HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 3 (2023)

Improving Alzheimer's Disease Diagnosis with Artificial Intelligence and Neuropsychological Tests

Akshita Raghuraman[1] and Emily Avery[2#]

[1]Evergreen Valley High School
[2]Yale School of Medicine
[#]Advisor

ABSTRACT

Alzheimer's Disease (AD) is a type of progressive dementia that causes loss of memory and function. It affects over 5.8 million people sixty years of age and older in the United States alone. When an individual has AD, neurodegeneration ultimately leads to severe impairment and eventually death. Throughout the stages of the disease, symptoms can range from mild forgetfulness to loss of motor control and bodily functions. Unfortunately, the disease may go unnoticed during initial stages if not tested for and detected early. Signs of Alzheimer's may accidentally be overlooked by physicians, especially if they are subtle. Artificial intelligence (AI) and machine learning models (ML) can be used to improve the identification process of early signs of AD. Some Neuropsychological impairments AD patients suffer from include memory loss, difficulty recalling words or phrases, changes in personality and behavior, and decline in bodily functions that affect day to day life. AI may be able to pick up certain details and data that human perception cannot. If trained with a proper data set, ML models would be helpful to physicians in this area of healthcare. In this systematic review, the different Neuropsychological impairments visible in the early stages of AD will be discussed as they relate to ML, deep learning, and AI applications intended to aid the diagnosis of AD.

Introduction

Dementia is a declining memory disorder affecting personality, language, attention, and social skills [1]. The most common subtype of dementia is Alzheimer's disease (AD), affecting sixty percent of individuals suffering from dementia [2]. The presence of amyloid β plaques and neurofibrillary tangles are an indicator of AD [3]. The disease is a neurodegenerative disorder that is characterized by a decline in memory, learning, and cognitive function. Affecting the elderly or individuals with head trauma, the illness in its early stages is known as mild cognitive impairment (MCI). MCI progresses to AD, which continues to worsen until all motor control and cognition have been lost, leading to death [4]. According to the World Health Organization, 55 million people are affected by dementia worldwide, and 10 million individuals are diagnosed every year. Symptoms that Alzheimer's patients experience throughout the progression of the disease include memory loss, decline in ability to participate in vital life tasks, changes in personality or behavior, and speech and language impairment [5]. The symptoms and the disease itself are irreversible, and there is no cure as of today. However, identifying the disease in its early stages and setting up patients with an appropriate care treatment plan can help slow the illness progression and provide comfort during the process.

Artificial intelligence (AI) and machine learning (ML) are popular in the field of cognitive science. Cognitive science researchers use AI to study the brain by making artificial neuron models to represent the mind [6]. Specifically for AD, AI methods can be applied to medical imaging (such as magnetic resonance imaging, MRI), genetic tests, cognitive tests, and lab-based analyses. AI can aid doctors in AD diagnosis by identifying abnormal data from Neuropsychological assessments. ML can be particularly useful for the early detection of AD. ML is a subfield in AI where models are built and "trained" to improve performance. It can be used to train programs to identify or classify

data, which is where it becomes useful for health applications [7]. Models can be trained with the results of specific cognitive and Neuropsychological tests in order to determine which patients may have or be at risk for AD.

AD diagnosis requires Neuropsychological testing to determine memory impairments. Tests used to aid diagnosis include Mini Mental State Examination (MMSE), Blessed-Orientation Mental-Concentration (BOMC) Test, and Montreal Cognitive Assessment (MOCA). These Neuropsychological tests provide quantitative diagnostic information about AD and dementia that are ripe for input into ML models. Ideally, high-performing AI and ML models can aid doctors in identifying signs of early dementia in patients.

In this review, an overview of ten prominent ML models for AD diagnosis will be provided. The processes of model training and model evaluation will be discussed along with the limitations and future applications of such models to the clinical management of AD.

Model Construction

Ten models were utilized in this analysis. Below is a description of the training data, patient population, and training methods used by each model. A summary of each model is provided in Table 1.

