İSTANBUL TİCARET ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

İstanbul Commerce University Journal of Science

http://dergipark.org.tr/ticaretfbd

Research Article / Araştırma Makalesi

EMPOWERING SELF-DETECTION: A GRAPHICAL USER INTERFACE POWERED BY MACHINE LEARNING FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE

KENDİNE TANIMANIN GÜÇLENDİRİLMESİ: ALZHEİMER HASTALIĞININ ERKEN TANISINA YÖNELİK MAKİNE ÖĞRENMESİ TEMELLİ BİR GRAFİKSEL KULLANICI ARABİRİMİ

Abdullah Talha KABAKUŞ[1] **Pakize ERDOĞMUŞ**[2]

https://doi.org/10.55071/ticaretfbd.416508

Corresponding Author (Sorumlu Yazar)
talhakabakus@duzce.edu.tr

Received (Geliş Tarihi)
18.01.2024

Revised (Revizyon Tarihi)
20.04.2024

Accepted (Kabul Tarihi)
02.05.2024

Abstract

Alzheimer's Disease (AD) is one of the most, if not the most, devastating neurodegenerative diseases that are incurable and progressive. Early diagnosis of AD comes with many promises in terms of medicine, sociology, and economics. Despite the existence of numerous studies that aim for early diagnosis of AD, to the best of our knowledge, there is not a publicly available tool that lets end-users assess AD. To address this gap, we propose a Graphical User Interface (GUI) powered by Machine Learning (ML) that makes self-assessment of AD possible – without any input from medical experts. The developed GUI lets end-users enter various information considering both commonly used features for the diagnosis of AD and the questions available in the gold standard screening tool for the diagnosis of AD, namely the Mini-Mental State Exam. In addition to employing 11 traditional ML algorithms, we propose a novel 1-dimensional (1D) Convolutional Neural Network (CNN). All ML models were trained on a gold standard dataset that comprised 373 records from three subject classes as follows: (i) non-demented, (ii) demented, and (iii) converted. Once the end-user enters the required input through the developed GUI, the previously trained ML model assesses the diagnosis of AD through this input in a couple of seconds. According to the experimental results, the proposed novel 1D CNN outperformed the state-of-the-art by obtaining an accuracy as high as 95,3% on the used gold standard dataset.

Keywords: Alzheimer's Disease, classification, cognitive disorder, deep learning, dementia, machine learning.

Öz

Alzheimer Hastalığı (AH), tedavi edilemeyen ve ilerleyici olan en yıkıcı nörodejeneratif hastalıklardan biridir, belki de en yıkıcı olanıdır. AH'nin erken teşhisi, tıp, sosyoloji ve ekonomi açısından birçok avantaj içermektedir. AH'nin erken teşhisine yönelik birçok çalışma olmasına rağmen, bilgimiz dahilinde olan son kullanıcıların AH değerlendirmesini yapmalarına olanak tanıyan açık erişimli bir araç bulunmamaktadır. Bu boşluğu doldurmak için, tıbbi uzmanlardan herhangi bir giriş olmadan AH'nin kendi değerlendirmesini mümkün kılan Makine Öğrenmesi temelli bir grafiksel kullanıcı arayüzü öneriyoruz. Geliştirilen grafiksel kullanıcı arayüzü, son kullanıcılara AH teşhisi için yaygın olarak kullanılan özniteliklerle birlikte AH teşhisi için altın standart tarama aracı olan Mini-Mental Durum Testi'ndeki soruları da dikkate alarak çeşitli bilgiler girmelerine izin verir. 11 geleneksel makine öğrenmesi algoritmasının kullanımının yanı sıra, benzersiz bir 1-boyutlu Konvolüsyonel Sinir Ağı (KSA) öneriyoruz. Tüm makine öğrenmesi modelleri, (i) bilişsel bozukluğu olmayan, (ii) bilişsel bozukluğu olan ve (iii) dönüştürülen olmak üzere üç konu sınıfından oluşan 373 örneklemli bir altın standart veri setinde eğitilmiştir. Son kullanıcı, geliştirilen grafiksel kullanıcı arayüzü aracılığıyla gerekli girişi yaptığında, daha önce eğitilmiş makine öğrenmesi modeli bu girdi üzerinden AH teşhisini birkaç saniye içinde değerlendirmektedir. Deneysel sonuçlara göre, önerilen benzersiz 1-boyutlu KSA, kullanılan altın standart veri setinde %95,3'e kadar yüksek bir doğruluk elde ederek en gelişkin modelleri geride bırakmıştır.

Anahtar Kelimeler: Alzheimer Hastalığı, bilişsel bozukluk, demans, derin öğrenme, makine öğrenmesi, sınıflandırma.

[1] Düzce University, Faculty of Engineering, Department of Computer Engineering, Düzce, Türkiye.
talhakabakus@duzce.edu.tr, Orcid.org/0000-0003-2181-4292.

[2] Düzce University, Faculty of Engineering, Department of Computer Engineering, Düzce, Türkiye.
pakizeerdogmus@duzce.edu.tr, Orcid.org/0000-0003-2172-5767.

1. INTRODUCTION

Alzheimer's Disease (AD) is an irreversible neurodegenerative condition, representing the most commonly diagnosed form of dementia. It is characterized by (i) disturbances in cerebral perfusion, vasculature, and cortical metabolism, (ii) activation of proinflammatory processes, and (iii) the accumulation of amyloid beta and hyperphosphorylated Tau protein (Hnilicova et al., 2023). Patients with AD or dementia (mild to severe) often face challenges in their daily activities and may require support to ensure their lives are safe. As indicated by a recent study, there is a widespread acknowledgment of the substantial impact of Alzheimer's Disease (AD) and dementia, prompting calls for action from opinion leaders and policymakers worldwide. The global estimate of individuals affected by dementia has surpassed 50 million, incurring an annual cost exceeding 1 trillion US dollars (Gustavsson et al., 2023). Projections suggest that the number of people affected by AD and other forms of dementia is expected to reach 152 million by the year 2050 (Nichols et al., 2022). The global trajectory of AD is illustrated in Fig. 1.

