A Regression Model For Alzheimer's Disease Progression Using The ADNI Database

Samuele Russo[1], Simone Tondi[2] and Valerio Ponzi[2]

[1]Neuroimaging Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy
[2]Institute for Systems Analysis and Computer Science, Italian National Research Council, Rome, Italy

Abstract

Alzheimer's disease is a neurodegenerative disorder and the most common form of dementia. It affects approximately 50 million people worldwide, and to date, no definitive cure has been found. As one of the leading causes of death among individuals over the age of 65, early diagnosis is crucial, as it can significantly improve life expectancy and quality of life. In recent years, numerous machine learning techniques have been applied to various biomarkers to support the early detection of the disease. The objective of this project is to conduct an in-depth analysis of the ADNI database in order to study the characteristics of individuals affected by Alzheimer's at different stages of the disease, using machine learning methods. The results of this study demonstrated that it is possible to distinguish four distinct stages of Alzheimer's progression—from cognitively healthy individuals to those severely affected—rather than the commonly discussed three. Notably, the analysis also revealed that women are disproportionately impacted by the disease, accounting for nearly 80% of the affected population.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder and the most common cause of dementia, accounting for approximately 50–70% of all diagnosed dementia cases. It can severely compromise an individual's ability to perform daily activities and alter their personality, as it affects brain regions responsible for memory, language, and cognitive function. In 2020, over 55 million people worldwide were affected by the disease, and projections estimate this number will rise to 139 million by 2050 [1].

The most frequently observed symptoms of AD include memory loss and behavioral changes, both of which are linked to the accumulation of specific biological substances [2]. These substances progressively disrupt neuronal function, ultimately leading to widespread brain atrophy. One of the most severely affected regions is the hippocampus, the brain structure responsible for the formation of new memories. This degeneration is caused primarily by the abnormal behavior of tau proteins and β-amyloid plaques, which form deposits in the brain and contribute to neuronal impairment.

Because of the irreversible damage caused to the neuronal population, Alzheimer's is a progressive and incurable condition that worsens over time. Despite its widespread impact, no definitive cure exists to date. However, early diagnosis is critical, as it allows for interventions that may slow the progression of the disease. Current treatments include medications that can help manage symptoms and maintain the patient's functional independence for longer. Moreover, lifestyle interventions such as physical activity and cognitive exercises have also been shown to delay the onset of brain damage.

Given the slow and progressive nature of the disease, longitudinal research is essential to better understand its development over time. Longitudinal studies are particularly valuable in Alzheimer's research because they involve repeated evaluations of the same individuals, allowing researchers to detect subtle changes and trends as the disease progresses.

Among the most prominent longitudinal resources available for Alzheimer's research is the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. One of the main objectives of ADNI is the early detection of Alzheimer's disease and the identification of reliable biomarkers to monitor its progression.

The ADNI study involves volunteers aged 55 to 90, recruited from various research sites across the United States and Canada. Launched in 2004, the study has evolved through several phases, each with specific objectives and participant cohorts. The first phase, ADNI-1, aimed to develop biomarkers for use as outcome measures in clinical trials. The cohort consisted of 200 cognitively normal elderly individuals, 400 participants with Mild Cognitive Impairment (MCI), and 200 patients diagnosed with Alzheimer's Disease. The second phase, ADNI-GO, focused on identifying biomarkers at earlier stages of the disease. It expanded the ADNI-1 cohort by including an additional 200 participants diagnosed with early MCI. The third phase, ADNI-2, built on previous efforts by refining biomarkers as predictors of cognitive decline and as clinical trial endpoints. This phase incorporated participants from both ADNI-1 and ADNI-GO, and enrolled an additional 150 cognitively normal individuals, 100 with early MCI, 150 with late MCI, and 150 with more advanced mild cognitive symptoms. Finally,

SYSTEM 2025: 11th Sapienza Yearly Symposium of Technology, Engineering and Mathematics. Rome, June 4-6, 2025
s.russo@hsantalucia.it (S. Russo)

ADNI-3 emphasized the use of tau PET imaging and other advanced functional imaging techniques in clinical trials. It continued with the ADNI-2 cohort and added 133 more cognitively normal elderly participants.

