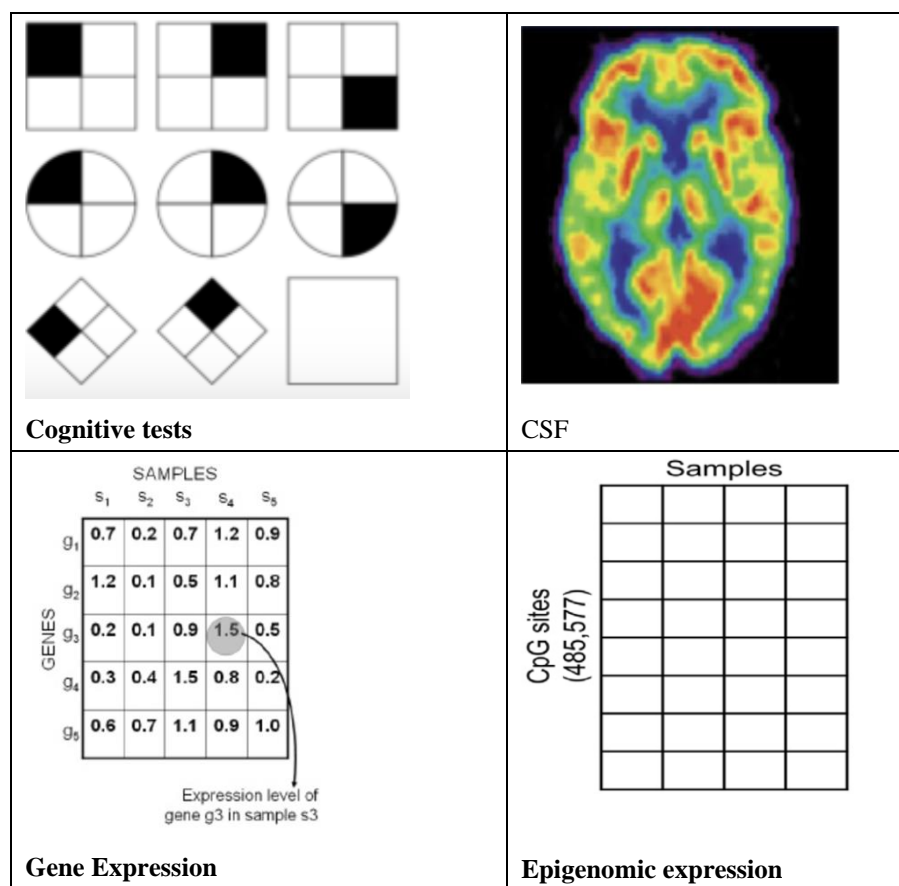
**Figure 1.** AD diagnosis – a system biology approach, source – research paper [29].

## 2.2. Alzheimer's Disease Biomarkers

Biomarkers can be defined based on the usefulness in diagnosis and prognosis of disease. As per National Institute of Health in the year 2001, "Biomarker is an indicator of certain objective measure and evaluation of biological process, pathogenic process, or pharmacological evaluation of therapeutic efficacy." As per 1998 Report of the "Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group" on "Molecular and Biochemical Markers of Alzheimer's Disease," an AD biomarker ideally has sensitivity and specificity greater than 80% while positive predictive value needs to be above 90% for detecting AD. Considering different factors leading to AD (as referred in Figure 1) and current clinical, neuro-physiological state, brain structure and functioning state for AD diagnosis, AD biomarkers (Figure 2) can be categorized as per usage in clinical and preclinical trials.

Biomarkers used for diagnosis of AD in clinical trials including disease risk identification at an early stage are Cognitive tests (like MMSE), neuroimaging biomarkers like MRIs, CSF (cerebro spinal fluid) , , DTI (diffusion tensor imaging) and blood plasma AD biomarkers. Biomarkers useful for preclinical stage are genome sequencing, gene / RNA expression profile, Protein expression profile, Epigenomic expression for DNA methylation, Interactome for gene-gene, gene-protein interaction.