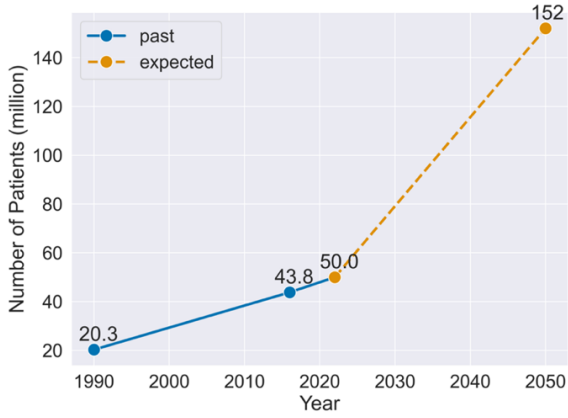

Figure 1. The Plot Depicting The Global Trend of AD (Erdogmus & Kabakus, 2023). By the year 2050, it is anticipated that the projected number of Alzheimer's disease (AD) patients will reach 152 million.

As per a 2019 study proposal, the annual global societal costs of dementia were estimated at approximately $1.313,4 billion for a population of 55.2 million individuals with dementia, equivalent to $23,796 per person with dementia. Among these costs, $213,2 billion was attributed to direct medical expenses, while $448,7 billion was associated with direct social sector costs, encompassing long-term care (Wimo et al., 2023). Therefore, the early diagnosis of AD is quite important in terms of sociology and economics. The advancement of AD can be more precisely characterized using the 7-stage model proposed by (Reisberg et al., 1982). These stages are categorized from 1 to 7, with corresponding values representing "*No impairment*", "*Very mild cognitive decline*", "*Mild cognitive decline*", "*Moderate cognitive decline*", "*Moderately severe cognitive decline*", "*Severe cognitive decline*", and "*Very severe cognitive decline*", respectively. The early identification of Alzheimer's Disease (AD) is of considerable importance, as

emphasized by (Rasmussen & Langerman, 2019). This is particularly crucial during Stages 2 or 3, as per the 7-stage model, where patients remain functionally independent and do not yet display evident signs of dementia. The potential for an early diagnosis extends up to 8 years prior to the manifestation of symptoms linked to dementia (Saxton et al., 2004).

The Mini-Mental State Exam (MMSE) (Folstein et al., 1975) is the gold-standard brief screening tool for the diagnosis of AD that was originally proposed back in 1975. The test consists of a series of questions and tasks that are designed to evaluate various aspects of cognitive functions such as arithmetic, memory, orientation, and language. The total score on the MMSE ranges from 0 to 30, with higher scores signifying better cognitive function. The test is commonly used by healthcare professionals as a quick and initial assessment of cognitive abilities.

Thanks to advances in computer software and hardware and the availability of big datasets, Artificial Intelligence (AI) has made an impact on every part of humans' lives and the field of medicine is no exception. As outlined in the literature, AI finds application in the medical domain through at least four distinct avenues: (i) evaluating the risk of disease onset and predicting treatment efficacy before commencement; (ii) addressing or mitigating complications; (iii) aiding in patient care during active treatment or procedural phases; and (iv) contributing to research endeavors focused on understanding the pathology, mechanisms, and optimal treatments for various diseases (Becker, 2019). As can be observed from Fig. 2, which illustrates the number of studies available on *PubMed* with the search term "*artificial intelligence*" by year for the last decade, the papers related to AI in the context of biomedical and life sciences literature consistently increase year-over-year.

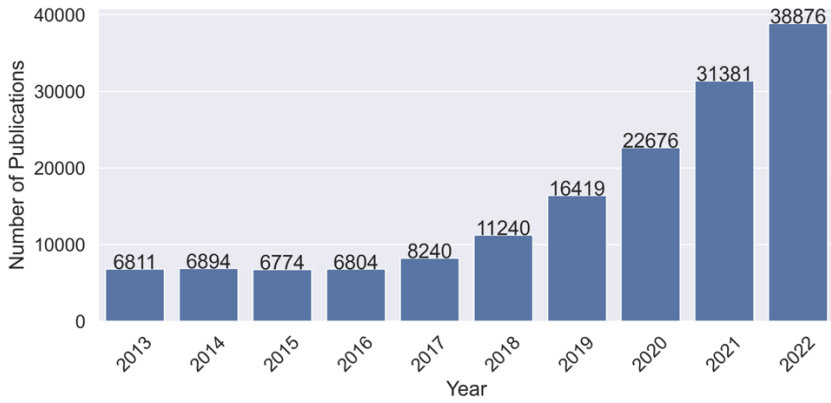

Figure 2. The Plot of the Number of Studies Available on *Pubmed* With The Search Term "*Artificial Intelligence*" For The Last Decade

In this study, we combine the promises of AI and early diagnosis of AD. To this end, we constructed a wide range of Machine Learning (ML) algorithms that were trained on a gold standard dataset that comprised the critical features for the diagnosis of AD such as MMSE score, age, gender, socioeconomic status based on the *Hollingshead*

Index of Social Position (Hollingshead, 1975), and years of education. The ML algorithms covered by this study include both widely used traditional ML algorithms and a proposed novel 1-dimensional (1D) Convolutional Neural Network (CNN). Finally, we propose a GUI for the end of real-life usage of the proposed model for early diagnosis of AD.