The dataset is composed of the results obtained from a series of tests that participants undergo at multiple time points during the study. These tests fall into two main categories: clinical tests and cognitive tests. Clinical tests are exploratory examinations that gather biological and physical data, aiming to confirm or exclude specific diagnoses and to detect potential anomalies in the patient's condition. Cognitive tests, on the other hand, are designed to evaluate mental functioning and to assess how the brain processes information. They usually consist of simple tasks or questions that the participant is required to complete, offering insight into memory, attention, language skills, and other cognitive domains.

In this project, we will use the ADNI database as it is one of the biggest and most used in Alzheimer's research. We will also analyze its longitudinal data in order to predict the evolution of the disease, to prevent further deterioration and to take earlier clinical and cognitive actions. In this context, several machine learning approaches have been applied to extract patterns from ADNI data, with particular success in classification problems. Among these, neural networks [3]have shown promising results. For example, probabilistic neural networks (PNNs) [4] have been used to classify patients with Alzheimer's versus those with mild cognitive impairment or normal cognition. Other models, such as elliptical basis neural networks (EBNNs) [5], and other hybrid neural network based systems [6] have been applied to model more complex decision boundaries and to integrate imaging data with neuropsychological scores.

Moreover, recent studies have started exploring the link between Alzheimer's disease and deficits in theory of mind [7], the capacity to understand others' beliefs, emotions, and intentions. This aspect, often investigated in social cognition, is found to deteriorate early in Alzheimer's and can provide sensitive behavioral markers of cognitive decline. In parallel, the role of sustained attention [8] has gained increasing attention in the literature, since impairments in maintaining focus over time are commonly observed in early phases of the disease. These deficits, measurable through specific cognitive tasks, may serve as early indicators and help in designing targeted interventions aimed at preserving attentional control.

2. Related Works

As there will be some acronyms that will appear throughout the literature review, it is possible to check their explanation in the table 1.

Alzheimer's disease represents a major challenge in the field of neurophysiological research, as no cure has yet been discovered, and its progression varies significantly from person to person. AD is typically diagnosed with high accuracy only in its later stages, which is why recent studies have focused on detecting the disease at earlier phases. One of the most recent contributions to this line of research [9] highlights ongoing efforts toward adopting a continuous model of disease progression.

Table 1
Acronyms

Acronym	Explanation
AD	Alzheimer's Disease
MCI	Mild Cognitive Impairment
CN	Cognitively Normal
ADNI	Alzheimer's Disease Neuroimaging Initiative
KNN	K Nearest Neighbours
NAN	Not A Number
RMSE	Root Mean Square Error
MMSE	Mini-Mental State Examination
ADAS	Alzheimer's Disease Assessment Scale
RAVLT	Rey Auditory Verbal Learning Test
TRABSCOR	Time to complete Part B of the Trail Making Test
FAQ	Functional Activities Questionnaire
CDR	Clinical Dementia Rating Scale

Until a few years ago, based on the degree of brain damage, it was common to characterize Alzheimer's progression using three main stages. The first stage, referred to as *Cognitively Normal*, includes individuals whose brains show no structural or functional damage and whose brain volume remains within the average range. The second stage, known as *Mild Cognitive Impairment* (MCI), corresponds to the early stages of Alzheimer's development. It is considered a transitional phase between normal aging and dementia-related decline and is typically associated with memory and language difficulties. The final stage is *Alzheimer's Disease*, in which the brain is clearly affected by the disease, both structurally and functionally.

As with many other illnesses, the known nature of Alzheimer's is primarily biological, involving both structural and molecular changes in the brain. Therefore, the study of AD relies heavily on the analysis of biomarkers. These biomarkers represent objective indicators of a patient's medical state and are characterized by their accuracy and reproducibility [10], making them essential for identifying the different stages of the disease.

According to the methodology described in [11], AD biomarkers can be broadly categorized into two groups. The first group includes imaging-derived biomarkers, which are extracted from techniques such as MRI or PET scans. The second group comprises biochemical biomarkers derived from cerebrospinal fluid (CSF), which reflect the molecular composition and alterations occurring in the brain. Both categories are crucial for understanding the progression of Alzheimer's and for improving

early-stage detection.

