# Enhancing Alzheimer's Disease Diagnosis: Insights from MLP and 1D CNN Models

Jonathan da Silva Bandeira
PPGEC-ECOMP
Universidade de Pernambuco
Recife, Pernambuco, Brazil
jsb2@ecomp.poli.br

Roberta Andrade de Araújo Fagundes
PPGEC-ECOMP
Universidade de Pernambuco
Recife, Pernambuco, Brazil
roberta.fagundes@upe.br

## Abstract

**Context:** Alzheimer's Disease (AD) is a complex neurodegenerative disorder that requires early diagnosis to improve patient outcomes. Recent advances in computational intelligence have sparked interest in leveraging machine learning to enhance diagnostic accuracy and efficiency. These innovations are crucial for transforming decision-making within Information Systems in clinical settings. **Problem:** Traditional methods like PET-scans and cerebrospinal fluid collection are highly accurate but costly and invasive, limiting accessibility. Developing data-driven, non-invasive solutions that retain diagnostic accuracy while handling complex biomedical data, such as plasma protein concentrations, remains a challenge. **Solution:** This study utilizes neural networks, specifically Multi-Layer Perceptron (MLP) and One-Dimensional Convolutional Neural Network (1D CNN). Preprocessing included Recursive Feature Elimination (RFE) for feature selection and Synthetic Minority Oversampling Technique (SMOTE) for data augmentation, addressing class imbalance. **SI Theory:** Grounded in Complexity Theory, the study examines how machine learning models can enhance data-driven medical systems by efficiently managing critical, highly sensitive datasets. **Method:** An experimental quantitative approach was used to evaluate binary and multiclass classifiers on a dataset with 120 protein features from 259 patients. **Summary of Results:** The MLP exhibited strong performance in specific subsets, achieving superior metrics in the binary classification after feature selection and data augmentation. Meanwhile, the 1D CNN excelled in multiclass classification, leveraging its convolutional layers to extract critical features from subtle protein variations, improving accuracy and robustness. **Contributions and Impact on IS Field:** This research enhances medical information systems by proposing machine learning models that can be integrated for accurate diagnostics, supporting clinical decision-making and advancing healthcare practices.

## CCS Concepts

• **Computing methodologies** → **Neural networks**; *Machine learning approaches*; • **Information systems** → **Health informatics**; • **Social and professional topics** → *Medical information systems*; • **General and reference** → Evaluation metrics.

## Keywords

Alzheimer's Disease, Machine Learning, MLP Neural Networks, Feature selection, Data augmentation, 1D CNN, Information Systems applications

## 1 Introduction

In recent decades, life expectancy has significantly increased, especially in developing countries like Brazil. With increased longevity, new challenges and concerns have emerged regarding the needs of the elderly population [1]. One of these concerns is related to the health of the elderly, particularly the development of Chronic Non-Communicable Diseases (NCDs) such as dementias, with Alzheimer's Disease (AD) being the most prominent, affecting approximately over 47 million people worldwide [8].

Dementia can be understood as the decline in an individual's cognitive functions, directly impacting their behavior and quality of life. Dementias can be reversible or irreversible, with the latter being degenerative and progressive (worsening over time). These range from mild conditions like Mild Cognitive Impairment (MCI), which are serious enough to be noticeable, to Alzheimer's Disease, discovered in 1907 by Alois Alzheimer [24].

Alzheimer's Disease is a progressive neurodegenerative disorder that causes irreversible brain damage, deteriorating an individual's cognitive abilities and memory (ability to recall old information or learn new information) and impairing their daily routine and behavior. AD is currently the most common form of degenerative dementia globally, mainly affecting elderly individuals over 65 years old in its late manifestation, though it can also rarely affect younger individuals in its early form [24].

According to the World Health Organization [27], in the last two decades, between 2000 and 2019, Alzheimer's Disease and other forms of dementia were ranked among the top ten leading causes of global mortality. In this ranking, AD and other dementias were the third leading cause of mortality in the Americas and Europe. Globally, women are the most affected, accounting for 65% of total deaths due to these diseases.

Like other chronic diseases, although irreversible, AD can be controlled if diagnosed at an early stage, slowing down its degenerative progression and ensuring a better quality of life for patients. Additionally, mild cognitive impairments like MCI increase the chances of developing more severe dementia, such as AD, by up to 15 times. Therefore, it is crucial that dementia, regardless of its severity, be diagnosed as early as possible [27].

Historically, among the most precise options to anticipate the final diagnosis of AD were Positron Emission Tomography (PET) scans and cerebrospinal fluid collection. However, these methods are either extremely expensive or invasive, limiting their accessibility. A breakthrough came with the discovery of a blood test capable of indicating beta-amyloid protein accumulation in the brain, a key marker of the disease, with 94% accuracy and the ability to predict AD up to twenty years before severe symptoms appear. Despite its

promise, this test is not yet widely available, as it requires extensive validation [19].

Existing diagnostic techniques for Alzheimer's Disease (AD) fall into three main categories: cognitive tests, neuroimaging, and biochemical analyses. Cognitive tests, such as the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA), are widely used due to their accessibility and low cost but are subjective and influenced by external factors like education and language proficiency. Neuroimaging methods, including PET scans and MRI, offer high accuracy in detecting AD-related brain abnormalities, yet their high cost and limited availability hinder routine use. Biochemical analyses measure biomarkers like beta-amyloid and tau proteins in cerebrospinal fluid (CSF) or blood plasma, providing valuable insights into AD progression, though CSF collection is invasive and blood-based tests still require further validation for widespread clinical application.

In this context, Information Systems and Computational Intelligence (CI) play an increasingly vital role. The integration of Machine Learning (ML) techniques within information systems is transforming healthcare diagnostics by leveraging data-driven approaches to support early detection of diseases. These technologies can process and analyze large volumes of clinical data efficiently, offering reliable support for medical diagnosis and enabling automation in clinical settings. By leveraging CI, information systems can facilitate not only diagnostic accuracy but also scalability and accessibility in healthcare.

For instance, advanced ML models are being developed to interpret biochemical and clinical data, identifying patterns that may not be evident through traditional methods. In 2018, an intelligent classification model demonstrated its potential by predicting AD diagnoses with an impressive accuracy of up to six years in advance compared to traditional approaches [11]. Such advancements highlight the significant role of Information Systems in addressing the challenges of early diagnosis and personalized treatment plans.

Given these limitations, research efforts are increasingly focused on identifying alternative biomarkers that could provide an accurate, non-invasive, and cost-effective means of diagnosing AD. This study contributes to this endeavor by analyzing a publicly available dataset developed by Ray et al. (2007), which consists of blood plasma samples from 259 individuals and contains the measured concentrations of 120 proteins. By applying machine learning techniques, the study seeks to identify new proteins correlated with AD progression, beyond the well-established biomarkers such as beta-amyloid and tau. The dataset's structure enables the exploration of potential diagnostic markers, making it a valuable resource for computational intelligence applications in healthcare.

