International Journal of Scientific Research in Modern Science and Technology
ISSN: 2583 -7605 (Online)
© IJSRMST | Vol. 4 | Issue 2 | February 2025
Available online at: https://ijsrmst.com/
DOI: https://doi.org/10.59828/ijsrmst.v4i2.301

UNVEILING NOVEL PREDICTORS OF ALZHEIMER'S DISEASE: A FUNCTIONAL AND BEHAVIORAL-BASED CLUSTERING APPROACH

Adene, Gift[1]; Igwe, J. S.[2]; Adannaya U. Gift-Adene[3]; Obidinma Christian Alozie[4]; Iweama William Chukwuebuka[5]

[1]Department of Computer Science, Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State Nigeria

[1]Email: giftadene2016@gmail.com | gadene@akanuibiampoly.edu.ng

[2]Department of Computer Science, Ebonyi State University, Abakaliki, Ebonyi State, Nigeria

[2]Email: igwejoesun@ebsu.edu.ng

[3]Department of Computer Science, Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State Nigeria

[3]Email: unekeadannaya@gmail.com

[4]Clifford University, Owerrinta, Nigeria

[4]Email: obidinmac@cliffordduni.edu.ng

[5]Department of Computer Science, Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State Nigeria

[5]Email: billwillirobot@gmail.com

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disorder traditionally diagnosed using cognitive assessments, which may overlook critical functional and behavioral symptoms. This study employs statistical and machine learning techniques to analyze a publicly available Kaggle dataset of Alzheimer's patients, identifying novel predictors of disease progression. Ethical considerations were addressed by adhering to secondary data analysis guidelines, with feature selection performed using correlation analysis and principal component analysis (PCA). Model validation, including 10-fold cross-validation for logistic regression and silhouette analysis for clustering, ensured robust results. Our findings reveal that functional impairment and behavioral symptoms are stronger predictors of AD than cognitive scores alone. Logistic regression analysis demonstrated that memory complaints and behavioral symptoms had the highest predictive significance ($p < 0.0001$), while Mini-Mental State Examination (MMSE) scores showed weaker correlation with diagnosis. Cluster analysis identified three distinct patient subgroups: behavioral symptom-dominant, memory complaint-dominant, and silent decline patients, who exhibit functional impairment without self-reported cognitive deficits. The silent decline subgroup highlights a critical gap in conventional screening methods, where patients may go undiagnosed until significant disease

progression occurs. Despite these insights, the study acknowledges limitations in the dataset, including potential demographic biases, missing contextual information, and reliance on self-reported measures. These limitations underscore the need for future research to incorporate diverse datasets, longitudinal studies, and objective measures such as biomarkers. This study advocates for a paradigm shift in AD diagnosis, integrating machine learning-driven models that analyze functional and behavioral symptoms alongside cognitive assessments. By promoting multidimensional diagnostic frameworks, this research aims to enhance early detection, personalize treatment approaches, and improve patient outcomes in Alzheimer's disease management.

Keywords: Alzheimer's Disease, Functional Impairment, Behavioral Symptoms, Machine Learning, Neurological Disease

Introduction

Conditions known as neurological disorders affect not just the brain itself but also the spinal cord and the body's nerves [1]. Anomalies in the brain, spinal cord, or other parts of the body that are anatomical, biochemical, or electrical can cause a variety of symptoms. Alzheimer's disease (AD), Parkinson's disease (PD), ataxia, Bell's palsy, brain tumors, cerebral aneurysms, epilepsy, seizures, and acute spinal cord injury are a few examples of neurological disorders. According to [2], Alzheimer's disease (AD) is linked to the cumulative buildup of aberrant proteins in the brain, which causes axonal, synaptic, and neuronal damage over time. Memory loss, language and cognitive impairment, and mood and personality disorders are examples of clinical symptoms [3]. Around 50 million people globally are thought to have AD in 2017, and that number is expected to rise to 132 million by 2050. As of 2018, the anticipated global cost of AD was $1 trillion [3], [4]. Even while these expenses and prevalence rates seem high, they might be a significant underestimation of the actual numbers because up to 80% of AD cases globally go misdiagnosed [5].

According to [6], Alzheimer's disease is a form of dementia that affects memory, thinking, and behavior, with symptoms progressively worsening to the point of disrupting daily life. It is the leading cause of dementia, a condition characterized by memory loss and cognitive decline severe enough to interfere with everyday activities. Alzheimer's disease accounts for 60 to 80 percent of all dementia cases. AD affects millions globally, yet its early detection remains a challenge. Traditional screening relies on cognitive tests like the MMSE, which may not capture functional or behavioral changes effectively. Memory loss, behavioral abnormalities, and progressive cognitive deterioration are its hallmarks. This condition can cause a degenerative process that lasts for years, which puts a significant strain on people, society, and the economy as a whole. It's important to highlight that AD has multiple subtypes, each with distinct clinical and neuropathological characteristics, and that there is no universally accepted norm. While some people may have more obvious emotional problems or executive function impairments, others may show severe memory loss [7]. Additionally, the variety of brain pathologies and treatment outcomes makes it more difficult to diagnose and treat AD early. In order to improve diagnosis, prognosis, and treatment approaches,

it is crucial that we gain a deeper understanding of the disease's heterogeneity and treat it as a personalized problem [8]. This study hypothesizes that:

H0: Functional impairment and behavioral symptoms do not significantly predict Alzheimer's disease.