In order to define a more continuous approach to the disease, a deep understanding of the different biomarkers is necessary, as some of them may vary significantly over the years due to age-related decline, while others may only manifest when the disease is in its later stages or actively progressing. Therefore, it is essential to understand the nature of different biomarkers prior to conducting the research.

Since Alzheimer's is a dynamic disease that primarily affects the brain and its neuronal population, numerous studies such as [12] and [13] have focused on identifying early stages of the disease using imaging-derived biomarkers. However, these abnormalities generally become apparent only in the later stages of Alzheimer's, meaning that MRI scans may appear normal during the early phases. Moreover, conventional machine learning techniques have shown limited effectiveness due to their reliance on expert users for complex feature extraction. These earlier studies aimed to build models that analyze anatomical and structural brain images obtained from MRIs, as well as assess brain function to identify defects and abnormalities. Additionally, training these models often required extensive image partitioning, increasing the time and complexity of the process. As a result, deep learning techniques have been increasingly adopted in such cases [14].

On the other hand, novel biomarkers such as cerebrospinal fluid (CSF) concentrations of β-amyloid, also analyzable via PET imaging, have gained attention [9]. When combined with traditional neuropsychological assessments, these biomarkers can better define disease progression. This evolution can then be compared to that of healthy individuals over time, offering a more accurate characterization than a single imaging result. By analyzing the progression of CSF biomarkers, it becomes possible to determine whether a patient's decline is faster than that of a healthy individual, which could support a continuous model of disease progression and potentially enable earlier detection, before reaching severe stages.

Although medical imaging is a valuable tool for Alzheimer's analysis—offering detailed views of brain volume, it remains a resource-intensive technique, as patients must undergo complex scanning procedures. In contrast, CSF biomarkers may offer a more accessible and interpretable alternative for continuous monitoring.

Today, thanks to technological advancements, particularly in artificial intelligence, computer-aided diagnosis systems have been developed [15]. These systems assist in the detection and diagnosis of medical data, serving as a "second opinion" for healthcare professionals. In Alzheimer's disease, CAD systems are mainly applied to medical image interpretation [16]. However, data processing becomes increasingly complex when dealing with high-dimensional feature spaces [17]. Consequently, machine learning algorithms are often preferred, as they allow for dimensionality reduction. Most studies applying ML algorithms to AD rely on supervised learning approaches to classify patients using either neuroimaging data or biomarkers, as shown in [18] and [19].

Based on this literature review, we observe Alzheimer's disease has been predominantly studied through medical imaging and classification algorithms. This led us to our research direction: focusing on the use of non-imaging biomarkers from the longitudinal ADNI dataset. Our analysis will employ unsupervised machine learning algorithms to group patients with similar medical conditions and subsequently examine the temporal evolution of key biomarkers.

3. Methodology

In this project, we aim to classify Alzheimer's disease into distinct stages, identify the most relevant characteristics associated with each stage, and perform a temporal analysis of how various parameters evolve. Our objective is to find common traits among patients who are at the same stage of the disease. Since the progression of Alzheimer's is not uniform across individuals, we decided to group patients based on their clinical test results, without relying on predefined diagnostic labels. For this reason, we work with unlabeled data, meaning that the final diagnosis of each patient is not used in the analysis.

Among the available unsupervised machine learning algorithms for classification, we selected the K-Means algorithm, as it allows us to partition patients according to similarities in their clinical profiles. However, K-Means is known to be sensitive to the choice of the number of clusters. To address this issue and optimize the algorithm's performance, we applied the Elbow Method [20] to determine the most appropriate number of clusters.

The algorithm was executed for up to 20 clusters, and the results showed a clear inflection point at $K = 4$, where further increases in the number of clusters led to only marginal improvements in performance. Therefore, we selected four clusters for the analysis. The corresponding Elbow graph can be seen in Figure 1.

Once the number of clusters has been defined, we apply the K-Means algorithm, which assigns each patient to the cluster with the closest centroid. The algorithm operates as follows: first, it initializes or updates the centroids; then, it calculates the distance between each patient and all centroids; each patient is assigned to the nearest centroid. If, during this process, any patient changes cluster, the algorithm restarts until convergence is reached.