## 1.1 Proposal and Scope

Currently, the set of technology-based diagnostic support solutions is primarily grounded in three approaches: cognitive tests, neuroimaging, and biochemical analyses. The first approach relies on exams such as the MMSE and has the main advantage of providing simpler data to understand, especially for researchers outside the health field. Lins [17] and Machado et al. [18] highlight researches that uses this as one of its approaches.

Neuroimaging is considered the current state-of-the-art among the three approaches, based on evidence and imaging exams, such as PET-Scan and Magnetic Resonance Imaging. The accuracy of the data used in training Machine Learning models is high because post-mortem brain autopsy images provide higher sensitivity and specificity values. The main drawback is that these values are only achieved post-mortem [8]. Research using this approach is the most common in the scientific community, with Silva et al. [23] being one of these studies.

Finally, biochemical analyses have gained prominence due to new insights into the disease's pathophysiology [8] and significant discoveries in the field, such as a new highly accurate and low-cost blood test capable of detecting AD development up to twenty years before onset [19], and advances in studies of a vaccine capable of preventing the disease [10]. This approach focuses on genetic, molecular, and protein characteristics present in the blood plasma, assessed through laboratory tests. Its main advantage is the potential cost-benefit ratio. However, the data are complex to understand, and unlike neuroimaging, which is a more established approach, biochemical analyses have seen significant development only in the past two years due to the mentioned studies.

From the perspective of the latter approach, various hypotheses, according to Uddin et al. [26], describe that the presence of certain proteins and genetic factors contributes to the development of AD and other dementias, with the amyloid hypothesis being one of the most well-known [25]. Based on this assertion, some studies have been developed, producing relevant work in computational diagnosis of AD and MCI using a biochemical approach, such as Ray et al. [22], Ravetti and Moscato [21], and Dantas and Valença [9]. These studies utilized a consistent dataset of 120 proteins found in the blood plasma of 259 patients who underwent laboratory tests throughout their lives and were diagnosed as healthy or with AD, MCI, or another type of dementia.

The collection and interpretation of a healthcare problem dataset is one of the main challenges in Computational Intelligence research applied in this field [2]. The main specific problems encountered in datasets from this field are inconsistencies, anomalies, instance imbalance, missing data, and possible dimensionality issues. Additionally, it is common for the volume of instances to be limited and small.

Based on these difficulties, Bai et al. [3] state that before using Machine Learning algorithms to infer patterns from a dataset with such characteristics, researchers need to address some challenges, such as dimensionality reduction, noise elimination, missing data handling, anomaly treatment, input variable selection, and correct data interpretation.

Studies like Ravetti and Moscato [21] and Dantas and Valença [9] demonstrate that selecting more representative features for the problem positively contributes to the performance of diagnosis by Machine Learning techniques. However, for techniques such as Artificial Neural Networks (ANNs), where diversity and quantity of instances impact both performance and model generalization, there is a hypothesis that, in addition to selection, feeding the model with more diverse and balanced information could improve performance.

Based on this hypothesis, the following research question is formulated: How do the selection and generation of new instances

influence the classification performance of an Artificial Neural Network for supporting AD and MCI diagnosis?

To answer this question, this work aims to investigate the impacts of selecting more representative features for the problem and artificially expanding data on the performance of AD and MCI classification models. For this purpose, the dataset developed by Ray et al. [22] was used.

## 1.2 Related Works

Alzheimer's Disease (AD) diagnosis has been the subject of extensive research, with various approaches focusing on biochemical, neuroimaging, and machine learning techniques. Several studies have explored different methodologies to improve early diagnosis and classification accuracy.

Ray et al. (2007) [22] pioneered the use of plasma protein biomarkers to predict AD, identifying 18 signaling proteins that achieved classification accuracy of up to 90%. Their study employed the Prediction Analysis of Microarrays (PAM) algorithm to select the most relevant biomarkers from a set of 120 proteins, emphasizing the role of hematopoiesis, immune response, and neuronal support in AD pathology.

Building upon this work, Ravetti and Moscato (2008) [21] proposed a refined 5-protein biomarker signature that maintained high accuracy (96%) while reducing computational complexity. Their approach utilized several machine learning classifiers available in the WEKA software, including Random Forest, Naïve Bayes, and Support Vector Machines (SVM), to optimize prediction models and evaluate classifier independence in biomarker selection.

In a related effort, Dantas and Valença (2013) [9] applied Reservoir Computing (RC) to AD diagnosis, comparing its performance with Multi-Layer Perceptron (MLP) and classifiers implemented in Weka. Their study demonstrated that alternative neural architectures could achieve competitive results in biomarker-based classification. Similarly, Silva et al. (2019) [23] investigated the use of Haralick texture descriptors extracted from MRI scans, applying Random Forest, SVM, and K-NN classifiers. Their findings indicated that combining textural and shape features could enhance AD classification accuracy.

Deep learning models have also gained traction in AD diagnosis. Ding et al. (2018) [11] developed a convolutional neural network trained on 18F-FDG PET scans, achieving an area under the ROC curve of 0.98. Their work highlighted the potential of deep learning for early AD detection, outperforming traditional radiological assessments.

Machado et al. (2021) [18] introduced DCARE, a computational model designed to support individuals with AD by analyzing context histories and predicting potential hazardous behaviors. The model processes physiological data received from external applications and utilizes an ontology specifically developed for AD-related contexts. A simulator, DCARE Dataset Simulator, was created to generate datasets for testing, producing 1,026 scenarios based on interviews with five AD care specialists. The model achieved an average accuracy rate of 97.44% in predicting these scenarios, highlighting its effectiveness in enhancing patient safety and independence. This study contributes to the cognitive test-based diagnosis of AD by incorporating historical context analysis for behavioral predictions.

Despite these advancements, many studies primarily emphasize accuracy without thoroughly examining the impact of preprocessing techniques, such as feature selection and class balancing, on model performance. Moreover, precision and recall metrics are often overlooked, limiting the interpretability of results. This work aims to address these gaps by evaluating the effectiveness of neural networks in identifying novel protein-based biomarkers, with a focus on feature selection and class balancing techniques.

# 2 Theoretical Foundation and Techniques

The integration of computational intelligence techniques, particularly in the domain of information systems, plays a crucial role in addressing complex problems in medical diagnostics. Given the intricate nature of biomedical data, especially in neurodegenerative disorders such as Alzheimer's Disease (AD), leveraging advanced machine learning models becomes essential for improving diagnostic accuracy. This section explores the theoretical foundations underpinning the methods used in this study, focusing on Complexity Theory, neural network architectures like Multi-Layer Perceptron (MLP) and One-Dimensional Convolutional Neural Networks (1D CNN), as well as data preprocessing techniques such as Recursive Feature Elimination (RFE) and Synthetic Minority Over-sampling Technique (SMOTE) to enhance classification performance.