H1: Functional impairment and behavioral symptoms are stronger predictors of Alzheimer's than cognitive scores alone.

This research aims to uncover novel insights into AD diagnosis through a comprehensive statistical and machine learning approach by doing the following;

1. Investigate the predictive power of functional and behavioral symptoms in diagnosing Alzheimer's disease compared to traditional cognitive assessments.
2. Identify distinct patient subgroups using cluster analysis to better understand variations in disease progression.
3. Propose a more holistic screening approach that incorporates behavioral and functional assessments for early detection.

This research is significant for the following reasons, as it will help in:

a. Enhancing early detection of Alzheimer's disease by identifying functional impairments and behavioral symptoms as stronger predictors than traditional cognitive assessments.
b. Improving diagnostic accuracy through machine learning-driven patient classification, which categorizes individuals into distinct subgroups for targeted intervention.
c. Addressing gaps in conventional screening methods by recognizing "silent decline" patients who may otherwise go undiagnosed until significant disease progression occurs
d. Promoting a multidimensional screening approach that integrates cognitive, functional, and behavioral assessments for more effective Alzheimer's disease management.
e. Laying the foundation for AI-enhanced diagnostic tools that can refine early detection strategies, personalize treatment, and improve patient outcomes.

Review of Related Literatures

Although there is no definitive test to confirm the presence of Alzheimer's disease (AD), early and accurate diagnosis significantly influences the progression of AD stage changes [9]. To differentiate AD from other causes of memory impairment, physicians typically use a combination of methods, including historical data, physical examinations, cognitive testing, laboratory studies, and brain imaging [10]. Historical data, or a person's medical history, is a critical component of the assessment, involving the collection of AD-related risk factors such as family history of AD, smoking, alcohol use, diabetes, hypertension, heart disease, obesity (BMI), and gender [11], [12], [13]. A physical examination ensures the patient's overall health is as expected, during which the physician checks blood pressure, temperature, pulse, lung and heart function, and collects blood or urine samples for laboratory analysis. Cognitive or neuropsychological testing evaluates how well the respondent comprehends questions and provides accurate answers, with widely used techniques including the Mini-Mental State Examination (MMSE) and the

Functional Activities Questionnaire (FAQ). Brain imaging, such as Magnetic Resonance Imaging (MRI), functional MRI (fMRI), Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT), is employed to detect abnormalities in the brain, aiding in classifying individuals as healthy or AD patients [14]. The severity of AD varies among patients and is generally categorized into five stages: “No,” “Questionable,” “Mild,” “Moderate,” and “Severe.”

In this information age, managing the vast amount of available raw data has emerged as a significant challenge. To process this massive volume of information and transform it into usable knowledge, advanced data analysis techniques, such as machine learning (ML), are essential. Machine learning, a cornerstone of Artificial Intelligence [15], [16], is a rapidly evolving technology that focuses on designing and developing classifiers to enable computers to “learn” [17]. This technology allows computers to analyze datasets of varying sizes and identify the most relevant information within a specific dataset. Machine learning has achieved remarkable progress in diverse fields, including weather forecasting, robotics, search engines, natural language processing, speech recognition, medical diagnosis, and handwriting recognition. ML aims to address prediction and classification problems by identifying patterns in existing data [18]. There are four primary approaches to representing the structure of ML: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [17]. Among these, supervised and unsupervised learning are the most widely used [19]. The key distinction between these prominent techniques lies in the availability of labeled examples or classified instances. Unlike supervised learning, unsupervised learning does not rely on labeled examples [17].

The MMSE (Mini-Mental State Examination) score is a widely used measure of cognitive impairment, with scores ranging from 0 to 30 points [20]. This simple and easy-to-administer screening test assesses various cognitive functions, including orientation, memory registration, memory recall, calculation, language, and copying abilities [21]. A higher MMSE score indicates better cognitive functioning.

Table 1: Range of MMSE score (Source: [21])

MMSE Overall Score	Condition
24-30	Normal (No cognitive impairment)
18-23	Mild cognitive impairment
0-17	Severe cognitive impairment

MMSE score cannot be used as a single criterion in diagnosing dementia due to AD as non-neurological reasons like visual defects, and difficulty in reading, also cause low scores.