Model 1 [5]

In the first model of consideration, researchers in Anhui, China, developed a ML method that helped aid diagnosis by using speech and language impairment identifiers rather than standard medical imaging. Participants in the study wore an IoT (Internet of Things) device that recorded their speech, which was then stored as data and used to train a ML model into identifying fluctuation and impairment in speech. Twenty-three elderly persons between the ages of 65-92, both affected by Alzheimer's and of general health, were monitored with the help of trained models that identified changes in voice and speech that doctors would have been unable to identify. Participants were instructed to speak with the IoT device in order to collect the data. The resulting data set and the "Dem@Care" dataset were utilized to accurately train the model. Dem@Care data consisted of the following: asking subject to describe a picture while looking at it, asking subject to describe a picture from memory after looking at it, asking subject to repeat a set of sentences, asking subject to repeat pronunciation of "pa-ka-ka." The use of a spectrograph algorithm aided the researchers in their collection of data and presenting results. The algorithm was inspired by the use of MM-SPDN algorithm that learned and fused multimodal data for AD diagnosis. The model was trained to classify patients as healthy or AD.

Model 2 [8]

In the second model of consideration, researchers in South Korea used the Seoul Neuropsychological Screening Battery (SNSB) tests for their research on Neuropsychological identifiers for AD diagnosis. The SNSB consists of Neuropsychological tests that measure attention, language, visuospatial function, memory, and frontal executive function. Data from 14,926 patients from various hospitals was collected and used. TensorFlow was used to train the ML algorithm to identify different cognitive states of patients in the study (AD vs MCI vs healthy) using data collected by trained psychologists conducting the SNSB. Neural network and logistic regression models were used to model the data.

Model 3 [9]

Researchers constructing the third model of consideration used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset as well as the Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) challenge to

collect patient data. Neuropsychological tests used to identify cognitive impairment related to AD were the Alzheimer's Disease Assessment Scale, MMSE [10], Rey's Auditory Verbal Learning Test [11], and Functional Activities Questionnaire [12]. 530 patients' data were extracted from the ADNI and TADPOLE data set to create models that predicted various stages of AD. ML algorithms used to correlate data were k-nearest neighbors (kNN), decision tree (DT), rule induction, Naive Bayes, generalized linear model (GLM), and deep learning algorithm. kNN algorithm is used for classification and regression based models and uses the value of K to define a number of neighbors. DT algorithms are prediction based and used for classification in ML. The algorithm uses leaf nodes to denote class labels and branches to denote input variable combinations. Rule induction uses data mining to extract a set of rules from training data. This model primarily uses 'if-then' statements. Naive Bayes Algorithm is another data mining technique that uses probability of attributes and Gaussian probability densities to model data. GLM is used for classification and regression models and classifies data based on maximum likelihood. It is useful for models that have limited predictors, in which the predictors follow a linear relationship. Deep learning models are based on neural networks, similar to the human brain's nervous system.

Model 4 [13]

In the fourth considered model, researchers from Germany and Russia conducted Neuropsychological assessments on patients from the neurological department of the hospital of Bremen-Ost, Germany. 158 patients were split into two groups based on their MMSE scores. The early group had MMSE scores between 25-28 and the late group had MMSE scores below 25. Both groups were administered the same Neuropsychological tests, which included the Consortium to Establish a Registry for Alzheimer's Disease—Neuropathological Assessment Battery (CERAD-NAB), Wechsler Memory Scale [14], and Beck Depression Inventory [15] along with recognition trials [16, 17, 18] and tests of verbal understanding [19]. CERAD-NAB consists of the Boston Naming Test, semantic word fluency test, word list learning, figure copying, and delayed figure recall [20]. Recognition trials were carried out to determine cognitive impairments that were specific to AD diagnosis. Tests of verbal understanding were differentiated from memory function impairment to eliminate frontotemporal and subcortical dementia. SVC was used alongside a leave-one-out procedure. A binomial test was concurrently used to compare correctly classified patients according to a Neuropsychological model (gender, education, age, CERAD-NAB score predictors) with full test battery (+ additional tests).