Working with high-dimensional data increases the complexity of the analysis. Although we reduced the number of features by removing those with insufficient data, the dimensionality remained significant, with 95

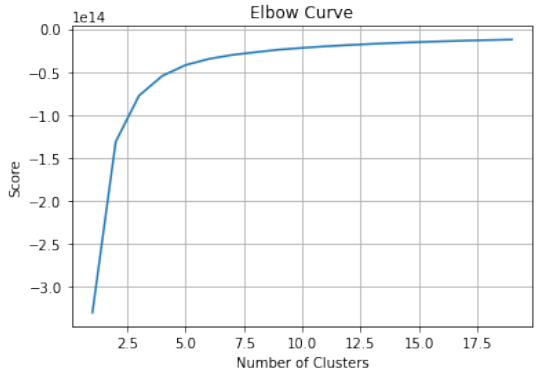

Figure 1: Elbow curve for K-Means' K selection

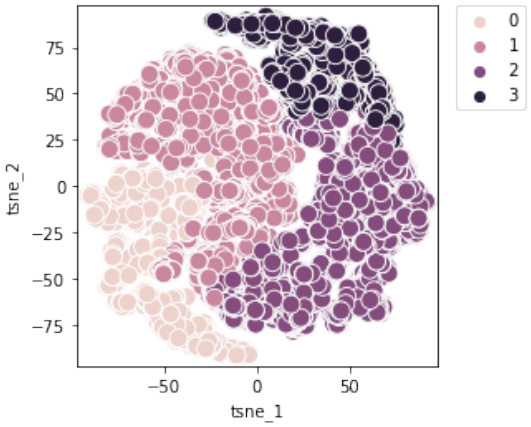

Figure 2: T-SNE cluster results

features still under consideration. Therefore, we considered the implementation of a dimensionality reduction algorithm such as PCA or t-SNE.

Following the conclusion of [21], which highlights that the t-SNE algorithm performs better when dealing with non-linear structures—unlike PCA, which is a linear algorithm—we decided to implement the t-SNE algorithm. The main characteristic of this algorithm is its ability to preserve clusters from high-dimensional spaces even after reduction to lower dimensions [22].

After applying the algorithm, the dataset's dimensionality was reduced to two components. The resulting clusters can be observed in Figure 2.

After analyzing these results, although the outcomes from feature reduction were satisfactory, as it was possible to group patients into four clusters—the main issue with this type of algorithm (and probably with most dimensionality reduction methods) is the loss of physical information. However, in this project, where the goal is to identify similar characteristics or features among patients at the same stage of Alzheimer's, dimensionality reduction does not provide us with meaningful insights. Therefore, we decided not to include feature reduction algorithms such as t-SNE or PCA in our analysis.

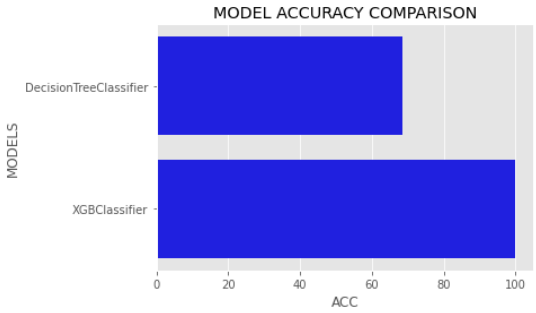

Figure 3: Regression models comparison

Table 2
Model Accuracy

	Tree Classifier	XGBoost
Accuracy	0.685	1.000
Cross Validation	0.685	0.999
Mean Squared Error	0.920	0.0

Once we identified the different stages of the disease, it became possible to train classification models that can assign new patient data to one of the defined stages. This classification provides insight into the patient's current condition and allows for a prediction of disease progression, as we can estimate how many stages remain before reaching the most severe form. The algorithms used for this task are the Regression Tree Classifier and the XGBoost Classifier.

We selected these two algorithms because, although they process data differently, both are capable of classifying new instances into the defined stages. Boosting algorithms are based on the idea of creating highly accurate prediction rules by combining multiple weak and imprecise rules. eXtreme Gradient Boosting (XGBoost) combines decision trees with gradient boosting to minimize execution time and maximize efficiency. On the other hand, regression trees integrate decision-making and prediction to classify new data effectively.