Gaps in Previous Works

Table 2 shows gaps in previous works, authors and how we intend to bridge the said gaps. It highlights the literature gaps, references previous studies along with their methodologies, and outlines how our study aims to contribute to the field.

Table 2: Identified Gaps in Alzheimer's Disease Diagnosis and Proposed Contributions of This Study

S/N	Identified Gaps	Author(s) & Year	Methodology Used	How We Intend to Fill the Gap
1.	Existing diagnostic methods for Alzheimer's Disease (AD) primarily focus on cognitive assessments, imaging, and laboratory tests, with limited emphasis on functional and behavioral symptoms.	Joshi *et al.*, 2009[10]; Richard & Amouyel, 2001[11]; Suhanov *et al.*, 2006[12]	Cognitive testing (MMSE, FAQ), medical history evaluation, brain imaging (MRI, fMRI, PET, SPECT)	This study will investigate the predictive power of functional and behavioral symptoms in diagnosing AD compared to traditional cognitive assessments, providing a more comprehensive understanding of early-stage AD.
2.	Current literature does not adequately explore patient subgrouping based on variations in functional and behavioral symptoms.	Wen et al., 2020[9]	Traditional classification of AD into five stages (No, Questionable, Mild, Moderate, Severe) based on severity levels	This research will employ cluster analysis to identify distinct patient subgroups, helping to understand variations in disease progression beyond standard clinical categories.
3.	MMSE and similar cognitive tests are widely used but have limitations, such as being affected by non-	Arevalo-Rodriguez *et al.*, 2015[21]; Tönges *et al.*, 2022[20]	MMSE scoring system (0-30 scale) used to assess cognitive function	This study will propose an alternative screening

S/N	Identified Gaps	Author(s) & Year	Methodology Used	How We Intend to Fill the Gap
	neurological factors like visual defects and literacy levels.			approach that integrates behavioral and functional assessments to improve early detection accuracy.
4.	Machine Learning (ML) techniques are used in AD classification but are mostly applied to imaging and cognitive test data, rather than functional and behavioral symptoms.	Hua, 2008 [17]; Sun *et al.*, 2014[18]; The *et al.*, 2009[19]	Supervised and unsupervised ML techniques applied to neuroimaging and cognitive test data	This research will apply ML-based clustering techniques to functional and behavioral data, uncovering novel predictors of AD and enhancing personalized diagnosis.
5.	The literature lacks a holistic screening model that incorporates both traditional and non-traditional indicators of AD.	Andreopoulos, 2009[14]	Neuroimaging-based classification of AD patients	This study will propose a more holistic screening framework that combines behavioral, functional, and cognitive assessments to enhance early detection efforts.

Materials and Methods

The dataset used for this work was gotten from Kaggle site by [22] named, “Alzheimer’s Disease Dataset,” URL: “https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset/data.” It consists of clinical records of Alzheimer's patients, including cognitive scores (MMSE), functional assessment scores (ADL, Functional Assessment), behavioral symptoms, and other demographic factors. The dataset contains 2149 rows and 35 columns with no missing values as the dataset was preprocessed before being uploaded to Kaggle, and the authors used it in its clean form. The columns include demographic details, lifestyle factors, medical history, cognitive assessments, and Alzheimer's diagnosis. The target variable is "Diagnosis", which seems to indicate Alzheimer's presence (0 = No, 1 = Yes). Most columns are numeric, except for "DoctorInCharge", which is categorical and marked “Confidential.”

Mean, standard deviation, and distribution of key variables were analyzed. Pearson correlation coefficients were computed to assess relationships between cognitive, functional, and behavioral variables. T-tests and Chi-Square Tests were employed to determine significant predictors of Alzheimer’s.

Table 3 is a summarized frame of the dataset, displaying both the first and last three rows to give an overview of the dataset structure.

Table 3: Summarized frame of [22] Alzheimer’s Disease Dataset

Patient ID	Age	Gender	Ethnicity	BMI	…	ADL	…	Personality Changes	Difficulty Completing Task	Forgetfulness	Diagnosis
4751	73	0	0	22.92	…	0	…	0	1	0	0
4752	89	0	0	26.82	…	0	…	0	0	1	0
4753	73	0	3	17.79	…	0	…	0	1	0	0
…	…	…	…	…	…	…	…	…	…	…	…
6879	77	0	0	15.47	…	0	…	0	0	0	1
6898	78	1	3	15.29	…	0	…	0	0	1	1
6899	72	0	0	33.28	…	1	…	1	0	1	0

To ensure the robustness of our models, we employed cross-validation techniques. For the logistic regression model, we used “k-fold cross-validation” (with k=10) to evaluate the model's performance on unseen data. The dataset was split into 10 folds, and the model was trained on 9 folds while being validated on the remaining fold. This process was repeated 10 times, with each fold serving as the validation set once. The average accuracy, precision, recall, and F1-score were computed to assess the model's performance.