Model 5 [21]

In the fifth model, researchers from Italy utilized the ADNI dataset to select 324 subjects for their study. Neuropsychological tests that were used included the MMSE, Clock Test [22], Logical Memory (LM) [14], Ray's Auditory Verbal Learning Test, Digit Span (DS) [14], Category Fluency Tests (Animals and Vegetables), Trail Making Test (TMT A-B) [23], Boston Naming Test, American National Adult Reading Test (ANART) [24], Alzheimer's Disease Assessment Scale-Cognitive Behavior (ADAS-Cog) [25], Geriatric Depression Scale (GDS) [26], and Functional Assessment Questionnaire. The Support Vector Machine (SVM) classification algorithm was used to create a predictive model that had the ability to perform binary group separation (normal vs MCI, MCI vs AD). The researchers fed Neuropsychological features into the classifier to obtain individual feature classification accuracy. The features were then optimized to determine the best performance of the different groups.

Model 6 [27]

In the sixth model, researchers from Washington D.C. used four ML algorithms to classify patients based on results from Neuropsychological tests. 106 patients from three different datasets were identified for study. Dataset 1 consisted of scores from nine standard Neuropsychological assessments. Dataset 2 consisted of responses and reaction times

from a spatial attention task, and Dataset 3 was a combination of Datasets 1 and 2. The ML techniques that were used were SVM, Random Forest (RF), Gradient Boosting (GB), and AdaBoost (AB). For the three experiments that were conducted, the ML algorithms were trained using the full original set of features, selected features after PCA, and original features after the feature selection. The feature selection techniques used to optimize predictive performance were SelectKBest, Sequential Forward Selection, Sequential Backward Selection, and Recursive Feature Elimination.

Model 7 [28]

In the seventh model, researchers from India compiled data from 466 subjects with the help of Neuropsychological tests. Tests that were done included the MMSE, BDIMC, COG, BOMC, MOCA, AD8, and GP CoG. The ML algorithms used included Naive Bayes, JRip, and RF. After the data set was evaluated, classification methods for the above ML algorithms were applied. The models were trained to predict healthy vs AD.

Model 8 [29]

In the eighth model, researchers from Taiwan collected data from 678 patients from the ADNI database. 226 had AD, 226 had mild cognitive impairment, and 226 were classified as clinically normal. The Neuropsychological assessments used to classify the patients into the three groups included the MMSE, CDR, ADAS, and neurophysiological scans from MRI imaging. The artificial neuron network Multilayer Perception (MLP) was used, along with SVM classifier. MLP was used as classifier and SVM was used to compare results. The models were trained to predict healthy vs MCI vs AD.

Model 9 [30]

In the ninth model, researchers used data from the ADNI dataset to construct a two-layer model using the classifier algorithm RF. Classification included multi-layer (first later) and binary (second layer). The multi-layer classification distinguished AD patients from individuals with CN and MCI. The binary second layer was used to determine the progression of AD patients in subsequent years. Data from the ADNI dataset was split into different groups including normal, mild MCI, progressive MCI, and AD. Neuropsychological tests used for the model included FAQ, MMSE, MOCA, ADAS, and RAVLT.

Model 10 [5]

In the final model, researchers utilized linguistic modeling techniques to determine indicators of AD. A subset of DementiaBank corpus from the TalkBank Project was used with 242 resulting control samples and 256 AD samples. The samples consisted of individual audio recordings and their transcriptions. The audio was analyzed for linguistic indicators of AD. Individuals were classified as “PossibleAD” and “ProbableAD,” which were later merged to form the overall AD group. Criteria that distinguished the two initial categories included symptoms of dementia, cognition deficits, and irregular laboratory results.