**Figure 2.** AD biomarkers, source – TADPLOE.

Biomarkers used in clinical AD diagnosis reflect the current neuro-psychological state, brain structure /function, are useful for classification of AD stages - CN, MCI and AD stage. Cognitive test, patient's age, MRIs for measuring of brain structural and functional integrity, invasive CSF biomarkers to measure tau levels and amyloid, FDG PET ROI averages to measure metabolism at cell, AV45 PET ROI averages to measure amyloid-beta load in brain, DTI ROI for measuring microstructural parameters of axons and other (like APOE status, Demographic information, Diagnosis) biomarkers are mostly used for clinical use to identify stage of Alzheimer. APOE status with increasing count of APOE4 allele etc. increases the risk of late onset Alzheimer [5].

Biomarkers used in preclinical AD diagnosis are the biomarkers that can detect genetic mutations, gene expression pattern and forecast much early about the accumulation of  $\beta$ -amyloid (A $\beta$ ) peptide, and neurofibrillary tangles (NFT), synaptic dysfunction at a later stage in patient's brain leading to Alzheimer. GWAS (Genome Wide Association Study) is able to detect single nucleotide polymorphisms, or SNP variations in genome of AD patients. "Although there is evidence of an important genetic component in AD, the majority of AD is probably caused by complex interactions between one or more susceptibility genes and different environmental factors" [25]. Epigenetics is the study of impact of behaviour and environment can on the way the genes work. This mechanism is important regulators of gene expression that ultimately impact on formation of different proteins including AD causing ones.

Research has shown that speech or utterances of subjects can help differentiate between cognitively normal and AD patients by leveraging different audio features and text features (like text features like LIWC, word vectors, BERT embeddings and CLAN features).

Manual analysis of biomarkers in AD diagnosis is time consuming and is prone to human error as well. So approaches like ML based data-driven disease diagnosis is being adopted by health industry. In the next section, we are covering our study on existing research on AD detection with ML based techniques.

**Table 1.** List of biomarkers and their utility in AD diagnosis.

<b>Biomarker</b>	<b>Collection</b>	<b>Early diagnosis scope</b>	<b>Diagnosis accuracy</b>
CSF fluid	Invasive	Medium - High	High
PET scan	Non-invasive	High	Medium - High
MRI	Non-invasive	Medium	High
DTI	Non-invasive	Medium - Low	Medium
Blood Plasma	Minimal invasive	Medium	High
MMSE	Non- invasive	Medium	Medium
APOE4 count	Non- invasive	High, pre-clinical	Medium - Low
Gene expression	Minimal- invasive	High, pre-clinical	Medium
Gene Sequencing	Minimal- invasive	High, pre-clinical	Medium - Low
Speech & Text	Non-invasive	High	Medium - Low
Combinational study	Invasive, Non-invasive	High	High

We have listed few biomarkers with information about their utility/contributions towards early AD diagnosis, possible accuracy level in diagnosis, invasiveness in experimenting with the biomarker data. While genomic biomarkers are primarily useful in AD diagnosis in pre-clinical stage of the disease, other biomarkers like speech and text, MMSE, blood plasma, PET scan etc. can be useful during both clinical and pre-clinical stage.

### 3. Related Work

Plenty of has already been carried out on Alzheimer's disease diagnosis using different ML (Machine Learning) techniques with various biomarkers. We have covered study on research papers on AD diagnosis from clinical, MRI, genomic, speech & text-based biomarker data.