Figure 3 shows the accuracy differences when both models are applied to the same dataset, and Table 3 confirms that the XGBoost algorithm performs better in classifying new cases into the identified disease stages.

Table 3
Errors from KNN

K	RMSE Error
3	0.2182
5	0.2112
7	0.2100
9	0.2088
11	0.2088
13	0.2087
15	0.2111
17	0.2096

4. Dataset and Treatments

The database used for the analysis was ADNI, specifically the ADNIMERGE dataset, which contains data from 2,419 different patients (1,150 females and 1,269 males) and includes 115 features for each patient. From this dataset, certain features or patients lacked sufficient data for proper analysis, so we removed those features whose proportion of missing values (NaNs) was 80% or higher.

After removing these features, we applied a supervised machine learning model—specifically, the K-Nearest Neighbors (KNN) algorithm—to fill in the remaining missing values in the dataset. The choice of this algorithm aimed to provide more accurate estimations for the missing values by imputing them based on the values of the most similar patients. Unlike the K in K-Means, in KNN the parameter K represents the number of neighbors considered for the imputation. To achieve greater accuracy, we computed the Root Mean Square Error (RMSE) for different values of K and selected $K = 13$ as it yielded the minimum error, with RMSE$(K = 13) = 0.2087$. All computed errors are presented in Table 3.

The pseudocode of the KNN algorithm is as follows:

1. Find the Euclidean distance to all training data points
2. Sort each distance
3. Select the first k values
4. Assign the value based on the selected points

In this project, we used the `KNNImputer` from the `sklearn` library in Python to complete the missing values using the K-Nearest Neighbors algorithm.

Once the dataset was completed, we carried out a feature analysis to identify the most relevant variables for the study. The features considered include categorical information such as age, gender, education, and ethnicity, as well as results from cognitive assessments like the Clinical Dementia Rating (CDR) and Functional Activities Questionnaire (FAQ). In addition, clinical test scores such as MMSE, ADAS, RAVLT, and TRABSCOR were analyzed to evaluate cognitive impairment levels.

Table 4
Alzheimer's level

Alzheimer's level	Description
0	Cognitively Normal
1	First stages in Alzheimer
2	Mild Cognitive Impairment
3	Alzheimer's Disease

These are the main features considered for the analysis and will serve as input for the K-Means algorithm. However, since our goal is to partition the patients based solely on their clinical data, we excluded categorical and socio-demographic features from the clustering process. Nevertheless, these features were analyzed after the clusters were formed and the patients were assigned to each cluster, in order to gain further insights into the characteristics of each group.

5. Experiments and Results

The implementation of the K-Means algorithm on the processed dataset allowed us to identify four Alzheimer's stages, which are presented in Table 4. Having defined these four stages of the disease, and considering that Alzheimer's progressively worsens over time, it is possible to arrange the pathological levels chronologically, with level 0 representing the earliest stage and level 3 the most advanced one.

This represents a novelty compared to the three main characterizations (AD, MCI, and CN), as it introduces a new stage that can aid in identifying the disease before it reaches a more severe and irreversible state. Since Alzheimer's causes brain shrinkage, Figure 4 shows the whole brain volume at each of the identified levels. In this image, it is clearly visible that the lower the brain volume, the more advanced the disease. Considering that the average cranial volume in men is greater than $1500\,\text{cm}^3$, a clear diagnosis of the disease can be established when the whole brain volume is $1.05 \cdot 10^3\,\text{cm}^3$ or less. Other types of dementia, such as Huntington's or Parkinson's disease, can also cause brain shrinkage. Therefore, since levels 1 and 2 in our classification may indicate early stages of Alzheimer's, it is necessary to examine additional biomarkers to ensure an accurate diagnosis.