Table 1. Model Comparison Summary

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Neuropsychological Tests	Speech and language	Measures for attention, language visuospatial	Clinical dementia rating sum of boxes, AD	MMSE, CERAD-NAB, Wechsler Memory Scale,	MMSE, Clock Test, LM, RAVLT, DS,

	observation—change in voice and speech	function, memory, frontal executive function	cognitive scale, MMSE, RAVLT, FAQ	Beck Depression Inventory, recognition Trials, Verbal understanding	Category Fluency Tests, TMT A-B, BNT, ANART, ADAS-Cog, GDS, FAQ
ML Algorithms used	Spectrogram feature to identify fluctuation algorithm 1 algorithm 2	Tensorflow, Neural Network, Logistic regression	k-NN, DT, Rule Induction, Naïve Bayes, GLM and deep learning algorithm	SVC	SVM classification algorithm
Number of subjects	23	14,926	530	158	324
Year published	2019	2019	2019	2017	2016
Average age	65-92	Not specified	75	69.8, 71.4	75.5
Type of prediction	Binary (normal vs AD)	Binary (NC vs CI), 3-way (NC vs MCI vs AD)	5-way classification (5 stages of AD)	Binary (AD vs non-AD)	Binary (normal vs AD vs MCI)
Performance accuracy	84.4	97.6	92.75	82	92
Country of origin	China	South Korea	Worldwide	Germany	Worldwide

	Model 6	**Model 7**	**Model 8**	**Model 9**	**Model 10**
Neuropsychological Tests	Standard Neuropsychological tests (9)	MMSE, BDIMC, COG, BOMC, MOCA, ADS, GP CoG	MMSE, CDR, ADAS, MRI Imaging	FAQ, MMSE, MOCA, ADAS, RAVLT	Boston Diagnostic Aphasia Examination, Linguistic analysis
ML Algorithms used	SVM, RF, GB<AB	Naïve-Bayes, JRip, RF	MLP, SVM	RF, FURIA, DT, FRBS	Logistic regression, SVM, DT, RF, k-NN
Number of subjects	106	466	678	1048	268
Year published	2019	2015	2019	2021	2019
Average age	Not specified	65-80+	76.12, 74.26, 75.56	73.864±7.107	61.3, 71.0
Type of prediction	Binary (healthy vs AD+MCI)	Binary (healthy vs AD)	3-way (AD vs MCI vs CN)	Binary (MCI vs AD), 3-way (normal vs MCI vs AD)	Binary (normal vs AD)
Performance accuracy	91.08	100	69	94.4, 86.80	80.7
Country of origin	USA	India	Taiwan	South Korea	China

Abbreviations: AD = Alzheimer's Disease, NC = Normal Control, MCI = Mild Cognitive Impairment, MMSE = Mini Mental State Examination, RAVLT = Ray's Auditory Verbal Learning Test, FAQ = Functional Activities Questionnaire, kNN = k-nearest neighbors, DT = Decision Tree, GLM = generalized linear model, CERAD-NAB: Consortium to Establish a Registry for Alzheimer's Disease Neuropathological Assessment Battery, SVC = Support Vector Classifier, LM = Logical Memory, DS = Digit Span, BNT = Boston Naming Test, ANART = American National Adult

Reading Test, ADAS-Cog = Alzheimer's Disease Assessment Scale–Cognitive Subscale, GDS = Geriatric Depression Scale, SVM = Support Vector Machine, RF = Random Forest, GB<AB = Gradient Boosting;A/B, BOMC = Blessed Orientation Memory Concentration, MOCA = Montreal Cognitive Assessment, AD8 = Dementia Screening Interview, GP COG = General Practitioner Assessment of Cognition, CDR = Clinical Dementia Rating, MLP = Multilayer perception, FURIA = Fuzzy Unordered Rule Induction Algorithm, FRBS = Fuzzy Rule-Based Systems for Classification and Regression.

Model Evaluation

Model 1

The results of the data on Dem@Care dataset indicate LogisiticRegressionCV to have the highest accuracy, with a score of 84.4%. LinearSVC and MLP performed relatively the same, with scores of 78.1% and 77.8% respectively. Decision Tree had a score of 71.9%. The worst performing algorithm was Bagging with a score of 46.9%. Experiments were carried out with different parameter levels to optimize LogisticRegressionCV algorithm. The number of components was optimal at 300, resulting in an accuracy of 86.1%, precision of 87.5%, recall of 91.3%, and f1 score of 89.4%.