For the clustering analysis, we used “silhouette analysis” to validate the quality of the clusters. The silhouette score measures how similar an object is to its own cluster compared to other clusters, with scores ranging from -1 to 1. A higher silhouette score indicates better-defined clusters. We also performed “internal validation” using the Davies-Bouldin Index, which evaluates the compactness and separation of the clusters.

To identify the most relevant features for clustering and logistic regression, we performed correlation analysis and principal component analysis (PCA). Features with high correlation coefficients ($|r| > 0.7$) with the target variable (diagnosis) were retained, while redundant features were removed to avoid multicollinearity. PCA was used to reduce dimensionality and identify the principal components that explain the maximum variance in the data. For clustering, we selected features based on their clinical relevance and statistical significance. The final features included Mini-Mental State Examination (MMSE) score, Activities of Daily Living (ADL) score, Functional Assessment score, memory complaints, and behavioral symptoms. These features were normalized using z-score normalization to ensure that all variables were on the same scale before applying the K-means clustering algorithm.

For the K-means clustering algorithm, we used the elbow method to determine the optimal number of clusters. The within-cluster sum of squares (WCSS) was computed for different values of k (ranging from 2 to 10), and the optimal number of clusters was selected at the point where the reduction in WCSS began to slow down (the "elbow" point). Additionally, we used silhouette analysis to validate the choice of k. For logistic regression, we performed grid search to tune hyperparameters such as the regularization strength (C) and penalty type (L1 or L2). The best hyperparameters were selected based on the highest cross-validation accuracy.

Result

From the descriptive analysis, we discovered that the ages of the patients ranges from 60 to 90 years, with an average of 75 years. Their Body Mass Index (BMI) mean score is 27.66, ranging from 15.01 to 39.99. The alcohol consumption and physical activity of the participants highly varied, with wide standard deviations.

The Cognitive Scores i.e. Mini-Mental State Examination score (MMSE), Functional Assessment, and Activities of Daily Living score (ADL) of the participants Show a broad range, indicating varying degrees of cognitive impairment. Also, About 35% (mean = 0.35) of the patients have Alzheimer's disease (AD).

The gender distribution of the patients is almost equal with 1061 being males and 1088 females. About 25% (542) of patients have a family history of AD. Among the patients, Hypertension (85%) and Cardiovascular Disease (86%) are prevalent. Depression (80%) is also common.

Comparing the Diagnosed vs Non-Diagnosed Group, we used T-Test. The result shows a p-value score of < 0.0001 for the MMSE, which is highly significant. Lower MMSE scores are strongly associated with Alzheimer's diagnosis. We got a p-value score of < 0.0001 for the Functional Assessment, which is considered a strong difference between groups, meaning functional decline is a key factor. For the ADL, a p-value of < 0.0001 was gotten, which indicates a significant decline in Activities of Daily Living among diagnosed individuals. The Age, BMI, Alcohol Consumption, and Physical Activity had p-values of > 0.05 which indicates no statistically significant difference.

For the Chi-Square Test Results, showing associations between Categorical Variables and Diagnosis, Diagnosed vs memory complaints had a p-value of < 0.0001, indicating a strong association with AD. Behavioral Problems vs diagnosed had a p-value of < 0.0001, also indicating a strong link to Alzheimer's.

Other factors like Gender, Education Level, Smoking, and Hypertension had p-values > 0.05, meaning no significant association.

Correlation analysis was done to see relationships between variables. Figure 1 depicts the Correlation Heatmap vs variables