## 2.1 Exploration of Complexity Theory

Complexity Theory has been widely applied in research involving information systems, particularly in contexts that require the integration of multiple interdependent components and the management of large volumes of heterogeneous data [13]. In the case of Alzheimer's Disease (AD) diagnosis, this theory provides a crucial framework for dealing with the inherent complexity of biomedical data, such as plasma protein concentrations, which exhibit non-trivial patterns that are difficult to model using traditional approaches.

In the study conducted, applying Complexity Theory enabled a robust approach to implementing neural networks, such as the Multi-Layer Perceptron (MLP) and the One-Dimensional Convolutional Neural Network (1D CNN). These architectures are particularly useful in capturing the complexity of biomedical data, where slight variations in protein levels can indicate the presence or progression of AD. As highlighted by Li and Chen [16], convolutional networks are capable of identifying local and sequential patterns, which are crucial in diagnostics based on temporal data or biochemical measurement series.

Complexity also manifests in the need to optimize feature selection, a challenge in high-dimensional and imbalanced datasets. Techniques such as Recursive Feature Elimination (RFE) and the Synthetic Minority Over-sampling Technique (SMOTE) were essential for handling data complexity and improving model accuracy. Recent studies, such as those by Rajasree and Rajakumari [20], have shown that recursive feature elimination can reduce dimensionality without compromising accuracy for AD diagnosis prediction.

The approach adopted in this study illustrates how Complexity Theory can be applied to understand not only the interaction between different data sources but also how to optimize machine learning models to achieve better results in high-uncertainty scenarios [28]. As pointed out by Bouamrane et al. [4], in healthcare systems, the ability to manage data complexity is essential for enhancing diagnostic accuracy and supporting clinical decision-making.

Thus, by using Complexity Theory as a theoretical foundation, this research not only contributes to the enhancement of medical information systems but also demonstrates the importance of flexible architectures and advanced data processing techniques in addressing the challenges of AI-based diagnostics.

## 2.2 Multi-Layer Perceptron Neural Networks (MLP)

Multi-Layer Perceptron (MLP) neural networks are a class of feedforward artificial neural networks. They consist of multiple layers of nodes, each layer fully connected to the next one. The primary advantage of MLPs is their ability to model complex non-linear relationships, which makes them suitable for various classification and regression tasks in medical diagnosis [23]. In the context of Alzheimer's Disease (AD) diagnosis, MLPs have been employed to differentiate between healthy individuals and those affected by AD or Mild Cognitive Impairment (MCI), leveraging biochemical data such as protein levels in blood plasma.

The architecture of an MLP typically includes an input layer, one or more hidden layers, and an output layer. The number of hidden layers and the number of neurons in each layer can significantly impact the performance of the model. For instance, a deeper network with more layers might capture more intricate patterns in the data but also risks overfitting, especially when the dataset is small [14]. Hyperparameter tuning and regularization techniques, such as dropout and L2 regularization, are crucial to optimizing MLP performance. The fundamental equations governing the operations of an MLP are presented below.

$$z^{(l)} = W^{(l)} x^{(l-1)} + b^{(l)} \tag{1}$$

$$a^{(l)} = f(z^{(l)}) \tag{2}$$

$$\hat{y} = f(W^{(L)} a^{(L-1)} + b^{(L)}) \tag{3}$$

In these equations:

- $z^{(l)}$ is the linear combination of inputs at layer $l$,
- $W^{(l)}$ and $b^{(l)}$ are the weights and biases for layer $l$,
- $a^{(l)}$ is the activation of the neurons at layer $l$,
- $f$ is the activation function (commonly ReLU or sigmoid), and
- $\hat{y}$ is the predicted output.

A significant challenge in training MLPs is the requirement for large datasets to achieve good generalization. Techniques such as data augmentation and synthetic data generation (e.g., using SMOTE) have been explored to mitigate this issue. This work also involves integrating MLPs with more sophisticated feature selection methods like Recursive Feature Elimination (RFE) to enhance predictive accuracy.

## 2.3 One-Dimensional Convolutional Neural Networks (1D CNN)

One-Dimensional Convolutional Neural Networks (1D CNN) are a type of deep learning model particularly effective for analyzing sequential data. Unlike traditional MLPs, CNNs use convolutional layers that apply filters to the input data, capturing local patterns and features [15]. In the context of AD diagnosis, 1D CNNs can process time-series data or sequential measurements from biochemical assays, identifying subtle temporal patterns indicative of disease progression.

The primary strength of 1D CNNs lies in their ability to automatically learn spatial hierarchies of features from input data. This makes them highly suitable for tasks where local dependencies are critical. For example, in this study, 1D CNNs were used to analyze sequences of protein expression levels, achieving higher accuracy in distinguishing AD from other forms of dementia compared to traditional machine learning methods. The convolutional layers extract high-level features that are then processed by fully connected layers to make the final prediction. As highlighted by Goodfellow et al. [12], a 1D CNN consists of convolutional, pooling, and fully connected layers. The fundamental operations of a 1D CNN are defined as follows.

$$z_i = (x * w)_i + b \tag{4}$$

$$a_i = f(z_i) \tag{5}$$

$$p_i = \max(a_{i:i+k-1}) \tag{6}$$

In these equations:

- $z_i$ is the result of the convolution operation at position $i$,
- $x$ is the input vector, $w$ is the filter (kernel), and $b$ is the bias,
- $a_i$ is the activation after applying the function $f$ (commonly ReLU),
- $p_i$ represents the result of max-pooling over a window of size $k$.

Implementing 1D CNNs involves selecting appropriate kernel sizes, number of filters, and the depth of the network. Too many filters or layers can lead to overfitting, particularly with small datasets. Regularization techniques, such as dropout and batch normalization, are essential to mitigate this risk. Moreover, combining 1D CNNs with feature selection methods like RFE can further enhance their performance by focusing on the most relevant features.

## 2.4 Recursive Feature Elimination (RFE)

Recursive Feature Elimination (RFE) is a feature selection technique that helps in identifying the most significant features for a given predictive model. It works by recursively removing the least important features and building the model iteratively until the optimal number of features is reached [6]. In the realm of AD diagnosis, RFE has been instrumental in refining datasets by reducing dimensionality and enhancing model interpretability.

The RFE process involves training a model, ranking features based on their importance, and eliminating the least significant ones. This is particularly useful in biomedical datasets where the number of features (e.g., various protein levels) can be very high. This work utilized RFE to select a subset of proteins that were most

predictive of AD, thereby improving the efficiency and accuracy of subsequent classification models.