Model 2

The authors of model 2 found that neural network algorithms performed more effectively compared to logistic regression, with the 3-layer neural network algorithm performing the best. When logistic regression predicted NC vs. Cl, the clinic-based dataset had an accuracy of 88.6% and the balanced dataset had an accuracy of 85.87%. Three-way prediction of NC vs. MCI vs. AD achieved 77.3 accuracy for the clinic-based dataset and 78.95 for the balanced dataset. For neural networks, 3-layer, 4-layer, 5-layer, and 6-layer were tested, with 3-layer having the best accuracies (97.6, 97.0, 96.7, 96.9). The sensitivity of the 3-layer neural network was 96.0 and the specificity was 96.8 and 97.4 for CI and MCI respectively.

Model 3

The generalized linear model performed best with an accuracy of 92.75 in the validation dataset and 88.25 in the testing dataset when predicting [whatever the model predicted, ex AD vs healthy control]. The algorithm notably incorporated the CDRSB cognitive assessment, which notes the volume of the brain compared to the age of the patient. Due to this, the generalized linear model was able to identify instances of AD, CN, EMCI, and LMCI classes and had recalls of 100.00%, 94.44%, 90.62%, and 89.16%. The class precision values were 100%, 89.16%, 98.12%, and 79.45%. By increasing the number of EMCI and LMCI classes in the dataset, the results can be improved for better efficiency and accuracy.

Model 4

Based on the Neuropsychological tests conducted, all the patients had cognitive impairment and were separated into two groups based on TBpm ratio. The early group scored better on reading comprehension and memory, while the late group had typical responses of AD patients with cognitive impairment. Total classification accuracy was 82%, with the early group having higher accuracy of 89% and the late group 72%. The specificity decreased between the early and late group which led to decreased accuracy in results. Accuracy also increased with the addition of the new

Neuropsychological tests the researchers conducted compared to just the standard ones that were already being used for evaluation.

Model 5

The authors of model 5 found that Gaussian RBF in the classification of CDR = 1 vs CDR = 0 had the highest accuracy with 92%, 90% sensitivity, and 93% specificity. The linear model had 91% accuracy with 89% sensitivity and 92% specificity. The quadratic model had 91% accuracy, 87% sensitivity, and 92% specificity. MLP model had 91% accuracy, 87% sensitivity, and 93% specificity. In the classification of CDR = 0.5 vs CDR = 0, linear, quadratic, and Gaussian RBF models all had similar accuracy of 86% and sensitivity of 85%, while MLP had accuracy of 85% and sensitivity of 83%. Linear, Gaussian RBF, and MLP had similar specificity of 87%, while quadratic had a slightly higher 88%. In the classification of CDR = 1, vs CDR = 0.5, linear and quadratic performed the best with 65% accuracy, while Gaussian RBF and MLP had accuracies of 64% and 63% respectively. MLP had the highest sensitivity with 62%, while Gaussian RBF had 61% sensitivity and linear and quadratic both had 59% sensitivity. Linear and quadratic both had specificity of 67%, while Gaussian RBF had 65% and MLP had 63% specificity. According to the above results, the best classification performance occurred when patients with moderate cognitive impairment and normal cognitive function were discriminated

Model 6

In this model, RF had the best performance with 89.67% accuracy, 75% sensitivity, and 98% specificity. The AdaBoost classifier had the highest sensitivity overall with 78.57%. The datasets also affected the model. Using SVM with feature selection, the dataset containing scores of Neuropsychological assessments and the dataset containing responses from spatial attention tasks provided the best accuracy of 91.08% with 85.71% sensitivity and 94% specificity. The ML algorithm performed better with the Neuropsychological assessment's dataset. This indicates higher success with Neuropsychological tests as opposed to spatial attention cognitive tasks.