Some important research has been already done on AD diagnosis using ML techniques with clinical datasets. R. Chaves et al., designed a classifier for AD diagnosis with "association rule mining" [6-9] using SPECT dataset. They achieved an accuracy of 95.87% with 100% sensitivity and specificity of 92.86%. But dataset was pathologically not proven while number of total samples were low. Liu, Zhang et al., had proposed classifier with "Sparse representation type" (SRC) [11] with an intention to create local patch based sub-classifiers. To improve, they applied ensemble technique on sub-classifiers. Muehlboeck et al. did research with CSF and baseline MRI together to enhance accuracy of classification [10]. Their dataset had 96 samples of 273 CN patients and 96 AD patients. Their proposed classification had 91.8% accuracies with combination of CSF and MRI. Veeramuthu et al. worked with "Fisher Discriminants ratio" for feature extraction to get the ROI [12]. They considered a final threshold for classification of AD and CN, had achieved 91.33% accuracy, 100% specificity, 82.67% sensitivity. But, they have not discussed regarding missing data handling, handling data with imbalanced class, validation. Lawrence V. Fulton et al., designed gradient boosted machine (GBM) to predict AD presence of as a function of gender, age, education, socioeconomic status (SES), and a mini-mental-state exam (MMSE). Their ResNet-50 framework predicted the clinical dementia rating (CDR) presence and severity from MRI's. While accuracy is good, MCI stage omission in investigation and generalizability in the findings are found as limitations [13]. Eufemia Lella et al., designed an ML framework for classifying AD and analyzing feature importance. ANN, SVM, RF techniques were applied on connectivity networks from ADNI data samples collected from AD and CN natives. They achieved AUC score 83 % and an accuracy of 75 % [16]. This score is not good for binary classification. They also skipped MCI stage in classification. Sarma et al. [15] applied different machine learning techniques to predict AD stages – CN, MCI and AD and achieved the best F1 score 89% for CN, 84% for MCI, 80% for AD stage identification using deep learning. They applied "Sequential Floating Backward Selection" (SFBS) and "correlation matrix" techniques for reduce dimensionality of the data from 113 biomarkers to 8 biomarkers. DL techniques are generally sensitive to the nature of the training data. They are also sensitive to the random initialization which is often overlooked and not addressed.

J.B. Bae et al. [29] in their research identified AD from T1 weighted MRI images of medial temporal lobe. Inception-v4, a 2D image classification CNN pretrained model was used. They used two datasets ADNI and SNUBH, trained the fine-tuned the model with 156 AD patients and 156 CN controls from each dataset and tested with 39 AD patents and 39 CN from each dataset. From each dataset, 5 model instances were constructed using 5-Fold cross validation. The average ensemble values of the average probabilities of the models generated from cross-validation were used as the final results in test. They got AUC score of 0.94 with ADNI trained model with ADNI test data and AUC score of 0.91 for SNUBH trained model with SNUBH test data. For ADNI trained, SNUBH test and SNUBH trained, ADNI test data, the AUC scores are 0.88 and 0.89 respectively. S. Mohanty et al. [30] developed an attention-based image classification algorithm that accurately detects recurring odontogenic keratocyst (OKC) from whole slide images. It uses state-of-the-art transformer architecture with long-term and short-term memory. This has helped detect essential features, assisting pathologists in focusing on detecting recurring OKC. The proposed model has a high classification accuracy of .98 and a recall of 1.0. This has a Matthews correlation coefficient of 0.96. This research suggests that methods other than CNN can be explored for image classification. However, the model may not work well the case for blurred images while all hospitals or archives may not have 100 % clear images every time. So classifying blurred images or having stains is a drawback in this model. This model can be extended using different zoom-level images to perform well in all scenarios. Data augmentation techniques like GAN can be used to increase sample size in these cases.

Xiaoyan Huang et al. [19] developed an SVM based method to classify Alzheimer disease (AD) genes from gene network data of human brain and gene expression in the whole genome. Candidate genes of AD was classified with an accuracy and ROC of 84.56% and 94%. The approach could provide a supplement for wide spectrum of 20,000 AD genes extracted. Taesic Lee et al. in their study, identified AD-related genes from DEG (Differentially Expressed Genes), TF (Transcription Factor) database, gene connectivity network data, and CFG (Convergent functional genomics) from blood gene expression data [26]. The AD related gene expression data is used to construct ML models from logistic regression (LR), L1-regularized LR (L1-LR), SVM, RF, and DNN. The best average values of the AUC (area under the curve) obtained were 0.657 for ADNI, 0.874 for ANMI and 0.804 for ANM2 dataset with the five-fold cross-validation in each dataset. The results suggest that gene expression data from blood samples are useful in predicting AD stages. However, high data imbalance in ADNI data where minority dementia sample size is very less compared to other two categories, leads to poor AUC score in case of ADNI. Additionally, multiclassification is avoided here even though datasets have three labels including MCI stage. Chihyun Park et al. in their research [27] proposes a DL (deep learning) based model that can classify AD stage from integrated large-scale DNA methylation and gene expression data sets. They increased the sample size by integrated two large-scale gene expression profiles GSE33000 and GSE44770 and got accuracy of 0.823. However, they did not use the multi-omics dataset from the same sample group.