The main contributions of this study can be outlined as follows:

- *Self-diagnosis of AD.* We provide an AI tool with a user-friendly GUI that can be easily used by end-users to assess AD.
- *A wide range of ML models.* A total of 11 ML models were employed for early diagnosis of AD. Additionally, a novel 1D CNN was proposed.
- *High accuracy.* The proposed novel 1D CNN outperformed the state-of-the-art by obtaining an accuracy as high as 95.8% on a gold standard dataset.
- *Broad audience.* The developed GUI was intentionally implemented in the Python programming language to ensure Operating System (OS) independence. Therefore, it can be used seamlessly on any modern OS which would let it be used by a broad audience.
- *A lightweight model.* The proposed system is a lightweight and nimble model capable of classifying a given input in as low as 3 milliseconds on average.
- *Comprehensive hyperparameter optimization.* We systematically conducted a comprehensive hyperparameter optimization task, encompassing a total of 254 trials, in an automated fashion. This process is crucial for determining the optimal value for each hyperparameter in the proposed model, ensuring the derivation of the most effective model that achieves the highest classification accuracy.

The remaining sections of the paper are organized as follows: Section 2 provides an overview of related work. Section 3 outlines the materials and methods employed in the proposed study. Section 4 presents the experimental results and subsequent discussion. Lastly, Section 5 concludes the paper, offering insights into future directions.

2. RELATED WORK

AD is a progressive neurodegenerative disorder that mainly affects memory and cognitive function. As per a recent report (Early-Onset Dementia and Alzheimer's Rates Grow for Younger Americans, 2022), individuals affected by any form of dementia have an average age of 49, with diagnosis rates witnessing a 200% increase from 2013 to 2017. The etiology of AD is still unclear. It is thought that the etiology of AD is a combination of a lot of factors, such as genetic, environmental, and lifestyle factors. Autosomal dominant mutations are effective in both late-onset AD and familial AD (Ringman et al., 2015). The development of both Early-Onset Alzheimer's Disease (EOAD) and Late-Onset Alzheimer's Disease (LOAD) is significantly influenced by genetic factors, particularly mutations in APP, PSEN1, and PSEN2. In addition to genetic factors, non-genetic factors, including occupational exposures (such as exposure to pesticides, organic solvents, electromagnetic fields and volatile anesthetics), pre-existing medical conditions (hypertension, diabetes, cerebrovascular

disease, dyslipidemia, traumatic brain injury, depression, and cancer), and lifestyle factors (such as coffee consumption, smoking, alcohol and physical activity, body mass index and cognitive activity), contribute to the environmental determinants of the conditions (Jiang et al., 2013). Since AD has become more widespread in recent years, early diagnosis of AD is also becoming increasingly important. Many studies in the literature aim for early diagnosis of AD, but these studies were not only proposed by medical researchers but also by computer science researchers. The diagnosis of AD is assessed by multiple evaluations, such as clinical evaluation using a person's medical history, cognitive assessment using some mental evaluation tests such as Mini-Mental State Examination (MMSE) and a Clock Drawing Test (CDT) (Brown et al., 2019; Kato et al., 2013), daily activities, genetic tests, imaging methods such as MRI, CT, PET and SPECT and some medical tests (e.g., blood test). Plasma and cerebrospinal fluid (CSF) p-tau217 demonstrated comparable diagnostic efficacy in identifying Alzheimer's Disease (AD) based on biomarker definitions. A recent study revealed that CSF p-tau biomarkers exhibited larger effect sizes compared to plasma p-tau biomarkers in distinguishing between amyloid-PET positive and negative groups (Therriault et al., 2023). In another recent study, it is suggested that phosphorylated ATM (pATM) and/or phosphorylated APOE (pAPOE) could potentially serve as biomarkers for an early and reliable diagnosis of Alzheimer's disease (AD) using fibroblast samples (Berthel et al., 2023).

Neuroimaging plays a pivotal role in the diagnosis and progression monitoring of AD. In contemporary contexts, the integration of neuroimaging with state-of-the-art AI and Machine Learning (ML) approaches holds significant promise in enhancing diagnostic and prognostic algorithms for identifying neurodegenerative disorders and assessing the effectiveness of treatment regimens (Hnilicova et al., 2023). CNNs are specifically employed in the diagnosis of AD. While, in some studies, 1D diagnostic features are used to train 1D CNNs, in others, imaging datasets are used to train 2D CNNs. If the dataset contains 1D signals and the training data is scarce, 1D CNNs have been proposed and achieved state-of-the-art performance levels (Kiranyaz et al., 2021). 1D features have been converted to 2D images to classify AD using 2D CNNs (Erdogmus & Kabakus, 2023). Some notable large neuroimaging datasets are as follows: *Alzheimer's Disease Neuroimaging Initiative* (*ADNI*), *Australian Imaging, Biomarkers and Lifestyle Aging Flagship Study* (*AIBL*), *DZNE-Longitudinal Cognitive Impairment and Dementia Study* (*DELCODE*), and *Open Access Series of Imaging Studies* (*OASIS*). Many novel CNN models were developed using these datasets for the diagnosis of AD. Most of the time, 3D CNNs are used in related work to make more accurate predictions. However, 3D CNNs require supplementary training data to enhance their performance. When faced with limited data, combining a state-of-the-art deep 2D CNN with a shallow 3D CNN might prove more effective than utilizing a standalone 2D CNN or 3D CNN alone, as suggested by (Xu et al., 2023). According to a recent study, neurological symptoms observed in COVID-19 patients may expedite the onset of AD. In this study, researchers use speech as a non-invasive diagnostic marker. Meaningful temporal speech features, directly extracted from the recordings of the Dementia Bank, have been utilized for the classification of the AD-related Dementia Group and the health control group. Support Vector Machine (SVM) outperformed other ML algorithms with an accuracy of 87% (Karande & Kulkarni, 2023). For automated AD diagnosis in the preliminary stages, imaging methods and DL approaches have been used together. (Sharma et al., 2023) submitted a comprehensive

review focusing on automated early AD diagnosis through Deep Learning (DL) methods, covering the period from 2009 to 2022.