Model 7

In this model, Naive Bayes, Random Forest, and JRIP algorithms performed with 100% accuracy, while J48 algorithm had 98.4% accuracy. Naive Bayes, Random Forest, and JRIP algorithms all had precision and recall values of 1.000. J48 had a precision and recall value of 0.984. Naive Bayes and Random Forest algorithms took 0.03 seconds to build the model, while JRIP took 0.06 seconds and J48 took 0 seconds. Based on the above results, the authors determined Naive Bayes to be the best out of the four classification techniques.

Model 8

In this model, MLP had the best performance with 60.7% orientation accuracy, 34.3% registration accuracy, 37.3% attention accuracy, 58.7% recall accuracy, 43.8% language accuracy, 68.1% MMSE total accuracy, 69.0% orientation recall accuracy, and 62.1% orientation + recall + language accuracy. The SVM algorithm had 59.8% orientation accuracy, 35.0% registration accuracy, 36.2% attention accuracy, 58.2% recall accuracy, 41.5% language accuracy, 66.7% MMSE total accuracy, 68.4% orientation recall accuracy, and 60.3% orientation + recall + language accuracy. Orientation recall provided the best accuracy in both models. The MLP algorithm consistently performed better than the SVM algorithm. The somewhat low performance accuracy of the models was due to similar MMSE scores between the normal and MCI group. SVM performance values were lower than the MLP performance due to SVM being a

binary classifier while MLP is a multilayer classifier. Therefore, SVM cannot consider information for more than two classes at a time.

Model 9

In the ninth model, the first and second layers of RFE were based on RF, SVM, and GB. In the first layer, RF-RFE performed the best with an accuracy of 94.40%, while SVM-RFE had an accuracy value of 92.40% and GB-RFE had an accuracy value of 92.60%. RF-RFE had 28 features, SVM-RFE had 12 features, and GB-RFE had 14 features. In the second layer, RF-RFE performed the best with an accuracy of 86.80%, while SVM-RFE had an accuracy value of 82.60% and GB-RFE had accuracy value of 86.00%. RF-RFE had 36 features, SVM-RFE had 17 features, and GB-RFE had 39 features. RF was determined the best algorithm to use. The performance of the RF model was compared with SVM, KNN, Naive Bayes, and DT models. RF consistently outperformed the other algorithms with accuracy value of 87.76% and precision and recall values of 87.50%.

Model 10

In the final model, logistic regression algorithm performed the best with accuracy of 80.7%. SVM performed with similar accuracy of 79.6%, DT with 74.8%, RF with 77.6%, and kNN with 75.8%. logistic regression was used for the experiment. Using perplexity features with the logistic regression algorithm, accuracy values of 76.8% and 85.4% were obtained. The final logistic regression AD detection model had accuracy value of 85.4%.

Discussion

Based on the results from each paper, Model 7 produced the most effective ML algorithm, as it achieved accuracy values of 100% in predicting patients with AD vs healthy controls. The model's impressive performance was due in part to its precision and recall. The use of the various Neuropsychological tests as well as the efficiency of the ML algorithms used resulted in the most accurate model. Future improvements to AI and ML models may find it beneficial to base research on the models and resources used by this group of researchers. For example, using similar parameters (Classification Accuracy, Precession, Recall, Time to build model) may help future developed models be more consistent and accurate.

Conclusion

The use of AI in AD as well as healthcare in general continues to grow as more ML techniques are developed and applied to medicine. While AI may seem like the future of medical diagnosis, ML models have limitations compared to traditional clinical practice. The ability to collect data from randomized trials and train ML models with the appropriate dataset may not be feasible for all medical diagnoses due to lack of data or unusual findings. General limitations of AI that could potentially cause a significant impact on findings in medicine and healthcare include discriminatory bias, generalization of population, and errors in newly developed algorithms [32]. It would be best to combine both AI and traditional clinical practice in aiding AD diagnosis. There are signs that humans may only be able to identify, such as clinical impression perceived by a neurologist and subtle behavioral indicators that are not quantified by Neuropsychiatric testing. Combining ML with the knowledge doctors possess would produce an efficient and effective way to combat AD diagnosis and help aid in treatment of individuals who face the struggles of dementia every day.