Sylvester O. Orimaye et al. [21] used features like syntax, n-gram etc. to build AD diagnostic models using SVM based ML algorithms. It learns linguistic biomarkers of verbal sound of aged persons that helps in AD clinical diagnosis. The linguistic features were from transcribed speeches from 99 control individuals and 99 AD patients of Dementia Bank dataset. However, the small size of the datasets is the main limitation. Elif Eyigoz et al. [22] used ML classification to predict AD at later stage. They used linguistic performance as early biomarker in subjects considered as cognitively normal. There are 703 samples from 270 participants in Framingham Heart Study participants. Test samples were collected when subjects were cognitively normal. An AUC value of 0.74 and accuracy of 0.70 was achieved with the use of linguistic variables. The linguistic variables which are relevant for prediction of AD onset were identified from literature as linked with cognitive decline. R'mani Haulcy et al. classified AD from non-AD and MMSE prediction based on audio and text data from ADReSS dataset. They trained five classifiers with text and audio features. The SVM classifier that is trained on BERT embedding and CLAN feature combination with PCA dimensionality reduction

technique applied, performed the best. This had accuracy of 0.898 on average. However, accuracy from speech data was poor.

### 3.1. Research Gap and Improvement Scope

Several gaps were found in earlier research and discussed. In general, lots of earlier research had to depend on pathologically unproven data because of unavailability of data. They used small datasets and generally void of the intermediate disease stage. Performance of some earlier research results looks impressive. But they were based on small unproven dataset. So, results are not much reliable. Some of the recent research have impressive result with binary classification with disease states AD and CN. It is observed that some of the researchers preferred to use binary classification by considering intermediate stage like MCI to be Dementia or AD. This way they may have achieved better result, but an important aspect of research is missed out by ignoring the intermediate stages of the disease. Some of the researchers focussed on GWAS based AD detection through identification of SNP alone. However, it is noticed that genes like APOE4 does not always cause AD. For AD with complex genetic mechanism and different pathway involved, need to consider other factors like gene expression regulation, epigenetic expression etc. in an integrated way. Another issue is noticed that researchers did not address data imbalance which is very common in real data. They used accuracy, AUC as performance measurement and omitted F1 score. F1 score is more accurate in measuring the performance of a model with all kinds of test data specially for imbalanced data.

With types of AD biomarker dataset, sample size, dimensionality different techniques need to be applied appropriately for feature selection, feature / dimensionality reduction and classification as evident from study of earlier work. With the popularity of generative AI and SMOTE (Synthetic Minority Oversampling Technique), data augmentation can be useful to train a model when sample size is small and count of minority is very less. LLM (Large Language Models) can help differentiate text / transcripts of an AD patient from a normal patient. Based on our study and gap analysis we have listed methods at broad level for AD detection and classification of disease stages.

**Table 2.** Recommended methods for AD diagnosis.

Biomarker Data sets	Feature selection Feature Reduction	/Data Imbalance with Sample size	LowSuggested supervised learning classifiers
Image (MRI / CT)	Filters (of CNN), Auto-encoder family of algorithms	Using pretrained model, Use F1 score for performance	CNN, Attention based model, pre-trained model, XGBoost, Ensemble with CNN as fusion
Non image clinical	Algorithms like SFS, Correlation Matrix	Increase weight of minority classes Use F1 score for performance	DL, SVM, RF, KNN, LR, XGBoost
Audio features: i-vectors and x-vectors	PCA	Use F1 score for performance	SVM, RF, NN, DT, Ensemble
Text features: word vectors, BERT embeddings, LIWC, CLAN	PCA	Use F1 score for performance	SVM, RF, NN, DT LLM for transcript analysis
Genome expression, Epigenetics data	Feature Ranking algorithm, like SFS, Encoding family of algorithms, selection	Merge different datasets of similar attributes / gene transcripts, Increase weight of minority classes	DL, CNN, XGBoost SVM (When sample size is low)

---

of Gene and epigeneticAugment Train data with  
data synthetic data, Use F1 score  
for performance

---

With availability of datasets both natural and synthetic way, DL (Deep learning) family of algorithms is getting popularity.