One of the key advantages of RFE is its compatibility with different types of models, including linear models, support vector machines, and neural networks. By integrating RFE with MLPs or 1D CNNs, researchers can enhance the predictive power of these models while reducing the computational cost.

## 2.5 Synthetic Minority Over-sampling Technique (SMOTE)

The Synthetic Minority Over-sampling Technique (SMOTE) is a popular method for addressing class imbalance in datasets. It generates synthetic examples of the minority class by interpolating between existing minority instances [5]. In AD diagnosis, where datasets often suffer from imbalanced class distributions (e.g., fewer AD cases compared to healthy controls), SMOTE can significantly improve model performance by providing a more balanced training set.

SMOTE works by selecting two or more similar instances from the minority class and generating new instances that lie between them. This synthetic generation helps the model to learn the characteristics of the minority class better, thus improving classification accuracy and reducing bias [7]. This study demonstrated that applying SMOTE to a dataset of protein levels from AD patients and healthy controls improved the performance of MLPs in distinguishing between these classes.

However, care must be taken when using SMOTE, as excessive synthetic generation can lead to overfitting, particularly if the original dataset is very small. Combining SMOTE with other techniques like RFE and robust cross-validation methods can help mitigate this risk and ensure the model generalizes well to unseen data.

# 3 Materials and methods

The methodology employed in this study was designed to ensure data preparation, model training, and evaluation, as illustrated in Figure 1. The key steps are described below:

(1) **Data collection and preprocessing:** The dataset used in this study consisted of 120 protein concentrations derived from blood plasma samples collected from patients diagnosed with Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Non-Dementia (ND). Initially, the data were organized with features presented as rows rather than columns due to its microarray origin. To prepare the data for machine learning models, the matrix was first transposed to align features as columns. Subsequently, data standardization was performed using the StandardScaler, according to the formula:

$$z = \frac{x - \mu}{\sigma} \tag{7}$$

where $x$ is the original feature value, $\mu$ is the mean, and $\sigma$ is the standard deviation of the feature.

(2) **Feature selection with RFE:** To optimize model performance and reduce dimensionality, RFE was employed using a Random Forest estimator. This iterative technique involves recursively removing the least significant features and training the model until an optimal subset is achieved. The use of RFE enhances model interpretability and reduces the risk of overfitting by focusing on the most relevant features.

(3) **Data Augmentation using SMOTE:** Given the class imbalance in the dataset, particularly the limited instances of AD cases, the SMOTE was applied to generate synthetic data points for the minority class. This approach ensures a more balanced dataset, allowing the classifiers to generalize better and avoid biases toward the majority class.

(4) **Model implementation and parameterization:** Two classifiers were implemented:
- A Multi-Layer Perceptron (MLP) with two hidden layers, using the ReLU activation function and a softmax output layer.
- A One-Dimensional Convolutional Neural Network (1D CNN) with a single convolutional layer, followed by max pooling, dropout, and fully connected layers.

Hyperparameters such as learning rate, number of neurons, and batch size were tuned to optimize model performance.

(5) **Experimentation in two scenarios:**
   (a) **Binary Classification:** Classifiers were trained on two subsets of data to distinguish between AD and ND patients, and MCI and AD patients.
   (b) **Multiclass Classification:** Ternary classifiers were trained on the full dataset to classify between AD, MCI, and ND patients.

(6) **Evaluation of models:** The performance of the models was assessed using metrics such as Accuracy, Precision, Recall, and F1-Score to provide a comprehensive evaluation of classifier effectiveness.

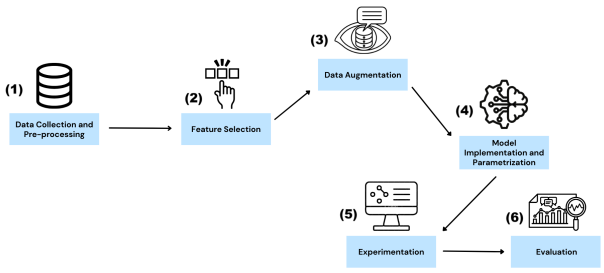


**Figure 1: Overview of the methodological approach used in this study.**

## 3.1 Dataset

The dataset employed in this study, curated and catalogued by Ray et al.[22], comprises blood plasma samples from 259 individuals, encompassing the relative concentration of 120 proteins, identified as potential biomarkers for Alzheimer's Disease (AD) and other neurological conditions. These samples were collected from specialized centers focusing on neurological and neurodegenerative diseases across the United States. These proteins were quantified

using filter-based, arrayed sandwich ELISAs, a high-throughput technique commonly used in proteomics studies to measure protein abundance.

The proteins measured in this dataset are involved in various biological processes, including immune response, hematopoiesis, apoptosis, and neuronal support. Some of the key proteins include CCL5, ICAM-1, TNF-alpha, IL-3, PDGF-BB, and EGF, which have been previously associated with neuroinflammatory and neurodegenerative processes. However, the focus of this study is to investigate whether neural network models can identify new proteins within this subset that may have a significant correlation with AD, beyond those already established in the literature.

All data files were originally provided in tabular format (.xlsx and .csv) and processed in Python for compatibility with machine learning models. The dataset was preprocessed to ensure consistency, removing missing values and normalizing protein expression levels for analysis.

This dataset was selected for three primary reasons. The first reason was to establish a real-world problem within the chosen application domain. The second reason pertains to the dataset's availability. Constructing a new dataset for the experiments would be infeasible in terms of cost and benefit, given the time required for sample collection, processing, and cataloging, as well as the necessity for ethical committee approval to authorize its use. Other datasets focusing on the same biochemical approach, which is the interest of this study, are harder to obtain, often requiring registrations, permissions, or releases from the organizations and researchers who compiled them. The dataset used is openly available, facilitating its immediate use.

The third reason relates to the content of the dataset. It is widely recognized that in biochemical or biomolecular approaches, the concentrations of beta-amyloid and tau proteins contribute to the development of Alzheimer's Disease (AD) [25]. Studies on the concentration of these proteins in patients have yielded useful results for early detection laboratory tests and even for potential vaccines to treat and prevent this condition. However, this dataset does not include information on these specific proteins. Instead, it comprises a subset of 120 proteins collected and analyzed in the study by Ray et al. (2007) [22], which explored potential biomarkers for AD diagnosis.

The objective of this research is to investigate whether neural networks can effectively identify other proteins, within this subset, that may have a significant correlation with AD, even if they have not yet been widely recognized in prior medical research. This approach allows for the discovery of new biomarker candidates that could contribute to further studies on AD diagnosis. Although this dataset does not include information on these specific proteins, it remains relevant because the etiology of AD is not fully understood, and there are no specific markers in laboratory and neuroimaging investigations for AD. Some findings, however, can support clinical diagnosis. Based on this premise, despite not being recently utilized, the investigation of the proteins present in this dataset is considered a viable element in the development of new solution models.