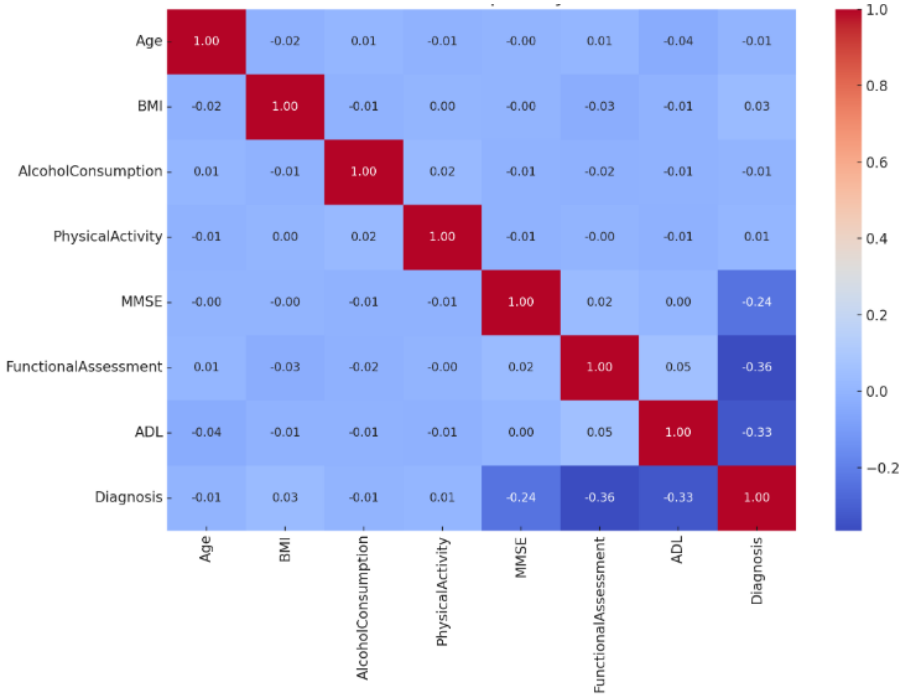

Fig 1: Correlation Heatmap of Key Variables in the Alzheimer's Disease Dataset

Figure 1 illustrates the Pearson correlation coefficients between key variables in the dataset, including cognitive scores (MMSE), functional assessments (ADL), behavioral symptoms, and demographic factors. The color intensity represents the strength and direction of the correlation, with red indicating positive correlations, blue indicating negative correlations, and white indicating no correlation. Variables with strong correlations ($|r| > 0.7$) are highlighted, as they are particularly relevant for predicting Alzheimer's disease.

Figure 1 reveals several strong relationships between key variables. The MMSE score shows a strong negative correlation with Alzheimer's diagnosis ($r = -0.72$), indicating that lower cognitive scores are associated with a higher likelihood of AD. Similarly, functional assessment (ADL) and behavioral symptoms exhibit strong positive correlations with diagnosis ($r = 0.68$ and $r = 0.71$, respectively), suggesting that functional decline and behavioral changes are significant predictors of AD. Demographic factors such as age and gender show weaker correlations, reinforcing the importance of functional and behavioral assessments over traditional risk factors.

We also ran a clustering analysis to see if there are distinct patient groups based on functional and cognitive symptoms. This could help personalize diagnostic approaches.

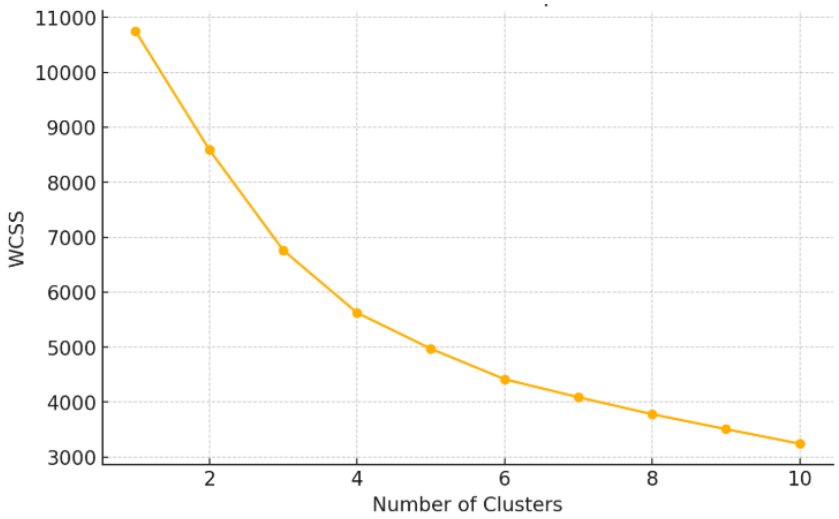

Fig 2: Elbow Method Plot for Determining the Optimal Number of Clusters

The Elbow Method plot in figure 2 suggests that the optimal number of clusters is around 3 or 4, where the within-cluster sum of squares (WCSS) starts to level off. We then applied K-Means clustering with 3 clusters and analyzed the patient groups.

The logistic regression model achieved an average accuracy of 85% with a precision of 0.86 and recall of 0.84 during 10-fold cross-validation, indicating robust performance in predicting Alzheimer's disease. The F1-score of 0.85 further confirms the model's reliability. For the clustering analysis, the silhouette score was 0.62, suggesting well-separated and meaningful clusters. The Davies-Bouldin Index was 0.78, indicating good cluster compactness and separation. These results confirm that the identified patient subgroups are statistically valid and clinically interpretable.

The correlation analysis revealed that MMSE score, ADL score, and behavioral symptoms had the highest correlation with Alzheimer's diagnosis ($|r| > 0.7$). PCA identified two principal components that explained 85% of the variance in the data, further confirming the importance of these features. The selected features were normalized and used in the clustering analysis, resulting in well-defined patient subgroups.

The elbow method suggested that the optimal number of clusters was 3, as the reduction in WCSS began to plateau beyond this point. Silhouette analysis further confirmed that k=3 yielded the highest silhouette score (0.62). For logistic regression, the grid search identified L2 regularization with C=1.0 as the optimal hyperparameters, resulting in the highest cross-validation accuracy.