Another objective of the project was to obtain information about patients in each group and to try to identify possible correlations. In image 5, it is possible to see the gender distribution of patients and their corresponding Alzheimer's level, showing a marked difference between males and females, with the female group being the most affected. The research results confirmed that women have a higher overall tendency to develop AD pathology

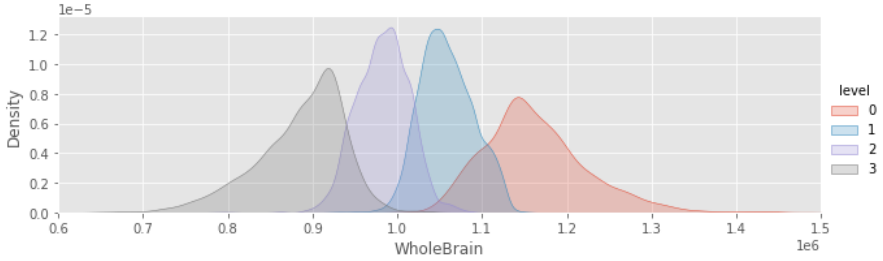

Figure 4: Whole brain volume and Alzheimer level distinction

Table 5
Categorical statistics of patients according to AD level

Feature	Subfeature	level 0	level1	level 2	level 3
Married status	Divorced	1.65 %	2.48 %	3.60 %	1.65
	Married	15.21 %	23.85 %	23.44 %	12.73 %
	Never married	0.41 %	1.03 %	1.90 %	0.66%
	Widowed	0.70 %	2.56 %	3.64 %	4.05 %
Ethnia	Hispanic/Latino	1 %	1.12 %	1.69 %	0.87%
	Non Hispanic/Latino	17.11 %	28.73 %	30.84 %	18.15 %
	Unknown	0.08 %	0.12 %	0.17%	
Race	White	16.99 %	27.28 %	27.74 %	16.49 %
	Black	0.58 %	1.488 %	3.31 %	1.86 %
	Asian	0.33 %	0.82 %	0.744 %	0.45%
	Other	0.29 %	0.29 %	0.86 %	0.37 %

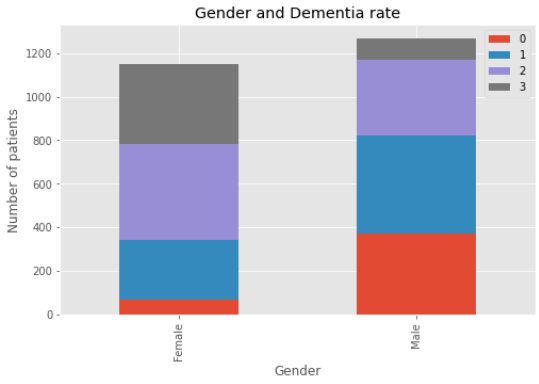

Figure 5: Alzheimer's level and gender of the patients

compared to men, as also reflected in our findings, reinforcing the idea that biological gender is a significant risk factor for Alzheimer's disease.

The distribution of categorical features across different AD levels is summarized in Table 5, expressed as percentages. The analysis revealed that neither race nor ethnicity appeared to be significant factors in the development of Alzheimer's Disease within our dataset. However, the findings of Liu et al. [23] align with our results, highlighting that married individuals are less prone to developing AD. This is likely due to the protective effect of close affective relationships, in contrast to individuals who are divorced, widowed, or living alone. In other words, stable couple relationships may act as a buffer against both the onset and progression of the disease.

Recent evidence from Rodriguez et al. [24] further supports our conclusions, showing that women are more likely than men to develop AD—an observation consistent with our own findings. However, their study also reported a higher prevalence of AD among Hispanic and African American populations, which they attribute to a greater predisposition to diabetes. This risk is notably heightened during pregnancy due to hormonal imbalances. Nonetheless, we were not able to confirm this association in our analysis.

Although our dataset includes both male and female participants, as well as Hispanic and non-Hispanic individuals, we did not incorporate diabetes as a variable. Specifically, we did not examine whether any of the biomarkers in the ADNI database are indicative of diabetic conditions. As such, diabetes-related risk factors were neither considered nor analyzed.