The majority of instances belong to the clinical diagnoses of Alzheimer Disease (AD), Non Dementia cases (ND), and Mild Cognitive Impairment (MCI), and these instances are used in both the experiments of this work and related studies. In their research, Ray et al. [22] subdivided the AD and ND samples into training and testing sets with a similar distribution. Out of the total 259 patient samples, 48 were not related to dementias and were not considered in the experiments. Of the 85 AD samples, 43 were used for training and 42 for testing. The ND samples were divided into 40 for training and 39 for testing. Originally, the dataset for the first experiment included a total of 164 samples (comprising all instances of ND and AD), divided into 83 training samples and 81 testing samples. The second experiment, which focuses on MCI data, considers another relevant piece of information: whether or not the evaluated cases of MCI progressed to other dementias. Therefore, the 47 MCI instances are divided into three classes for final diagnosis: AD, MCI, and Other Dementias. For this study, data related to Other Dementias were excluded, leaving 39 instances in this subset.

In both datasets, there are 120 features, which correspond to the concentrations of proteins in the patients' plasma. The first step was to analyze the layout and types of data. The AD dataset is organized into 121 columns and 164 rows. The rows represent the total number of instances or examples. The columns are the features that describe each example. Of these columns, 120 are input data, consisting of continuous values (in computing, floating-point numbers or floats), while 1 column corresponds to the class, label, or desired output. The MCI dataset has 121 columns and 39 rows. Out of the 121 columns, 1 column represents the classes or labels expressed as categorical objects, and 120 columns are input data. These latter columns represent the concentrations of the same proteins identified in the AD dataset, expressed as continuous numerical values. In the MCI dataset, all patients were initially diagnosed with MCI, with some progressing to AD while others maintained their initial diagnosis. Thus, the classes refer to the final diagnoses, which vary between AD or MCI.

## 3.2 Data Preprocessing and Feature Selection

To prepare the data for machine learning models, the original matrix, as seen in the Figure 2, was transposed to align features (proteins) as columns and samples as rows. This transformation ensured compatibility with common machine learning frameworks. Data normalization was applied using Z-score transformation, which standardizes each protein's expression levels by subtracting the mean and dividing by the standard deviation across all samples.

| CLASS | AD | AD | AD | AD |
|---|---|---|---|---|
| ANG | 4,520104 | 4,430224 | 4,244593 | 2,90625 |
| BDNF | 2,692648 | 0,94586 | 2,916182 | -0,26235 |
| BLC | 0,788355 | 0,243013 | 1,029638 | -0,65143 |
| BMP-4 | 1,142045 | 0,407405 | 1,966551 | -0,03289 |
| BMP-6 | -0,37695 | 0,480273 | 2,39777 | -0,45564 |
| CK b8-1 | 0,312388 | 0,082394 | 0,99577 | -0,54466 |
| CNTF | 0,036673 | -0,23741 | 0,54394 | -0,37042 |

**Figure 2: Sample of Original Data Matrix**

The dataset used has few instances (164 and 39 in the AD and MCI subsets, respectively) and many input attributes, totaling 120. This characterizes a high-dimensional problem. To mitigate this issue, a feature selection was applied to maintain a minimal protein

signature without compromising the classifier's final performance. The feature selection was performed using RFE technique, which requires a Machine Learning algorithm as an estimator. The algorithm used for this purpose was Random Forest, chosen for its effective visualization of selections based on the importance of features.

Ten selections were executed, and the most balanced one in terms of the smallest number of features and accuracy level in tests was chosen. Of the ten executions with the estimator, five used the Gini criterion, which is based on impurity levels, and five used Entropy, which is based on information gain. Generally, models using entropy performed better according to the observed criteria of the number of selected features and accuracy rates. The Figure 3 shows the graph of this relationship for the Random Forest estimator. The selection that returned 8 features and had an accuracy above 90% was chosen.

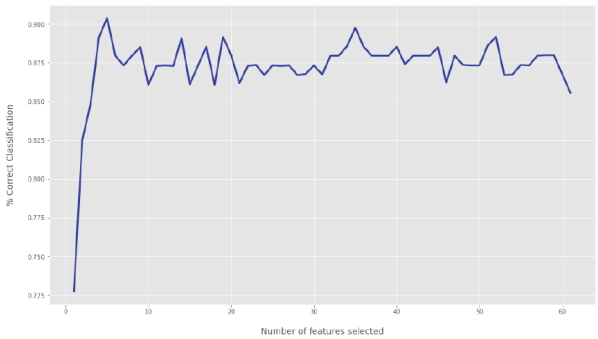


**Figure 3: Graph of selected features x performance**

As highlighted, tree-based algorithms work with the importance level of the features in a dataset. The importance level of the 8 selected features is presented in the Figure 4.

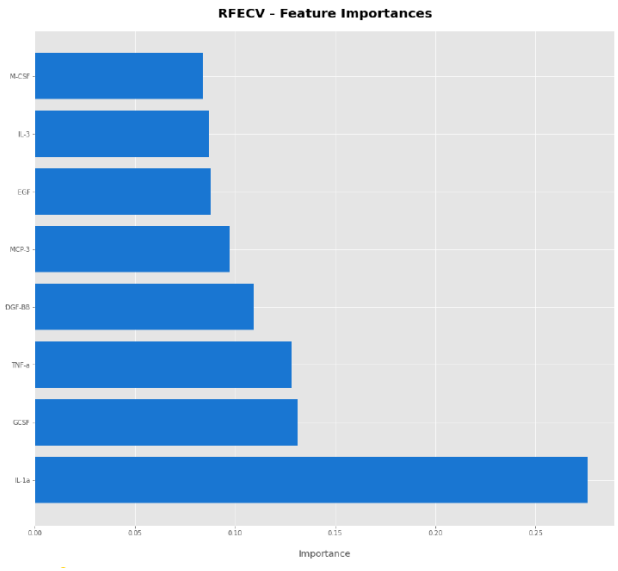


**Figure 4: Ranking of features importance**

Among the selected features for this 8-protein signature, the "IL-1 alpha" protein is considered the most relevant to the problem. Interestingly, both this protein and "PDGF-BB" appeared in all selections, regardless of the algorithm, with "IL-1 alpha" consistently being the most relevant. It is also noted that the importance levels are not very high, as the importance level is proportional to the number of selected features.

The number of selected proteins and the importance level of a chosen variable can vary. This variability occurs because the dataset is vast and contains highly diverse features. Additionally, Random Forest is an ensemble method that uses a majority voting system, meaning the different trees generated in each run can alter the selection of a particular protein or its importance ranking. Nevertheless, the most relevant features will always be selected, as previously noted.