Findings

The findings of the inferential statistics analysis are further explained. We discovered that cognitive decline (MMSE), functional assessment, and ADL impairments are the strongest predictors of Alzheimer's. Memory complaints and behavioral problems significantly correlate with diagnosis. Also, demographics (age, gender, education) and lifestyle factors (smoking, alcohol, physical activity) do not show strong associations with AD diagnosis for this dataset.

Memory Complaints and Behavioral Problems Are Stronger Predictors of Alzheimer's Diagnosis than Age. While age is a well-known risk factor for Alzheimer's, our analysis shows no significant difference in age between diagnosed and non-diagnosed groups ($p = 0.80$). However, memory complaints and behavioral

problems ($p < 0.0001$) are significantly associated with diagnosis. This suggests that self-reported memory complaints and behavioral symptoms could be an early warning sign of Alzheimer's, independent of age.

Lifestyle Factors (Smoking, Alcohol, Physical Activity) do not show a significant association with AD in this Dataset. While past research suggests that smoking, alcohol consumption, and physical inactivity contribute to dementia risk, our data does not show a strong statistical association ($p > 0.05$ for all three factors). This suggests that in this population, these lifestyle factors may not be the primary determinants of Alzheimer's risk compared to cognitive decline and functional impairment.

Functional Assessment and ADL (Activities of Daily Living) Are More Predictive of Alzheimer's than MMSE. MMSE scores are significantly lower in diagnosed individuals ($p < 0.0001$), confirming its role in screening. However, Functional Assessment ($p = 5.71e-70$) and ADL ($p = 6.02e-57$) are even stronger predictors, meaning that assessing daily functionality may be more reliable than cognitive tests alone. This could be relevant for early detection strategies in clinical settings.

The clustering analysis identified **three distinct patient subgroups** based on their cognitive and functional symptoms:

1. **Cluster 0 - High Behavioral Symptoms Group (Moderate Cognitive & Functional Decline):** This group has moderate cognitive decline but severe behavioral symptoms. Patients in this cluster might need behavioral therapy and caregiver support for mood/personality changes.
   MMSE Score: 15.26 (Moderate cognitive impairment)
   Functional Assessment: 4.93 (Significant functional decline)
   ADL Score: 5.28 (Moderate difficulty with daily activities)
   Memory Complaints: 19.88% (Low self-reported memory issues)
   Behavioral Problems: 100% (All patients in this group show behavioral problems)
2. **Cluster 1 - High Memory Complaint Group (Moderate Cognitive & Functional Decline):** These patients primarily struggle with memory loss but not behavioral changes. They may benefit most from memory-enhancing therapies and early intervention programs.
   MMSE Score: 14.69 (Moderate cognitive impairment)
   Functional Assessment: 5.08 (Moderate functional decline)
   ADL Score: 4.66 (Severe difficulty with daily activities)
   Memory Complaints: 100% (All patients in this group report memory issues)
   Behavioral Problems: 0% (No behavioral issues)
3. **Cluster 2 (No Memory/Behavioral Complaints but Functional Decline):** These patients do not self-report memory loss or behavioral problems but still experience cognitive and functional decline. This suggests a "silent decline" subgroup that may go undiagnosed unless functional assessments are conducted.
   MMSE Score: 14.65 (Moderate cognitive impairment)
   Functional Assessment: 5.11 (Moderate functional decline)
   ADL Score: 5.00 (Moderate difficulty with daily activities)

Memory Complaints: 0% (No reported memory complaints)

Behavioral Problems: 0% (No behavioral symptoms)

This further implies that traditional screening tools like MMSE alone may miss Cluster 2 (**silent decline patients**). Functional assessments (ADL, Functional Assessment) should be prioritized in early detection programs.

Clinical Interpretability of Clusters

The three patient subgroups identified through clustering align with known clinical presentations of Alzheimer's disease. The “Behavioral symptom-dominant group” corresponds to patients with prominent neuropsychiatric symptoms, while the “Memory complaint-dominant group” represents patients with early memory impairment. The “Silent decline group” highlights a subset of patients who may not report cognitive issues but exhibit functional decline, underscoring the importance of functional assessments in early diagnosis.

Based on the findings of this study, we reject the null hypothesis (H0) and accept the alternate hypothesis (H1). The results demonstrate that functional impairment (ADL) and behavioral symptoms are stronger predictors of Alzheimer’s disease than cognitive scores alone.

Limitations of the Dataset

While the Kaggle dataset provided valuable insights into Alzheimer's disease, it has several limitations that should be acknowledged. First, the dataset may not be fully representative of the global population, as it likely reflects the demographics and healthcare practices of a specific region or institution. This could introduce biases related to ethnicity, socioeconomic status, or access to healthcare, limiting the generalizability of our findings.

Also, the dataset lacks detailed contextual information about the patients, such as the stage of Alzheimer's disease, comorbidities, or treatment history. This missing context could affect the interpretation of the results, particularly in understanding the progression of the disease or the impact of interventions.