There are also many recent studies using cognitive tests for the diagnosis of AD. In a recent study, a novel architecture incorporating auditory, cognitive, and linguistic aspects was employed to propose a multimodal ensemble system. Scores from the MMSE have been used to assess AD severity (Sahu et al., 2022). In another study, authors proposed a system that concentrates on the MMSE and Functional Activities Questionnaire (FAQ) cognitive tests for the diagnosis of various stages of dementia using both traditional ML algorithms and Neural Network (NN) (Joshi et al., 2009). In another study, the neuropsychological dataset from the ADNI was used to classify MCI (Mild-Cognitive-Impairment) from CN (cognitively normal) using SVM, Random Forest, Gradient Boosting (GB), and AdaBoost (AB) classifiers. The SVM classifier utilizing the Radial Basis Function (RBF) kernel demonstrated the highest performance in MCI detection, achieving an average classification accuracy of 88.06% (Almubark et al., 2020). Another study proposes an analysis of the performance of supervised learning techniques to classify the data patterns of a real dataset of older adults from Pernambuco, Brazil. The study made use of genetic markers (CYP46A1 and APOE), demographic information (gender, age, and level of education, represented by study time), cognitive test scores (MMSE, Clinical Dementia Rating, Ascertaining Dementia, and Semantic Verbal Fluency Test). The objective was to develop a classifier capable of discriminating healthy and pathological aging using ML techniques. The methods applied included Artificial Neural Networks (ANNs), Random Forest, SVM, and Stochastic Gradient Boosting, as detailed in the work by (Lins et al., 2019). In a recent study, a novel multimodal deep regression model was introduced to forecast cognitive test scores by leveraging diverse data types. The multimodal dataset included cerebrospinal fluid (CSF) levels of tau and beta-amyloid, structural measurements from magnetic resonance imaging (MRI), functional and metabolic data from positron emission tomography (PET), and cognitive scores derived from neuropsychological tests (Cog). The primary objective was to attain exceptionally accurate predictions of future MMSE test scores for up to five years following the baseline biomarker collection, as outlined by (Morar et al., 2020). A comparison of the related work is listed in Table 1.

Table 1. A Comparison of The Related Work

Related Work	**Employed Method(s)**
(Kiranyaz et al., 2021)	CNN
(Erdogmus & Kabakus, 2023)	CNN
(Xu et al., 2023)	CNN
(Karande & Kulkarni, 2023)	Traditional ML algorithms
(Joshi et al., 2009)	Traditional ML algorithms and NN
(Almubark et al., 2020)	Traditional ML algorithms
(Lins et al., 2019)	Traditional ML algorithms and ANN

3. MATERIAL AND METHOD

In this section, (*i*) the used dataset and the employed data preprocessing, (*ii*) the proposed models, and (*iii*) the developed GUI are described in the following subsections, respectively.

3.1. Used Dataset and Employed Preprocessing

An ML model's success is greatly bound up with the dataset it was trained on. Therefore, we used a gold standard dataset to train the proposed ML models. The used dataset, namely *OASIS Longitudinal MRI* (Marcus et al., 2010) data, comprised 373 records from 150 subjects, where 64 of them were *demented*, 72 of them were *non-demented*, and the remaining 14 subjects were regarded as *converted*, which means they were not diagnosed with AD at the time of their initial visit; but later were diagnosed with AD. The subjects' ages range from 60 to 98. The distribution of the records available in the used dataset grouped by target class and gender is plotted in Fig. 3.

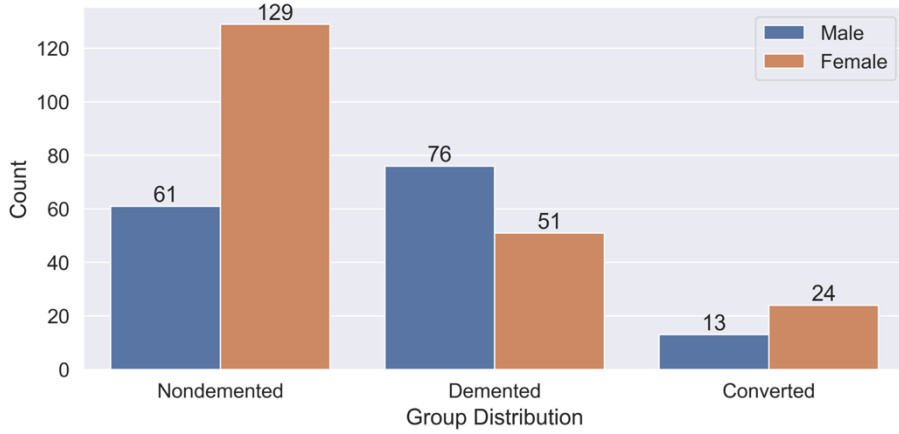

Figure 3. The Plot of the Distribution of the Records Available in the Used Dataset Grouped By Target Class and Gender

As a part of the employed preprocessing, the records with at least 1 NA (Not Available) value, which represents a missing or undefined value, were dropped from the dataset using pandas (The pandas development team, 2020), a widely used, open-source Python library for data manipulation and analysis. Consequently, the preprocessed dataset comprised 354 records, where 150 of them and 204 of them came from male and female subjects, respectively. The original dataset consists of 15 features as listed with their data types and descriptions in Table 2. The features Subject ID and MRI ID were dropped during the preprocessing since they are identically distributed. Also, the feature Hand was dropped since all subjects in the dataset were right-handed. Consequently, the preprocessed dataset consisted of the remaining 12 features. As a part of the preprocessing employed, categorical features were converted into numerical labels. All numerical features were normalized into the range of (0,1) using the

MinMaxScaler utility of scikit-learn (Pedregosa et al., 2011), which is a widely used, open-source ML library for the Python programming language. The data preprocessing employed for the proposed study is illustrated in Fig. 4.