### 3.3 Metrics

All models were evaluated using four metrics valid for classification problems: accuracy, precision, recall (sensitivity), and F1-score. These metrics were chosen because they all vary based on the values in a confusion matrix.

The use of accuracy, precision, recall, and F1-Score is important for comprehensively evaluating the performance of classification models. Accuracy provides an overall view of the proportion of correct predictions, while precision and recall are crucial for understanding the model's behavior in correctly detecting positive classes, minimizing critical errors. The F1-Score balances precision and recall, offering a robust metric when it is important to optimize both detection and reliability in classifications.

## 4 Experiments and Evaluation

The hyperparameter configurations used in the models were empirically tested, considering values that are multiples of 8 (related to the number of inputs) for the number of neurons per layer, and for the others, manual tests were performed with the values available in the Keras framework.

In both experiment scenarios, a simple MLP neural network model with two hidden layers was tested. The first hidden layer had 64 neurons and the second had 32, with one neuron in the output layer and 8 in the input layer, receiving the 8 inputs corresponding to the proteins from the signature selected with RFE and Random Forest. The model used the ReLU activation function in the hidden layers and the sigmoid logistic activation function in the output layer. The optimization algorithm was Adam with learning rate of 0.001, and loss function was binary crossentropy. The AD subset, with its 164 instances distributed as in Ray et al.'s original work, was used for training and testing. The model was trained for 10 epochs with a mini-batch size of 32.

The same model was then evaluated with the application of the SMOTE technique to generate synthetic instances and expand the training samples in 500 samples by class in both subsets of AD and MCI. The original 81 samples of AD subset were reserved for testing. For MCI subset, 12 samples were reserved for testing. Compared to the model's performance without data expansion, the model using this technique showed slightly better performance.

Additionally, a more robust architecture was employed, considering the expanded dataset for training a 1D CNN. Given that it is more robust and requires more data compared to the simple MLP, it made no sense to test this architecture with a small sample set. This architecture was considered due to its more robust feature extraction capability provided by convolution, performed in one dimension (considering feature vectors, not matrices). The 1D CNN model was built with a single convolutional layer with 64 neurons, a kernel size of 2, and ReLU activation function, followed by a max-pooling layer of size 2, and three fully connected hidden layers with 128, 64, and 64 neurons, respectively, all with ReLU activation function. Between these layers, a dropout of 25% was used, and in the output layer, a single neuron with sigmoid logistic activation function was used. The same mini-batch size, number of epochs, optimizer, and loss function as the MLP model were utilized. The goal of using this model was to evaluate whether its performance would be significantly superior to the traditional MLP model.

Experiments were extended to the MCI subset with a total of 39 samples. MLP tests without SMOTE indicated that the performance of machine learning models with this subset was very low due to the small number of samples. Tests were conducted on the original subset compared to the expanded and features selected subset to validate the SMOTE and RFE approach. The primary objective of this work was to use a more robust architecture on this subset, something not previously done in other studies. The same data expansion to 500 samples was applied, and the performance of both proposed models for this subset was also evaluated.

## 4.1 First Scenario: Binary Classifiers

In this first scenario, considering binary classifiers in two subsets of data, the proposed MLP model without data expansion and features selection was tested and compared to the same model with data expansion and features selection. The results, shown in the Table 1, indicate that the model tested with RFE selection on the SMOTE-expanded dataset was slightly superior when tested on the AD original subset.

**Table 1: Results of first scenario experiments - binary classifiers AD subset**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| MLP (Original data) | 89% | 88% | 90% | 89% |
| MLP (RFE-SMOTE) | 94% | 97% | 90% | 94% |
| 1D CNN (RFE-SMOTE) | 90% | 90% | 90% | 90% |

The results of these models were also compared with the results obtained by the proposed 1D CNN model. It was observed that the results of the MLP model were slightly superior to those of the 1D CNN model. For the MCI results, shown in the Table 2, comparisons were made between the MLP model without features selection on original subset and MLP model with RFE on expanded dataset. Comparisons were also made with the 1D CNN on the same expanded dataset, where the MLP model also proved better.

**Table 2: Results of first scenario experiments - binary classifiers MCI subset**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| MLP (Original data) | 58% | 83% | 56% | 67% |
| MLP (RFE-SMOTE) | 70% | 88% | 58% | 70% |
| 1D CNN (RFE-SMOTE) | 65% | 78% | 58% | 67% |

## 4.2 Second Scenario: Multiclass Classifiers

For this scenario, the idea was to combine the two subsets of 164 and 39 instances into a single dataset, with 102 samples reserved for testing. The remaining samples were artificially expanded for training. The two originally proposed models were adapted to perform classification of the three existing classes simultaneously: AD, MCI, and ND.

The MLP neural network model had the number of neurons in the output layer changed to 3 neurons, and its output-layer activation function was modified to the softmax function. The loss function was changed to categorical crossentropy. The remaining hyperparameters remained the same as in the first experiment scenario. The same changes were applied to the 1D CNN model, maintaining the configurations of the previous scenario for the other hyperparameters.

In this second scenario, considering multiclass classifiers in a single expanded dataset, the MLP and 1D CNN models were tested. The results, shown in the Table 3, indicate that the 1D CNN model outperformed the MLP model in all metrics considered.

**Table 3: Results of second scenario experiments - multiclass (ternary) classifiers**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| MLP (RFE-SMOTE) | 78% | 87% | 76% | 81% |
| 1D CNN (RFE-SMOTE) | 97% | 98% | 96% | 97% |

## 4.3 Discussion

In the binary classification scenario on AD subset, MLP had almost the same performance than 1D CNN, but it was slightly better, achieving a 94% accuracy, 90% recall, and 89% F1-Score with the expanded dataset. This demonstrates the strength of the MLP in capturing global relationships within the protein expression data, making it particularly effective in distinguishing between Alzheimer's Disease and healthy controls. While the 1D CNN reached a competitive 90% accuracy, 90% recall, and 90% F1-Score. These results shows that in this binary context, overall data distribution was key to achieving high classification performance.

In the binary classification scenario for MCI subset, the MLP model showed significant improvement when utilizing RFE and SMOTE. The MLP trained on the original dataset achieved a low accuracy of 58% and an F1-score of 67%, highlighting its limited ability to generalize without features selection and data expansion. However, with RFE and SMOTE, the MLP's accuracy increased to 70%, accompanied by an F1-score of 70% and a precision of

88%. In contrast, the 1D CNN with RFE trained on expanded data achieved an accuracy of 65% and an F1-score of 67%, demonstrating slightly lower performance in terms of precision compared to the MLP. These results suggest that while both models benefited from feature selection and data augmentation, the MLP was better suited for handling the complex patterns associated with the MCI subset in this binary classification task.