Limitations

It is important to note that tests for certain models were conducted in the native language spoken in the country of origin. Having Alzheimer's patients speak in their first language ensures there is no discrepancy regarding learning English at a different age or not being as fluent in one language versus another. Testing patients and training models using English, or another language may alter the data and provide inaccurate predictions due to the different abilities or speaking levels that participants may have prior to developing AD.

References

Almubark, I, Chang, L. -C., Nguyen, T., Turner, R. S., & Jiang, X. (2019). Early Detection of Alzheimer's Disease Using Patient Neuropsychological and Cognitive Data and Machine Learning Techniques. *2019 IEEE International Conference on Big Data (Big Data)*, 5971–73. https://doi.org/10.1109/BigData47090.2019.9006583.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Battista, P., Salvatore, C., & Castiglioni, I. (2017). Optimizing Neuropsychological Assessments for Cognitive, Behavioral, and Functional Impairment Classification: A Machine Learning Study. *Behavioural Neurology*, Article e1850909. https://doi.org/10.1155/2017/1850909.

Beck, A. T. H., and Bailer, M. (1985). *BDI: Beck-Depressions-Inventar*. Göttingen: Hogrefe.

Bhagya Shree, S. R., and Sheshadri, H. S. (2014). An Initial Investigation in the Diagnosis of Alzheimer's Disease Using Various Classification Techniques. *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 1–5. https://doi.org/10.1109/ICCIC.2014.7238300.

Critchley, M. (1953). *The Parietal Lobes*. Hafner Publishing Company.

El-Sappagh, S., Alonso, J. M., Islam, S. M., Sultan, A. M., & Kwak, K. S. (2021). A Multilayer Multimodal Detection and Prediction Model Based on Explainable Artificial Intelligence for Alzheimer's Disease. *Scientific reports*,11(1), 1-26. https://doi.org/10.1038/s41598-021-82098-3.

Emmady, P. D., School, C., & Tadi, P. (2022). *Major Neurocognitive Disorder (Dementia)*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK557444/.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State. A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research*, 12(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6.

Gold, C. A., Marchant, N. L., Koutstaal, W., Schacter, D. L., & Budson, A. E. (2007). Conceptual fluency at test shifts recognition response bias in Alzheimer's disease: Implications for increased false recognition. *Neuropsychologia*, 45(12), 2791-2801. https://doi.org/10.1016/j.Neuropsychologia.2007.04.021.

Guo, Z., Ling, Z., & Li, Y. (2019). Detecting Alzheimer's Disease from Continuous Speech Using Language Models. *Journal of Alzheimer's Disease*, 70(4), 1163–1174. https://doi.org/10.3233/JAD-190452.

Gurevich, P., Stuke, H., Kastrup, A., Stuke, H., & Hildebrandt, H. (2017). Neuropsychological Testing and Machine Learning Distinguish Alzheimer's Disease from Other Causes for Cognitive Impairment. *Frontiers in Aging Neuroscience* ,9(114), https://www.frontiersin.org/articles/10.3389/fnagi.2017.00114.

Haldenwanger, A., Eling, P., Kastrup, A., & Hildebrandt, H. (2010). Correlation between cognitive impairment and CSF biomarkers in amnesic MCI, non-amnesic MCI, and Alzheimer's disease. *Journal of Alzheimer's Disease*, 22 (3), 971-980.

Hildebrandt, H., Haldenwanger, A., & Eling, P. (2009). False recognition helps to distinguish patients with Alzheimer's disease and amnestic MCI from patients with other kinds of dementia. *Dementia and Geriatric Cognitive Disorders*, 28(2), 159-167. https://doi.org/10.1159/000235643.