Table 2. The List of the Features of the Used Dataset

Feature	Data Type	Description
Group	Categorical (Nominal)	The class of the record
Visit	Numerical (Interval)	The number of visits
MR Delay	Numerical (Ratio)	The days between visits
M/F	Categorical (Nominal)	Gender
Age	Numerical (Interval)	Age
EDUC	Numerical (Ratio)	Years of education
SES	Categorical (Ordinal)	Socioeconomic status assessed by the *Hollingshead Index of Social Position*
MMSE	Numerical (Ratio)	MMSE score
CDR	Numerical (Ratio)	Clinical dementia rating
eTIV	Numerical (Interval)	Estimated total intracranial volume
nWBV	Numerical (Interval)	Normalized whole-brain volume
ASF	Numerical (Interval)	Atlas scaling factor

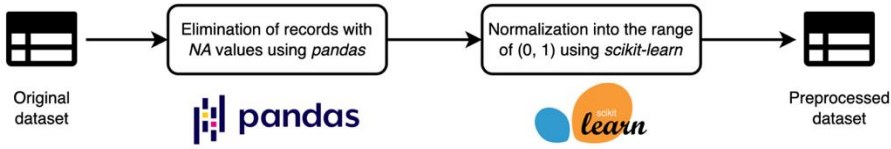

Figure 4. The Illustration of the Employed Data Preprocessing for the Proposed Study

3.2. Proposed Model

A total of 12 models were implemented to cover a wide range of different classifiers for the diagnosis of AD. More specifically, 11 of the covered models were traditional ML algorithms, which are listed in Table 3. The implementations of these algorithms, apart from Learning Vector Quantization, Light Gradient Boosting Machine (GBM), and eXtreme Gradient Boosting (XGBoost), are available in *scikit-learn* and were employed for the proposed study. For the implementation of the aforementioned three missing ML algorithms, namely (i) Learning Vector Quantization, (ii) Light GBM, and (iii) XGBoost, open-source Python packages, namely (i) *sklvq* (van Veen et al., 2021), (ii) *lightgbm* (Ke et al., 2017; *Light Gradient Boosting Machine*, 2024), and (iii) *xgboost* (Chen & Guestrin, 2016; *XGBoost Documentation*, 2024), were employed, respectively.

Table 3. The List of The Employed Traditional ML Algorithms With Their Hyperparameters

Traditional ML Algorithm	**Employed Hyperparameters**
Support Vector Machine (SVM)	- Kernel type: Linear
Logistic Regression	- Penalty: L2 - Tolerance for stopping criteria: $1xe^{-4}$ - Inverse of regularization strength: 1,0
Stochastic Gradient Descent (SGD)	- Loss function: Linear SVM - Penalty: L2 - Regularization multiplication constant: 0,0001
Naïve Bayes	- Smoothing: $1xe^{-9}$
Random Forest	- Number of trees: 100 - Quality measurement function: Gini
Decision Tree	- Quality measurement function: Gini
k-Nearest Neighbors (kNN)	- Number of neighbors (k): 3
Linear Discriminant Analysis (LDA)	- Solver: Singular value decomposition - Threshold: $1xe^{-4}$
Learning Vector Quantization	- Distance function: Squared Euclidean - Activation function: Sigmoid - Discriminant function: Relative Distance - Solver: Steepest Gradient Descent
Light GBM	- Boosting type: Gradient Boosting Decision Tree - Learning rate: 0,1 - Number of trees: 100
eXtreme Gradient Boosting (XGBoost)	- Number of trees: 100 - Objective: Binary Logistic

In addition to the employed 10 traditional ML algorithms, a novel 1D CNN, which comprised 5 layers, was proposed to benefit from the power of CNNs. *TensorFlow* (Abadi et al., 2016), a widely used ML suite from *Google*, was employed as the DL backend. *Keras* (Chollet, 2017), a high-level Application Programming Interface (API) for *TensorFlow*, was employed for the implementation of the proposed CNN. The proposed CNN starts with a 1D Convolutional layer ($Conv1D$) with 16 filters and a kernel size of 3 which refers to the dimension of the filter. $Conv1D$ layers are responsible for the convolution operation on the input through its filters. Another $Conv1D$ layer with 32 filters and a kernel size of 5 followed the first $Conv1D$ layer. A third $Conv1D$ layer with 32 filters and a kernel size of 7 followed the second $Conv1D$ layer. Then, a *Flatten* layer, which is responsible for reshaping the given input into a vector, followed. A $Dense$ layer, which is a deeply connected layer whose neurons are connected to every neuron in the preceding layer, was employed as the final layer of the proposed CNN. This $Dense$ layer was responsible for the classification of the given input into one of the three target classes, namely (i) *demented*, (ii) *non-demented*, and

(iii) *converted*. To this end, this $Dense$ layer was configured (i) to have as many units as the number of target classes (3) and (ii) $Softmax$ was employed as the activation function which converts a vector of real numbers into a probability distribution and the class with the highest probability is predicted as the output of the network. The loss function of a Deep Neural Network (DNN) is responsible for estimating the loss of the model. $Categorical\ Cross-Entropy$ was employed as the loss function of the proposed CNN. Fig. 5 illustrates an overview of the architecture of the proposed novel CNN.

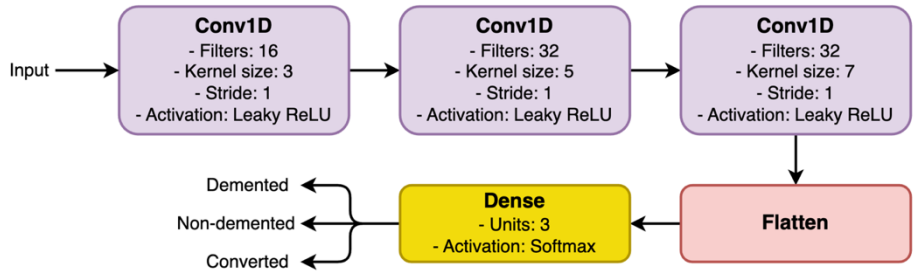

Figure 5. The Illustration of A High-Level View of the Architecture of the Proposed Novel CNN