#### 4. Deep Learning in AD Diagnosis

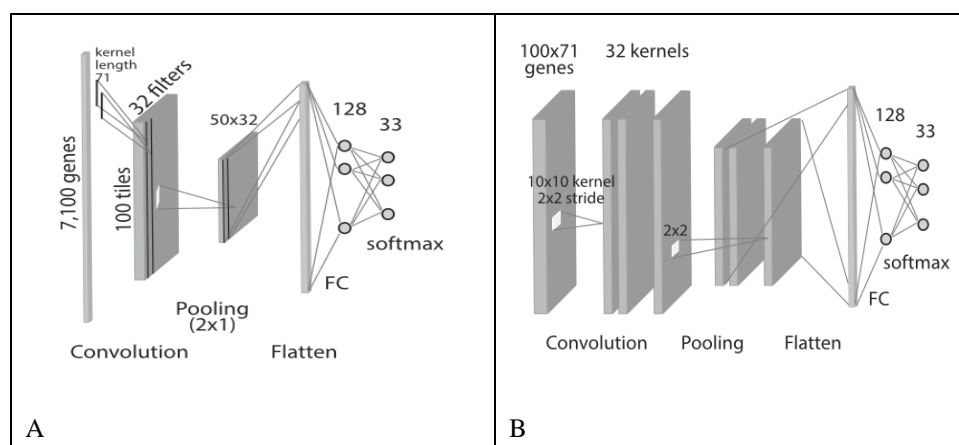
From various research work in AD detection using ML techniques, it is found that deep learning techniques produces the best result in general. Genomics etc. generates huge amounts of data used in experiments. For example, more than six billion bases are included in a single human genome sequence. Deep learning is uniquely suited to process this kind of data.

Neural networks like 'Deep learning' and CNN ('Convolution Neural Network') in general can be used as an AD classifier for neuroimaging, gene expression, DNA sequencing, TF (Transcription factor binding), sound and text data (Figure 3). Deep learning based auto-encoders can be used to transform high-dimensional data (like genomic and image) to a low dimensional data and then appropriate classifier can be used to identify the AD stage from low dimensional data.

CNN is already in use for two dimensional image classification, object recognition etc. PET, MRI image samples from patients can be input to train the AD classifier and detect AD stages accordingly. CNN can be used to predict SNP sites in a single dimensional DNA sequence (like Figure 3A) and accordingly identify AD risk genes / gene alleys [27].

Regulatory regions in single dimensional DNA, DNA / RNA binding sites with protein (TF) can be identified and predicted using CNN. TF binding sites influence gene expression of target genes.

CNN models are already in use to classify emotions, identify tones etc. from speech and text data. Accordingly, can be used to predict from verbal symptoms. Before high dimensional AD biomarkers is passed to Ad classifier for stage detection, dimensionality reduction of the data is a major step. Feature extraction and selection can be done with the help of convolution and pooling layer CNN. The reduced biomarker data set can be passed to the classifier to train the classifier and predict the AD stage accordingly. The classifier at the end is fully connected deep classifier.



**Figure 3.** Different CNN architecture diagrams – A: CNN with single dimensional input applicable for text, genomic data, B: CNN with two-dimensional input. A, B – source research paper [27].

Deep learning models can offer flexibility, proportionally scale to the amount of training data. However, they are generally sensitive to the nature of the training data. They are also sensitive to the random initialization as they learn via a stochastic training algorithm. So, there may be different set of weights every time of training, resulting different predictions. One solution to such high variance is to train multiple DL models and then combine individual predictions. This helps reduce prediction variance and improvise performance as well. An auto encoder (AE) based on deep learning based can also be used to reduce the dimension of the data before transmitting to the classifier at the end.