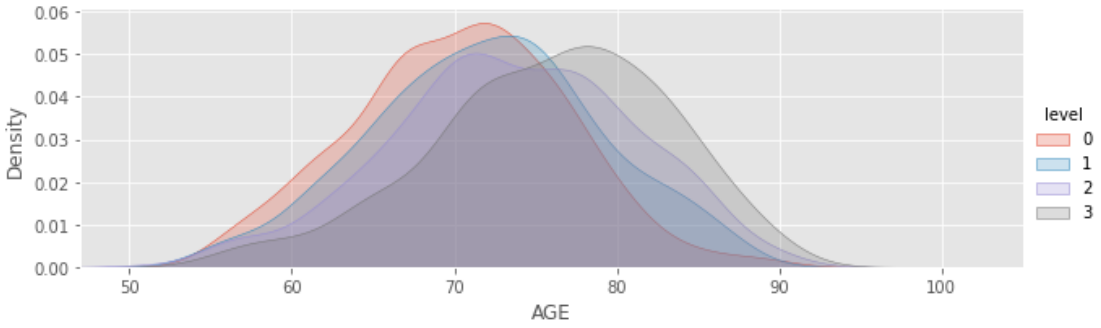

Figure 6: Alzheimer's level and age of the patients

Regarding age, the participants range from 55 to 90 years old. As shown in Figure 6, the mean age of individuals diagnosed with AD is 76.08 $\pm$ 6.91 years. We mention this in connection to the diabetes-related discussion in [24]. While it is known that women face an increased risk of developing diabetes during pregnancy, most women in our dataset are likely post-menopausal and beyond childbearing age, making pregnancy-related diabetes an unlikely contributing factor. Furthermore, since the ADNI dataset does not include variables related to pregnancy history, we were unable to investigate any potential link between pregnancy-related diabetes and AD development.

6. Conclusion

Affecting a large part of the population and expected to reach 139 million people by 2050, Alzheimer's disease is one of the most common forms of neuropsychological dementia. It is a disease that is widely known and, at the same time, extremely misunderstood. Although it has been studied since the beginning of the 20th century, it still has neither a cure nor a full understanding of its progression, despite numerous studies and the substantial funding dedicated to the cause. Longitudinal databases such as ADNI facilitate research in this area by collecting data from patients, both healthy and affected by the disease, over the years; additionally, its public availability allows scientists to work freely on their projects.

Here, we have seen both in the literature review and in practice that machine learning algorithms, such as K-Means, applied to the ADNI database can help us understand relationships among features of different natures, from biological to physical, cognitive, and linguistic, when it comes to the detection of Alzheimer's. It has been shown that women are more likely to suffer from the disease, representing 78.66% of diagnosed individuals. Additionally, the average age of those affected is 76.08 $\pm$ 6.91 years. These results confirm findings from other studies on the subject, including the observation that marital status is not a significant factor in the development of the disease.

One of the key aspects of this project has been the identification of four stages of Alzheimer's disease, ranging from a healthy brain to one that is severely affected. These four stages allow the identification of two intermediate phases between health and severe illness, which makes it possible to classify patients at an early stage of the disease. This can facilitate treatment before it worsens, thereby slowing its progression — one of the main objectives of this project. These four stages are associated with total brain volume, as it can be observed that the higher the disease stage, the lower the brain volume.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, C. P. Ferri, The global prevalence of dementia: a systematic review and metaanalysis, Alzheimer's & dementia 9 (2013) 63–75.

[2] E. Joe, J. M. Ringman, Cognitive symptoms of alzheimer's disease: clinical management and prevention, Bmj 367 (2019).

[3] G. Capizzi, F. Bonanno, C. Napoli, A wavelet based prediction of wind and solar energy for long-term simulation of integrated generation systems,

in: SPEEDAM 2010 - International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2010, p. 586 – 592. doi:10.1109/SPEEDAM.2010.5542259.

[4] F. Bonanno, G. Capizzi, C. Napoli, Some remarks on the application of rnn and prnn for the charge-discharge simulation of advanced lithium-ions battery energy storage, in: SPEEDAM 2012 - 21st International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2012, p. 941 – 945. doi:10.1109/SPEEDAM.2012.6264500.

[5] G. Lo Sciuto, G. Capizzi, R. Shikler, C. Napoli, Organic solar cells defects classification by using a new feature extraction algorithm and an ebnn with an innovative pruning algorithm, International Journal of Intelligent Systems 36 (2021) 2443 – 2464. doi:10.1002/int.22386.

[6] G. Capizzi, F. Bonanno, C. Napoli, Hybrid neural networks architectures for soc and voltage prediction of new generation batteries storage, in: 3rd International Conference on Clean Electrical Power: Renewable Energy Resources Impact, ICCEP 2011, 2011, p. 341 – 344. doi:10.1109/ICCEP.2011.6036301.