Hyperparameters of a DNN are the parameters that are set empirically and affect the learning process (Chollet, 2017; Zhang et al., 2019). The official tuner of *Keras*, *Keras Tuner* (O'Malley et al., 2019), was employed for the hyperparameter optimization of the proposed CNN. The optimization algorithm chosen for the hyperparameter optimization task was *Hyperband* (Li et al., 2018), a widely utilized method that combines random search with aggressive early stopping to efficiently explore the hyperparameter space. *Hyperband* offers several advantages, including efficiency, scalability, flexibility, and the ability to provide state-of-the-art performance. The objective of the optimization task was to minimize the loss obtained for the validation set. An activation function is a mathematical operation that introduces non-linearity into the model. The optimization algorithm of a DNN is responsible for adjusting the parameters of a model to minimize the error or loss of the model calculated by the employed loss function. The learning rate determines the size of the steps made to the model's weights during the training. As listed in Table 4, the hyperparameters (i) activation function, (ii) optimization algorithm, (iii) learning rate, and (iv) batch size, which is the number of training examples utilized in one iteration, were optimized through the employed hyperparameter optimization task. The obtained optimum value for each hyperparameter is given in bold in Table 4. As can be seen in this table, $Leaky\ Rectified\ Linear\ Unit\ (ReLU)$, $Adamax$ (Kingma & Ba, 2015), $1xe^{-2}$, and 8 were employed as the activation function, optimization algorithm, learning rate, and batch size of the proposed CNN, respectively. $Leaky\ ReLU$ is a variant of $ReLU$ that has a small slope for negative values instead of a flat slope. $Adamax$ is a variant of $Adam$ based on the infinity norm (Kingma & Ba, 2015).

Table 4. The list of hyperparameters optimized through the employed hyperparameter optimization task. The obtained optimum value for each hyperparameter is given in bold.

Hyperparameter	**Evaluated Values**
Activation function	$ReLU$, eLU, $\boldsymbol{Leaky\ ReLU}$, $PReLU$
Optimization algorithm	$Adam$, $\boldsymbol{Adamax}$, $Adadelta$, SGD, $RMSprop$
Learning rate	$\boldsymbol{1xe^{-2}}$, $1xe^{-3}$, $1xe^{-4}$
Batch size	**8**, 16, 32, 64

3.3. Developed GUI

For the self-assessment of the AD, we developed a GUI using the Python programming language, which was intentionally chosen since it is platform-independent. Therefore, the developed GUI can be run in any environment that supports the execution of Python. More specifically, *CustomTkinter* (Schimansky, 2024), a modern and customizable Python UI library based on the standard Python module *tkinter*, was employed for the GUI implementation. The developed GUI assesses AD through the features of the used dataset. The features that require medical experts' input were eliminated since we aim to provide a GUI capable of self-assessment of AD in a couple of minutes. These features are (i) age, (ii) gender, (iii) years of education (EDUC), (iv) socioeconomic status (SES), and (v) the score of the MMSE, which requires the exact questions of MMSE that investigates the cognitive impairment of the subject. Therefore, the answers to the questions in MMSE were taken through the keyboard and microphone. To vocalize the sentences of MMSE, we employed a Python library, namely *gTTS* (Durette, 2023), that interfaces with the *Google Text-to-Speech API*. The employed text-to-speech mechanism is illustrated in Fig. 6.

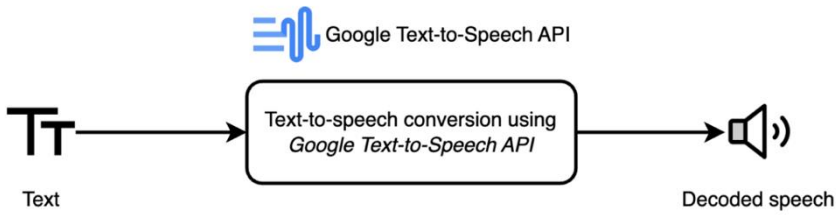

Figure 6. The Illustration of the Employed Text-to-Speech Mechanism

To evaluate the accuracy of the voice input taken from the microphone, it is necessary to convert it to text. To this end, we employed OpenAI's Whisper (Radford et al., 2023), a versatile speech recognition model trained on a comprehensive dataset of diverse audio inputs. Whisper computes the log-mel spectrogram of the given sound to make the input ready to be yielded into the proposed 1D CNN. The output of the model is the decoded text in the declared natural language, which was English in our case. It is worth mentioning that Whisper requires the command-line tool FFmpeg (Bellard, 2023) to be installed on the operating system. The employed speech-to-text mechanism is illustrated in Fig. 7.