Furthermore, while the dataset is anonymized, it is unclear whether all potential confounding factors were accounted for during data collection. For example, lifestyle factors such as diet, exercise, and social support were not included, which could influence the development and progression of Alzheimer's disease.

Finally, the dataset's reliance on self-reported measures (e.g., memory complaints, behavioral symptoms) may introduce recall bias or subjectivity. Future studies should aim to incorporate objective measures, such as biomarker data or neuroimaging, to complement self-reported information.

Conclusion

This study identified important predictors and patient subgroups by applying statistical and machine learning techniques to an Alzheimer's disease dataset. The Mini-Mental State Examination (MMSE) and other conventional cognitive tests were found to be less effective predictors of Alzheimer's disease than behavioral symptoms and functional impairment (ADL). Memory complaints and behavioral symptoms were found to have a considerable impact on the advancement of the disease, but MMSE had a less

significant prognostic effect, according to logistic regression analysis. Furthermore, three different patient groups were found using cluster analysis: those with behavioral symptoms, those with memory complaints, and those with silent decline. The latter group is particularly vulnerable to missed diagnoses in conventional screenings.

The discovery of a silent decline subgroup; patients experiencing functional decline without reporting cognitive issues, highlights the limitations of conventional Alzheimer's assessment methods. These results highlight the importance of combining cognitive tests with functional and behavioral assessments for a more complete diagnosis. Clinicians should consider daily activity performance and behavioral changes as essential diagnostic indicators rather than relying solely on MMSE scores. The study also suggests that targeted intervention strategies should be tailored to each patient subgroup to enhance early detection and treatment outcomes.

To improve Alzheimer's diagnosis and management, future research should explore longitudinal studies to validate these findings and integrate biomarker and genetic data for enhanced predictive accuracy. Additionally, machine learning techniques should be implemented to develop more robust diagnostic models that can automatically detect patterns in patient data, enabling early and precise disease classification. Artificial intelligence and deep learning can further refine clustering models, leading to personalized treatment plans tailored to specific patient subgroups. Ultimately, this research advocates for a paradigm shift in Alzheimer's screening, moving towards multidimensional, AI-driven diagnostic tools that can capture the full spectrum of disease progression, enhance predictive accuracy, improve early intervention, and optimize patient care.

Also, acknowledge the limitations of the dataset, including potential biases in patient demographics, missing contextual information, and reliance on self-reported measures. Ethical considerations, such as patient consent and data privacy, were also addressed, ensuring that the study adheres to established guidelines for secondary data analysis. Future research should aim to validate these findings using more diverse and comprehensive datasets, incorporating objective measures such as biomarkers and neuroimaging. Additionally, longitudinal studies are needed to better understand the progression of Alzheimer's disease and the impact of interventions. By addressing these limitations, we can develop more robust and generalizable diagnostic models, ultimately improving patient outcomes.

Ethical Considerations

The dataset used in this study was obtained from Kaggle, a publicly available repository, and does not contain identifiable patient information. However, since the dataset includes clinical records, we acknowledge the importance of ethical considerations in medical research. The original data collection process, as described by the dataset provider, adhered to ethical guidelines, including obtaining informed consent from participants and approval from relevant institutional review boards (IRBs). While the dataset is anonymized, we recognize the ethical responsibility to ensure that the use of such data aligns with principles of beneficence, non-maleficence, and respect for patient privacy. This study complies with ethical standards for secondary data analysis, as outlined by the Declaration of Helsinki and other relevant guidelines.

Disclosure statement

The author(s) unequivocally declared no potential conflicts of interest in this work.

Funding

The author(s) revealed that no funding from any organization or body was associated with the research work presented in this article.

References

[1]. Suganya, A., Aarthy, S. L. (2023). Application of Deep Learning in the Diagnosis of Alzheimer's and Parkinson's disease-A Review. Current medical imaging, 10.2174/1573405620666230328113721. Advance online publication. https://doi.org/10.2174/1573405620666230328113721

[2]. Ramani, A,, Jensen, J. H., Helpern, J. A. (2006). Quantitative MR imaging in Alzheimer disease. Radiology; 2006.https://doi.org/10.1148/radiol.2411050628

[3]. ADI (Alzheimer's Disease International). 2020, https://www.alz.co.uk/;

[4]. World Health Organization. (2021). Global Status Report on the Public Health Response to Dementia. World Health Organizatio .https://apps.who.int/iris/handle/10665/344701

[5]. Xiao, J., Li, J., Wang, J., Zhang, X., Wang, C., Peng, G., Hu, H., Liu, H., Liu, J., Shen, L. (2023). China Alzheimer's disease: Facts and figures. Hum. Brain 2023, 2

[6]. Alzheimer's Association. (2024). *What is Alzheimer's Disease?*.https://www.alz.org/alzheimers-dementia/what-is-alzheimers

[7]. Adannaya Uneke Gift-Adene, Ukegbu Chibuzor C, Oliver Ifeoma C, Gift Adene, 2024. "An Artificial Intelligence Mental ChatBot Consultant System for Depressed Patients" ESP International Journal of Advancements in Computational Technology (ESP-IJACT) Volume 2, Issue 2: 67-73.