Using an unsupervised approach, regular auto-encoder (AE) extracts a representation of the input data which is low-dimensional.

CNN is good for handling matrix data, but many biomarkers are based on network connectivity were CNN struggles. Graph convolutional neural networks (GCNNs) is an emerging technology based on CNN. For a complex neurodegenerative disease like AD with synaptic dysfunction involved, learning of irregular graph-structured data on non-Euclidian domain is essential where GCNN can play an effective role.

## 5. Conclusion

In this paper we have covered discussion on factors causing LOAD, biomarkers like MMSE, MRI, genomic data, TF binding data, audio /text data etc. used for clinical and preclinical diagnosis of the stages of the disease and utility of ML techniques. We have covered study on literatures on usage of ML techniques for LOAD detection with different biomarker data at clinical and preclinical stages. At the end, we have mentioned the utility of DL and CNN techniques for all types of biomarker data. We hope paper will provide readers a basic understanding on cause and of symptoms of LOAD, different biomarkers used at the clinical and preclinical stage of the disease, ongoing research on AD diagnosis. It is expected that research community would understand different aspects of LOAD where complex mechanisms are involved and take up proper approach with ML based techniques for early diagnosis of the disease.

**Author Contributions:** Manash Sarma contributed to literature review, conceptualization, and original draft writing. Subarna Chatterjee contributed to reviewing and supervising.

**Conflicts of Interest:** All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

**Consent to Participate:** All authors agree to be accountable for all aspects of the work.

## References

1. Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15(3), 321-387.
2. The need for early detection and treatment in Alzheimer's disease. *EBioMedicine* 2016; 9:1
3. Cummings, JL., Morstorf, T., Zhong, K.: Alzheimer's disease drug development pipeline: few candidates, frequent failures. *Alzheimer's Res Ther* (2014)
4. Cortez M. Merck will end Alzheimer's trial as alternative approach fails. <https://www.bloomberg.com/news/articles/2018-02-13/merck-will-end-alzheimer-s-trial-as-alternative-approach-fails>
5. Schmechel, D., et al.: Association of apolipoprotein E allele  $\epsilon$ 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*.43:1467-72.;63:287-303(1993)
6. Chaves, R., J. Ramírez, et al.: Effective Diagnosis of Alzheimer's Disease by Means of Association Rules. *Hybrid Artificial Intelligence Systems*, Springer, 1, 452- 459 (2010).
7. Chaves, R., J. Górriz, et al.: Efficient mining of association rules for the early diagnosis of Alzheimer's disease. *Physics in medicine and biology* 56(18): 6047 (2011)
8. Chaves, R., J. Ramírez, et al.: Association rule- based feature selection method for Alzheimer's disease diagnosis. *Expert Systems with Applications* 39(14): 11766-11774 (2012)
9. Chaves, R., J. Ramírez, et al.: Functional brain image classification using association rules defined over discriminant regions. *Pattern Recognition Letters* 33(12): 1666-1672 (2012)
10. Westman, E., et al.: Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage* 62(1),229-238 (2012)
11. Liu, Zhang., et al.: Ensemble sparse classification of Alzheimer's disease. *Neuroimage* 60(2), 1106-1116 (2012)
12. Veeramuthu, A., et al.: A New Approach for Alzheimer's Disease Diagnosis by using Association Rule over PET Images. *International Journal of Computer Applications* 91(9), 9-14 (2014)
13. Fulton LV, Dolezel D, Harrop J, Yan Y, Fulton CP. Classification of Alzheimer's Disease with and without Imagery using Gradient Boosted Machines and ResNet-50. *Brain Sci.* 2019 Aug 22;9(9):212. doi: 10.3390/brainsci9090212. PMID: 31443556; PMCID: PMC6770938.