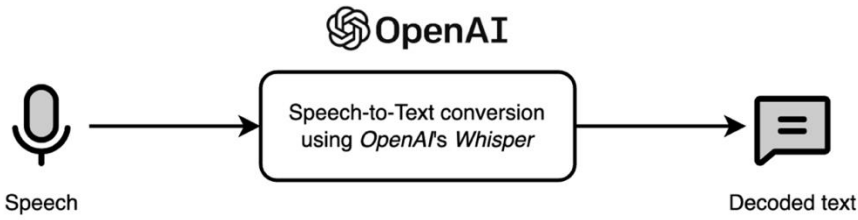

Figure 7. The Illustration of the Employed Text-to-Speech Mechanism

As can be observed from Fig. 8, which presents the developed GUI for early diagnosis of AD, the developed GUI takes answers from end-users to the provided 21 questions. While 4 of them, namely (i) "*Select your gender*", (ii) "*Enter your age*", (iii) "*Enter your years of education*", and (iv) "*Enter your socioeconomic status*" were taken from the used dataset, the remaining 17 questions were taken from the gold standard MMSE. The ground truths to the questions "*Enter the current year*", "*Enter the current season*", "*Enter the current month*", "*Enter the current day of month*", "*Enter the name of the current day of month*", "*Enter the country you live in*", and "*Enter the city you're in*" were obtained programmatically. More specifically, the date and time were obtained from the system. The location information was obtained from the *IPInfo* website (*IPinfo*, 2023). It is worth mentioning that *IPInfo* provides the ISO-standard country code of the current location instead of the country name. The retrieved country code was mapped to the corresponding country name using a widely used, open-source Python library, namely *pycountry* (Theune, 2023). The ground truths to the questions "*Enter the state you're in*", "*Enter the name of the apartment you're in*", and "*Enter the floor you're on*" were retrieved from a JSON file located in the same folder of the developed GUI's source code with the keys "*state*", "*apartment*", and "*floor*", respectively. Once the end-user clicks the "*Evaluate Mini-Mental State Test*" button after entering the required inputs for the assessment of AD, a dialog window, such as the one presented in Fig. 9, is displayed to inform the end-user regarding the diagnosis result. This sample result was obtained when the input was evaluated with the best-performing model, the proposed 1D CNN.

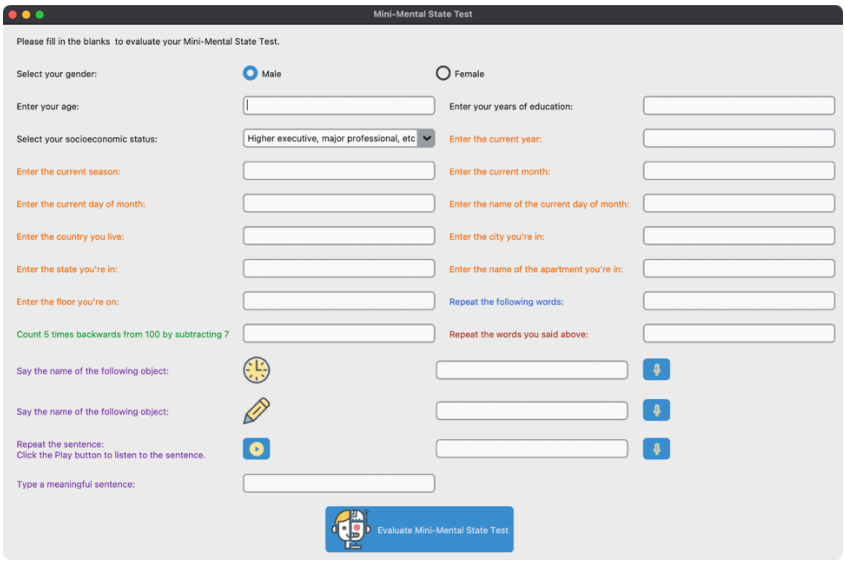

Figure 8. The Main UI of the Developed Tool for Early Diagnosis of AD

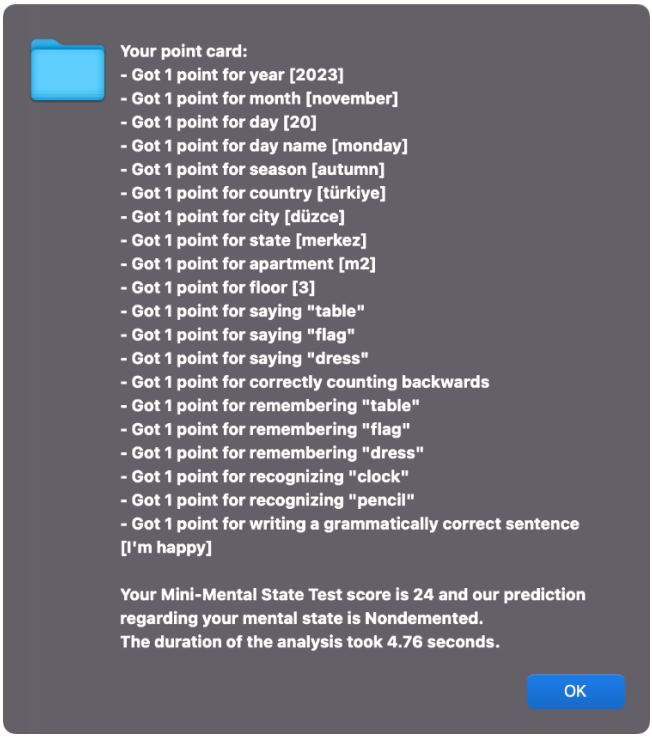

Figure 9. The UI of the Developed Tool That Displays the Diagnosis Result of AD

4. EXPERIMENTAL RESULTS AND DISCUSSION

The metrics used to evaluate the classification performance of the proposed ML models are described in the following subsection. Then, in the next subsections, we present the experimental results and discussion for the proposed two approaches: (i) the ML models based on the feature subset of the GUI, and (ii) the ML models based on all the features available in the used gold standard dataset, respectively.

4.1. Evaluation Metrics

The *de facto* standard metrics that are used to evaluate the performance of classifiers, namely (i) $Accuracy$, (ii) $Precision$, (iii) $Recall$, and (iv) $F1-score$ were employed for the evaluation of the proposed ML models. Let P denote positives, representing samples labeled with the target class, and N represent negatives, indicating samples labeled with the complementary class of the target. TP, TN, FP, and FN denote correctly predicted *positives*, correctly predicted *negatives*, *positives* incorrectly predicted as *negative*, and *negatives* incorrectly predicted as *positive*, respectively. $Accuracy$ is the ratio of correctly predicted samples to the total number of samples. The equation of $accuracy$ is given in Eq. 1. $Precision$ is defined as the ratio of accurately predicted positive instances to the total number of instances predicted as positive. The equation of $precison$ is given in Eq. 2. $Recall$ is the ratio of correctly predicted positive instances to the total number of actual positive instances. The equation of $recall$ is given in Eq. 3. $F1-score$ is the harmonic mean of the $Precision$ and $Recall$ and is more useful than $Accuracy$ when dealing with imbalanced datasets. The equation of $F1-score$ is given in Eq. 4.

$$Accuracy = (TP + TN) \,/\, (P + N) \tag{1}$$

$$Precision = TP/\,(TP + FP) \tag{2}$$

$$Recall = TP \,/\, (TP + FN) \tag{3}$$

$$F1-score = 2\,x\,(Precision\ x\ Recall) \,/\, (Precision + Recall) \tag{4}$$

When it comes to visualization of the evaluation results, we employed a confusion matrix, which is a specific table that provides a comprehensive summary of the model's predictions compared to the actual outcomes. The confusion matrix serves as a valuable tool for understanding the strengths and weaknesses inherent in a classification model.