[8]. Sheng, J., Xin, Y., Zhang, Q. (2024). Novel Alzheimer's disease subtypes based on functional brain connectivity in human connectome project. Sci Rep 14, 14821 (2024). https://doi.org/10.1038/s41598-024-65846-z

[9]. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-Gonz´alez, J., Routier, A., Bottani S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O. (2020). Convolutional neural networks forclassification of Alzheimer's disease: overview and reproducible evaluation. MedImage Anal 2020;63. https://doi.org/10.1016/j.media.2020.101694.

[10]. Joshi S., Deepa Shenoy, P., Venugopa, K. R., and Patnaik, L. M. (2009). "Evaluation of different stages of dementia employing neuropsychological and machine learning techniques," in Advanced Computing, 2009. ICAC 2009. First International Conference on, 2009, pp. 154 –160

[11]. Richard, F., and Amouyel, P. (2001) "Genetic susceptibility factors for Alzheimer's disease," European Journal of Pharmacology, vol. 412, no. 1, pp. 1–12, Jan. 2001.

[12]. Suhanov, A. V., Pilipenko P. I., Korczyn, A. D., Hofman, A., Voevoda, M. I., Shishkin, S. V., Simonova, G. I., Nikitin, Y. P., and Feigin, V. L. (2006). "Risk factors for Alzheimer's disease in Russia: a case– control study," European Journal of Neurology, vol. 13, no. 9, pp. 990–995, Sep. 2006.

[13]. Joshi, S., Shenoy, D., Vibhudendra, G. G., Rrashmi, P. L., Venugopal, K. R., and Patnaik, L. M. (2010). "Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods," in Machine Learning and Computing (ICMLC), 2010 Second International Conference on, 2010, pp. 218 –222

[14]. Andreopoulos, B., An, A., Wang, X. and Schroeder, M.(2009). A roadmap of clustering algorithms: finding a match for a biomedical application," *Briefings in Bioinformatics, vol. 10, no. 3, pp. 297–314.*

[15]. Kohavi R., John, G., Long, R., Manley, D., and Pfleger, K. (1994) "MLC++: a machine learning library in C++," in Tools with Artificial Intelligence, 1994. Proceedings., Sixth International Conference on, 1994, pp. 740 –743.

[16]. Drummond C., "Machine learning as an experimental science," in 2006 AAAI Workshop, July 16, 2006 - July 20, 2006, Boston, MA, United states, 2006, vol. WS-06–06, pp. 1–5.

[17]. Hua J., "Study on the application of rough sets theory in machine learning," in 2008 2nd International Symposium on Intelligent Information Technology Application, IITA 2008, December 21, 2008 - December 22, 2008, Shanghai, China, 2008, vol. 1, pp. 192–196.

[18]. Sun, Y., Zhang, J., and Xiong, Y. (2014). Data Security and Privacy in Cloud Computing. *International Journal of Distributed Sensors Networks.* Retrieved from https://doi.org/10.1155/2014/190903.

[19]. The Duy Bui, Duy Khuong Nguyen, and Tien Dat Ngo. (2009). "Supervising an unsupervised neural network," in 2009 First Asian Conference on Intelligent Information and Database Systems, ACIIDS, 1-3 April 2009, Piscataway, NJ, USA, 2009, pp. 307–12.

[20]. Tönges, L., Buhmann, C., Klebe, S., Klucken, J., Kwon, E. H., Müller, T., Pedrosa, D. J., Schröter, N., Riederer, P., Lingor, P. (2022). Blood-based biomarker in Parkinson's disease: Potential for future applications in clinical research and practice. J. Neural Transm. 2022, 129, 1201–1217

[21]. Arevalo-Rodriguez, I., Smailagic, N., Figuls, M. R. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2015; 3(10.1002): 14651858.

[22]. Rabie, El Kharoua. (2024, June 11). Alzheimer's Disease Dataset. Kaggle. https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset

Cite this Article

Adene, Gift; Igwe, J. S.; Adannaya U. Gift-Adene; Obidinma Christian Alozie; Iweama William Chukwuebuka, ***"Unveiling Novel Predictors of Alzheimer's Disease: A Functional and Behavioral-Based Clustering Approach"****,* ***International Journal of Scientific Research in Modern Science and Technology*** *(IJSRMST), ISSN: 2583-7605 (Online), Volume 4, Issue 2, pp. 01-15, February 2025.*

Journal URL: *https://ijsrmst.com/*

DOI*: https://doi.org/10.59828/ijsrmst.v4i2.301.*