14. Nguyen, M., Sun, N., Alexander, DC., Feng, J., Yeo, BT.: Modeling Alzheimer's disease progression using deep recurrent neural networks. In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI), IEEE (2018), p. 1–4. <https://doi.org/10.1109/PRNI.2018.8423955>.
15. Sarma, M., Chatterjee, S. (2020). Identification and Prediction of Alzheimer Based on Biomarkers Using 'Machine Learning'. In: Bhattacharjee, A., Borgohain, S., Soni, B., Verma, G., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. [https://doi.org/10.1007/978-981-15-6318-8\\_23](https://doi.org/10.1007/978-981-15-6318-8_23)
16. Lella, E., Lombardi, A., et al.: Machine Learning and DWI Brain Communicability Networks for Alzheimer's Disease Detection (2019)
17. Kienzler, R.: The lightweight IBM Cloud Garage Method for data science. <https://developer.ibm.com/technologies/artificial-intelligence/articles/the-lightweight-ibm-cloud-garage-method-for-data-science/>
18. Wu, Q., et al.: Deep Learning Methods for Predicting Disease Status Using Genomic Data. J Biom Biostat (2018)
19. Huang, X., Liu, H., Li, X. et al. Revealing Alzheimer's disease genes spectrum in the whole-genome by machine learning. BMC Neurol 18, 5 (2018). <https://doi.org/10.1186/s12883-017-1010-3>
20. Asif M, Martiniano HFMCM, Vicente AM, Couto FM. Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. PLoS One. 2018 Dec 10;13(12):e0208626. doi: 10.1371/journal.pone.0208626. PMID: 30532199; PMCID: PMC6287949.
21. Orimaye SO, Wong JS, Golden KJ, Wong CP, Soyiri IN. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. BMC Bioinformatics. 2017 Jan 14;18(1):34. doi: 10.1186/s12859-016-1456-0. PMID: 28088191; PMCID: PMC5237556.
22. Eyigoz E, Mathur S, Santamaria M, Cecchi G, Naylor M. Linguistic markers predict onset of Alzheimer's disease. EClinicalMedicine. 2020 Oct 22;28:100583. doi: 10.1016/j.eclinm.2020.100583. PMID: 33294808; PMCID: PMC7700896.
23. Haulcy R, Glass J. Classifying Alzheimer's Disease Using Audio and Text-Based Representations of Speech. Front Psychol. 2021 Jan 15;11:624137. doi: 10.3389/fpsyg.2020.624137. PMID: 33519651; PMCID: PMC7845557.
24. (2023), 2023 Alzheimer's disease facts and figures. Alzheimer's Dement.. <https://doi.org/10.1002/alz.13016>
25. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 2005 Jul;1(1):55-66. doi: 10.1016/j.jalz.2005.06.003. PMID: 17476317; PMCID: PMC1864941.
26. Lee T, Lee H. Prediction of Alzheimer's disease using blood gene expression data. Sci Rep. 2020 Feb 26;10(1):3485. doi: 10.1038/s41598-020-60595-1. PMID: 32103140; PMCID: PMC7044318.
27. Mostavi, M., Chiu, YC., Huang, Y. et al. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics 13 (Suppl 5), 44 (2020). <https://doi.org/10.1186/s12920-020-0677-2>
28. Saykin AJ, Shen L, Yao X, Kim S, Nho K, Risacher SL, Ramanan VK, Foroud TM, Faber KM, Sarwar N, Munsie LM, Hu X, Soares HD, Potkin SG, Thompson PM, Kauwe JS, Kaddurah-Daouk R, Green RC, Toga AW, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. Alzheimers Dement. 2015 Jul;11(7):792-814. doi: 10.1016/j.jalz.2015.05.009. PMID: 26194313; PMCID: PMC4510473.
29. Bae, J.B., Lee, S., Jung, W. et al. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci Rep 10, 22252 (2020). <https://doi.org/10.1038/s41598-020-79243-9>
30. Mohanty, S.; Shivanna, D.B.; Rao, R.S.; Astekar, M.; Chandrashekar, C.; Radhakrishnan, R.; Sanjeevareddygari, S.; Kotrashetti, V.; Kumar, P. Development of Automated Risk Stratification for Sporadic Odontogenic Keratocyst Whole Slide Images with an Attention-Based Image Sequence Analyzer. Diagnostics 2023, 13, 3539. <https://doi.org/10.3390/diagnostics13233539>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.