[7] N. Brandizzi, S. Russo, R. Brociek, A. Wajda, First studies to apply the theory of mind theory to green and smart mobility by using gaussian area clustering, in: CEUR Workshop Proceedings, volume 3118, 2021, p. 71 – 76.

[8] F. Fiani, V. Ponzi, S. Russo, Keeping eyes on the road: Understanding driver attention and its role in safe driving, in: CEUR Workshop Proceedings, volume 3695, 2023, p. 85 – 95.

[9] R. J. Jutten, L. I. Thompson, S. A. M. Sikkes, P. Maruff, J. L. Molinuevo, H. Zetterberg, J. Alber, D. Faust, S. Gauthier, M. Gold, J. Harrison, A. Lee, P. J. Snyder, A neuropsychological perspective on defining cognitive impairment in the clinical study of alzheimer's disease: Towards a more continuous approach., Journal of Alzheimer's disease : JAD (2022).

[10] K. Strimbu, J. A. Tavel, What are biomarkers?, Current Opinion in HIV and AIDS 5 (2010) 463.

[11] A. Lloret, D. Esteve, M.-A. Lloret, A. Cervera-Ferri, B. Lopez, M. Nepomuceno, P. Monllor, When does alzheimer's disease really start? the role of biomarkers, International journal of molecular sciences 20 (2019) 5536.

[12] A. S. Fleisher, W. S. Houston, L. T. Eyler, S. Frye, C. Jenkins, L. J. Thal, M. W. Bondi, Identification of alzheimer disease risk by functional magnetic resonance imaging, Archives of Neurology 62 (2005) 1881–1888.

[13] K. A. Johnson, N. C. Fox, R. A. Sperling, W. E. Klunk, Brain imaging in alzheimer disease, Cold Spring Harbor perspectives in medicine 2 (2012) a006213.

[14] D. AlSaeed, S. F. Omar, Brain mri analysis for alzheimer's disease diagnosis using cnn-based feature extraction and machine learning, Sensors 22 (2022) 2911.

[15] M. L. Giger, K. Suzuki, Computer-aided diagnosis, in: Biomedical information technology, Elsevier, 2008, pp. 359–XXII.

[16] L. Billeci, A. Badolato, L. Bachi, A. Tonacci, Machine learning for the classification of alzheimer's disease and its prodromal stage using brain diffusion tensor imaging data: A systematic review, Processes 8 (2020) 1071.

[17] S. Gupta, V. Saravanan, A. Choudhury, A. Alqahtani, M. R. Abonazel, K. S. Babu, Supervised computer-aided diagnosis (cad) methods for classifying alzheimer's disease-based neurodegenerative disorders, Computational and Mathematical Methods in Medicine 2022 (2022).

[18] T. Jo, K. Nho, A. J. Saykin, Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data, Frontiers in aging neuroscience 11 (2019) 220.

[19] S. Grueso, R. Viejo-Sobera, Machine learning methods for predicting progression from mild cognitive impairment to alzheimer's disease dementia: a systematic review, Alzheimer's Research & Therapy 13 (2021) 1–29.

[20] M. Syakur, B. Khotimah, E. Rochman, B. D. Satoto, Integration k-means clustering method and elbow method for identification of the best customer profile cluster, in: IOP conference series: materials science and engineering, volume 336, IOP Publishing, 2018, p. 012017.

[21] J. Pareek, J. Jacob, Data compression and visualization using pca and t-sne, in: Advances in Information Communication Technology and Computing, Springer, 2021, pp. 327–337.

[22] L. Van der Maaten, G. Hinton, Visualizing data using t-sne., Journal of machine learning research 9 (2008).

[23] H. Liu, Z. Zhang, S.-w. Choi, K. M. Langa, Marital status and dementia: Evidence from the health and retirement study, The Journals of Gerontology: Series B 75 (2020) 1783–1795.

[24] K. Rodriguez, M. Trevino, G. Castro, H. R. Rodriguez, G. Ortiz, J. Garza, S. Aleman, A. Alvarez, et al., The effects of alzheimer's: Gender, race, & ethnicity, DHR Proceedings 2 (2022) 15–19.