In the multiclass classification scenario, where AD, Mild Cognitive Impairment (MCI), and Non-Dementia (ND) were analyzed collectively, the 1D CNN excelled across all metrics, achieving an impressive 97% accuracy, 96% recall, and 97% F1-score. This performance highlights the superior ability of convolutional architectures to differentiate between subtle class variations, such as those between MCI and ND. These results underline the value of leveraging both global and local data representations, depending on the complexity and class granularity of the problem, to maximize the effectiveness of diagnostic models in heterogeneous and imbalanced clinical datasets. This study demonstrates the importance of tailoring machine learning approaches to specific diagnostic challenges to enhance precision and support clinical decision-making.

The study concludes by comparing our best-proposed model, the MLP with RFE and SMOTE, with the best results from previous studies on this dataset Ray et al. [22], Ravetti and Moscato [21], Dantas and Valença [9]. It was found that our best binary classifier outperformed in accuracy, the only metric used in all related works, compared to the two previous other machine learning works listed in the Table 4 for MCI.

**Table 4: Comparison between the performance of the best binary classifier model of the present work with the related works**

| Research | Acc. (AD) | Acc.(MCI) |
|---|---|---|
| Ray et al. (2007) | 89% | 81% |
| Ravetti & Moscato (2008) | 93% | 65% |
| Dantas & Valença (2013) | 94% | 66% |
| This research's MLP-RFE-SMOTE (2024) | 94% | 70% |

As a novel contribution, we highlight the proposal of a multiclass model, made possible by artificial data expansion, which was not considered in previous studies, along with the use of a deep learning architecture.

## 4.4 Limitations and Future Work

While the proposed models demonstrated strong classification performance in the diagnosis of Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI), several limitations must be acknowledged to ensure a comprehensive understanding of their applicability and future improvements.

The dataset used in this study comprises 259 patient samples with 120 protein features. Despite the robust feature selection and augmentation techniques applied, the relatively small sample size may limit the generalization of the results to broader populations. Future studies should consider expanding the dataset by incorporating additional patient data from diverse demographics to improve model robustness.

The reliance on protein biomarkers for classification introduces challenges in feature interpretation. Although Recursive Feature Elimination (RFE) successfully identified a subset of the most relevant proteins, further research is needed to validate these findings in clinical settings. Additionally, integrating explainable AI techniques could enhance the interpretability of model decisions, aiding clinicians in understanding model outputs.

The proposed models were specifically trained on biochemical data related to AD and MCI. However, their potential for application in broader contexts, such as diagnosing other neurodegenerative disorders or complex medical conditions, remains an open question. Future research should explore the adaptability of these models to different datasets and medical conditions, assessing their ability to identify patterns beyond AD pathology.

The use of deep learning models, particularly the One-Dimensional Convolutional Neural Network (1D CNN), increases computational requirements. While performance gains were evident, the feasibility of deploying such models in real-time clinical applications or resource-constrained environments warrants further investigation. Optimizations, such as model pruning or lightweight architectures, could improve deployment feasibility.

Before clinical implementation, extensive validation is required to ensure ethical compliance, data privacy, and real-world effectiveness. Collaboration with healthcare professionals and institutions will be crucial for bridging the gap between computational models and practical diagnostic applications.

Approaches such as ensemble learning (e.g., Gradient Boosting, XGBoost, and Random Forest), transformer-based architectures, and self-supervised learning frameworks could provide deeper insights into the dataset while potentially improving classification accuracy and generalization. Additionally, hybrid approaches that integrate deep learning with traditional machine learning algorithms could be explored to leverage interpretability and performance.

By addressing these challenges, the proposed models can evolve into more comprehensive and adaptable tools, contributing to improved diagnostic accuracy and patient outcomes in the management of neurodegenerative diseases.

# 5 Conclusions

The increase in population longevity, especially in developing countries like Brazil, has brought new challenges in managing the health of the elderly, particularly with the rising prevalence of Non-Communicable Chronic Diseases (NCDs) such as dementias. Alzheimer's Disease (AD), a progressive neurodegenerative disorder, is the most common form of dementia, affecting millions globally and significantly impacting both patients and their caregivers' quality of life.

Traditional diagnostic approaches, such as the Mini-Mental State Examination (MMSE), are limited by variability and accuracy, often influenced by external factors like patients' educational background. More precise techniques, including Positron Emission Tomography (PET) scans and cerebrospinal fluid analysis, though effective, remain costly and invasive, limiting accessibility. Although promising advancements, such as blood tests that detect AD years before severe symptoms appear, have been made, these innovations are not yet widely available.

In this context, Computational Intelligence (CI), leveraging Machine Learning (ML) techniques, offers a promising alternative to support early and accurate AD diagnosis. This study explored the use of Multilayer Perceptron (MLP) and One-Dimensional Convolutional Neural Networks (1D CNN) to classify patients using biochemical data, demonstrating high accuracy in distinguishing between healthy individuals, those with Mild Cognitive Impairment (MCI), and those with AD.

Using Complexity Theory as a basis, this research contributes to the field of Information Systems by providing a model for managing the intricate relationships inherent in biomedical datasets. Feature selection using Recursive Feature Elimination (RFE) and data augmentation with Synthetic Minority Over-sampling Technique (SMOTE) significantly improved the models' performance. The MLP, for instance, saw an increase in accuracy from 58% to 70% and in F1-score from 67% to 70% for the MCI subset. Similarly, the 1D CNN benefited from these techniques, achieving a higher balance in metrics for imbalanced datasets.

The results showed that while the MLP demonstrated strong performance in specific subsets, the 1D CNN excelled overall, achieving superior accuracy and robustness in multiclass classification tasks. These findings underscore the potential of convolutional architectures to capture subtle patterns in biochemical data critical for diagnosing AD and MCI, even in complex and heterogeneous datasets.

By integrating Complexity Theory, this study not only improves diagnostic accuracy but also enriches the Information Systems domain by illustrating how advanced ML models can be effectively applied in healthcare. Future research should continue to refine these technologies, aiming to enhance scalability, accessibility, and precision in clinical settings, ultimately supporting better decision-making processes and patient outcomes in the management of neurodegenerative diseases.

## Acknowledgments

The authors would like to acknowledge the assistance of OpenAI's ChatGPT in the structuring, proofreading, and summarizing of English text throughout the development of this paper. The AI-based language model contributed to improving the clarity and coherence of the writing.