Kang, M. J., Kim, S. Y., Na, D. K., Kim, B. C., Yang, D. W., Kim, E. J., Na, H. R., Han, H. J., Lee, J., Kim, J. H., Park, K. H., Park, K. W., Han, S., Kim, S. Y., Yoon, S. J., Yoon, B., Seo, S. W., Moon, S. Y., Yang, Y., . . .Youn, Y. C. (2019). Prediction of Cognitive Impairment via Deep Learning Trained with Multi-Center Neuropsychological Test Data. *BMC Medical Informatics and Decision Making,* 19(231), https://doi.org/10.1186/s12911-019-0974-x.

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Medicine*, 17, 1-9. https://doi.org/10.1186/s12916-019-1426-2.

Kinney, J. W., Bemiller, S. M., Murtishaw, A. S., Leisgang, A. M., Salazar, A. M., & Lamb, B. T. (2018). Inflammation as a Central Mechanism in Alzheimer's Disease. *Alzheimer's & Dimentia: Translational Research & Clinical Interventions*, 4(1), 575-590. https://doi.org/10.1016/j.trci.2018.06.014.

Lee, G. G., Huang, P. -W., Xie, Y. -R., & Pai, M. -C. (2019). Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Normal Based on Neuropsychological Data via Supervised Learning. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 1808–1812. https://doi.org/10.1109/TENCON.2019.8929443.

Liu, L., Zhao, S., Chen, H., and Wang, A. (2020). A New Machine Learning Method for Identifying Alzheimer's Disease. *Simulation Modelling Practice and Theory*, 99. https://doi.org/10.1016/j.simpat.2019.102023.

Llano, D. A., Laforet, G., & Devanarayan, V. (2011). Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Disease & Associated Disorders*, 25(1), 73-84. https://doi.org/10.1097/WAD.0b013e3181f5b8d8.

Luber, S. (2011). Cognitive Science Artificial Intelligence: Simulating the Human Mind to Achieve Goals. *In 2011 3rd International Conference on Computer Research and Development*, 1, 207–210, https://doi.org/10.1109/ICCRD.2011.5764005.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Series in Computer Science.

Morris, J. C., Mohs, R. C., and Rogers, H. (1989). Consortium to establish a registry for Alzheimer's Disease (CERAD) clinical and Neuropsychological. *Psychopharmacology*, 24(4), 641-652.

Nelson, H. E., and Willison, J. (1991). *National adult reading test (NART)*. Windsor: NFER-Nelson.

Pfeffer, R. I., Kurosaki, T. T., Harrah Jr, C. H., Chance, J. M., & Filos, S. (1982). Measurement of Functional Activities in Older Adults in the Community. *Journal of Gerontology*, 37(3), 323–329, https://doi.org/10.1093/geronj/37.3.323.

Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, 8(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271.

Rey, A. (1964). *L'examen clinique en psychologie. 2e éd. Psychologue*. Presses universitaires de France.

Rohrer, J. D., Knight, W. D., Warren, J. E., Fox, N. C., Rossor, M. N., & Warren, J. D. (2008). Word-finding difficulty: a clinical analysis of the progressive aphasias. *Brain*, 131(1), 8-38. https://doi.org/10.1093/brain/awm251

Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., & Umer, A. (2019). Classification of Alzheimer's Disease Using Machine Learning Techniques. *In Proceedings of the 8th International Conference on Data Science, Technology and Applications*, 296–303. https://doi.org/10.5220/0007949902960303.

Wechsler, D. (1987). *Wechsler memory scale—Revised: Manual*. Psychological Corporation.

Yesavage, J. A., & Sheikh, J. I. (1986). 9/Geriatric Depression Scale (GDS). *Clinical Gerontologist*, 5(1–2), 165–73. https://doi.org/10.1300/J018v05n01_09.

Yiannopoulou, K. G., & Sokratis G. P. (2013). Current and Future Treatments for Alzheimer's Disease. *Therapeutic Advances in Neurological Disorders*, 6(1), 19–33. https://doi.org/10.1177/1756285612461679.