## References

[1] Alzheimer's Association. 2023. 2023 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia* 19 (2023), 1598–1695. https://doi.org/10.1002/alz.13016

[2] A. Arora, J. E. Alderman, J. Palmer, S. Ganapathi, E. Laws, M. D. McCradden, L. Oakden-Rayner, S. R. Pfohl, M. Ghassemi, F. McKay, D. Treanor, N. Rostamzadeh, B. Mateen, J. Gath, A. O. Adebajo, S. Kuku, R. Matin, K. Heller, E. Sapey, N. J. Sebire, et al. 2023. The value of standards for health datasets in artificial intelligence-based applications. *Nature Medicine* 29, 11 (2023), 2929–2938. https://doi.org/10.1038/s41591-023-02608-w

[3] B.M. Bai, N. Mangathayaru, and B.P. Rani. 2015. An Approach to Find Missing Values in Medical Datasets. In *International Conference on Engineering & MIS (ICEMIS '15)*. ACM, 1–7. https://doi.org/10.1145/2832987.283308

[4] M-M. Bouamrane, C. Tao, and I. N. Sarkar. 2015. Managing interoperability and complexity in health systems. In *Methods of Information in Medicine*, Vol. 54. 01–04.

[5] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16 (2002), 321–357.

[6] T. Chen and C. J. Lin. 2021. Feature Selection for Support Vector Machines: An Application to Medical Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 4 (2021), 729–741.

[7] D. Dablain, B. Krawczyk, and N. V. Chawla. 2023. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. In *Proceedings of the IEEE Transactions on Neural Networks and Learning Systems*, Vol. 34. 6390–6404. https://doi.org/10.1109/TNNLS.2021.3136503

[8] A.P. Dalmagro, C.A. Cazarin, and F.S. Zenaide. 2020. An Update about the Biochemical and Molecular Basis of Alzheimer's Disease. *Brazilian Applied Science Review* 4, 1 (2020), 118–130.

[9] L. Dantas and M. Valença. 2013. Utilizando Reservoir Computing para Realizar Diagnóstico Precoce de Mal de Alzheimer. In *BRICS Symposium on Biocomputing (SBIC)*.

[10] H. Davtyan, A. Hovakimyan, and S. et al. Kiani Shabestari. 2019. Testing a MultiTEP-Based Combination Vaccine to Reduce Beta and Tau Pathology in Tau22/5xfad Bigenic Mice. *Alzheimer's Research & Therapy* 11 (2019), 107.

[11] Y. Ding, J.H. Sohn, and M.G. et al. Kawczynski. 2018. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. *Radiology* 290, 2 (2018), 456–464.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

[13] Forat Falih Hasan and Muhamad Shahbani Abu Bakar. 2021. An Approach for Data Transformation In Homogeneous and Heterogeneous Information Systems. In *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. 1–5. https://doi.org/10.1109/HORA52670.2021.9461287

[14] S. Haykin. 2022. *Neural Networks and Learning Machines* (4th ed.). Pearson.

[15] S. Kiranyaz, O. Avcı, O. Abdeljaber, T. Ince, M. Gabbouj, and D.J. Inman. 2019. 1D Convolutional Neural Networks and Applications: A Survey. *arXiv preprint* arXiv:1905.03554 (2019). Available at https://arxiv.org/abs/1905.03554.

[16] X. Li and Y. Chen. 2023. Attention-guided CNN for Early Alzheimer's Detection Using Neuroimaging Data. *IEEE Transactions on Medical Imaging* 42 (2023), 145–157. https://doi.org/10.1109/TMI.2023.10898765

[17] Anthony Lins. 2018. *Aplicação de Aprendizagem de Máquina no Diagnóstico de Declínio Cognitivo e Demência de Alzheimer Baseado em Testes Cognitivos e Marcadores Genéticos*. Ph. D. Dissertation. Universidade Federal Rural de Pernambuco (UFRPE), Recife.

[18] Savanna Machado, Jorge Luis Barbosa, João Tavares, and Márcio Martins. 2021. DCARE: Um Modelo Computacional para Acompanhamento de Pessoas com Doença de Alzheimer baseado na Análise de Históricos de Contextos. In *Anais do XVII Simpósio Brasileiro de Sistemas de Informação* (Uberlândia). SBC, Porto Alegre, RS, Brasil. https://sol.sbc.org.br/index.php/sbsi/article/view/17700

[19] S. Palmqvist, S. Janelidze, and Y.T. et al. Quiroz. 2020. Discriminative Accuracy of Plasma Phospho-Tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA* 324, 8 (2020), 772–781.

[20] R. S. Rajasree and S. B. Rajakumari. 2024. Exploring Impact of Feature Selection on Classification Models for Detection of Alzheimer's Disease. In *Proceedings of the International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. Chikkaballapur, India, 1–6.

[21] M.G. Ravetti and P. Moscato. 2008. Identification of a 5-Protein Biomarker Molecular Signature for Predicting Alzheimer's Disease. *PLOS ONE* 3 (2008), e3111.

[22] S. Ray, M. Britschgi, C. Herbert, and Y. et al. Takeda-Uchimura. 2007. Classification and Prediction of Clinical Alzheimer's Diagnosis Based on Plasma Signaling Proteins. *Nature Medicine* 13 (2007), 1359–1362.

[23] I.R.R. Silva, G.S.L. Silva, R.G. de Souza, W.P. dos Santos, and R.A.A. Fagundes. 2019. Model Based on Deep Feature Extraction for Diagnosis of Alzheimer's Disease. In *International Joint Conference on Neural Networks (IJCNN)*. IEEE, Budapest, Hungary, 14–19.

[24] A. A. Tahami Monfared, M. J. Byrnes, L. A. White, and Q. Zhang. 2022. Alzheimer's Disease: Epidemiology and Clinical Progression. *Neurology and Therapy* 11 (2022), 553–569. https://doi.org/10.1007/s40120-022-00338-8

[25] M. Timmers, I. Tesseur, J. Bogert, H. Zetterberg, K. Blennow, A. Börjesson-Hanson, M. Baquero, M. Boada, C. Randolph, and L. Tritsmans. 2019. Relevance of the Interplay Between Amyloid and Tau for Cognitive Impairment in Early Alzheimer's Disease. *Neurobiology of Aging* 79 (2019), 131–141. https://doi.org/10.1016/j.neurobiolaging.2019.03.016

[26] M. S. Uddin, A. Al Mamun, M. Rahman, T. Behl, A. Perveen, A. Hafeez, M. N. Bin-Jumah, M. M. Abdel-Daim, and G. M. Ashraf. 2020. Emerging Proof of Protein Misfolding and Interactions in Multifactorial Alzheimer's Disease. *Current Topics in Medicinal Chemistry* 20 (2020), 2380–2390. https://doi.org/10.2174/1568026620666200601161703

[27] World Health Organization (WHO). 2020. WHO Reveals Leading Causes of Death and Disability Worldwide: 2000-2019. World Health Organization, Switzerland. Available at https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019.

[28] T. Xia, T. Dang, J. Han, L. Qendro, and C. Mascolo. 2024. Uncertainty-Aware Health Diagnostics via Class-Balanced Evidential Deep Learning. In *Proceedings of the IEEE Journal of Biomedical and Health Informatics*, Vol. 28. 6417–6428. https://doi.org/10.1109/JBHI.2024.3360002