4.2. Experimental Results and Discussion of ML Models Based on the Developed GUI

70% of the constructed dataset was allocated as the training set, while the remaining 30% was designated as the test set. Consequently, the training set comprised 283 samples, while the test set comprised 71 samples. *Scikit-learn* was employed for the dataset splitting with shuffling. All the employed ML models were trained and tested on the same designated subsets. More specifically, the proposed novel 1D CNN was trained for 50 epochs and 30% of the training set was used as the validation set, which

is a distinct subset of the data that the model has not seen during training. The accuracy values for both the training and validation sets during the model's training are depicted in Fig. 10, which confirms neither overfitting nor underfitting was the case for the model.

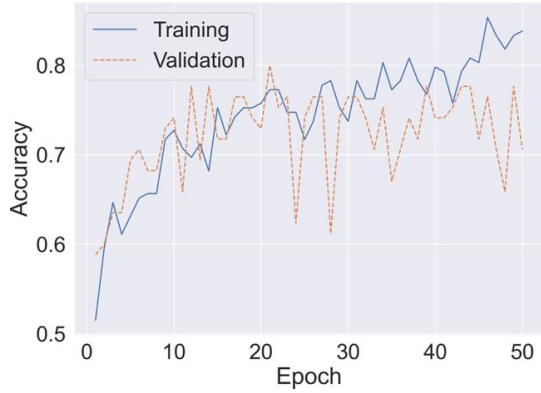

Figure 10. The Graph Illustrating The Accuracy Values Acquired for Both The Training and Validation Sets Throughout The Training Process of the Proposed Novel 1D CNN Based on the Developed GUI

As can be observed from Fig. 11, which presents the obtained accuracy values when the employed ML models were evaluated on the test set, both Random Forest and kNN achieved the best accuracy, an accuracy of as high as 87,32%, among all the employed ML models. The employed kNN achieved a higher F1-score (84.97%) compared to the employed Random Forest (81,71%). Therefore, the employed kNN was regarded as the best model for the early diagnosis of AD. As the confusion matrix of the evaluation result of the employed kNN is presented in Fig. 12, 62 of 71 test samples were correctly classified. The employed kNN was persisted into a file, allowing it to be utilized within the developed GUI.

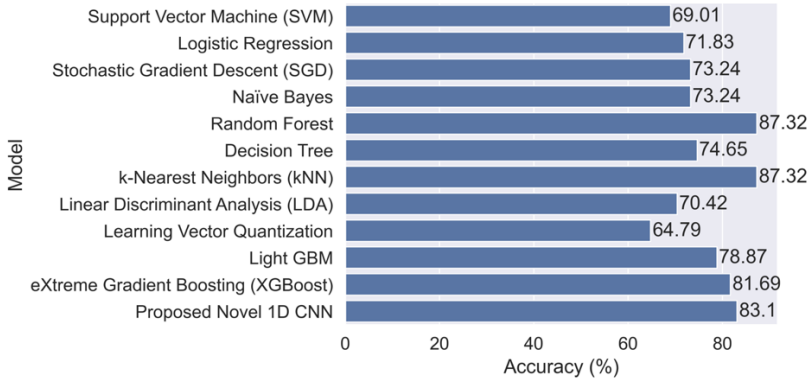

Figure 11. The Evaluation Results of the Employed ML Models Using The Dataset That Covers The Feature Subset of the Developed GUI

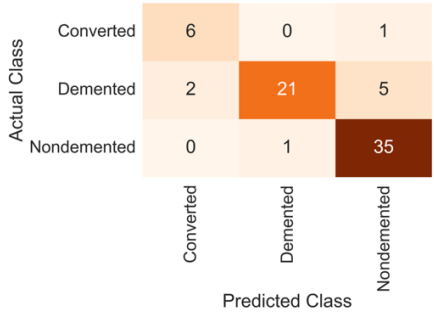

Figure 12. The Confusion Matrix of the Employed kNN, Which Achieved The Best Classification Performance Among All The Employed ML Models

4.3. Experimental Results and Discussion of ML Models Based on Used Gold Standard Dataset

Our study aimed to assess the performance of the proposed ML models in comparison to the state-of-the-art. To provide a quantitative comparison of the proposed ML models with the state-of-the-art, we also conducted experiments using all the features available in the used gold standard dataset. The preprocessing employed in the previous approach was also used for this experiment. Therefore, the training set comprised 283 samples and the test set comprised the remaining 71 samples. Unlike the previous approach, this time each sample of the dataset comprised 11 features instead of 5. All the employed ML models were trained and tested on the same designated subsets. More specifically, the training of the proposed novel 1D CNN utilized the Early Stopping callback. This component actively monitored the model's performance on a validation set throughout the training process and halted the training when signs of performance degradation were detected. The Early Stopping callback was employed with the following two hyperparameters: (i) the loss calculated for the validation set was monitored, and (ii) the *patience*, which is the number of consecutive epochs with no improvement on the monitored criterion before the training process is halted, was set 20. The training of the proposed novel 1D CNN was extended for a total of 58 epochs, concluding when the Early Stopping callback intervened and terminated the training process. Consistent with the prior approach, 30% of the training set served as the validation set. The accuracy values for both the training and validation sets were plotted during the model's training, as depicted in Fig. 13. The results indicate that neither overfitting nor underfitting occurred in the model.

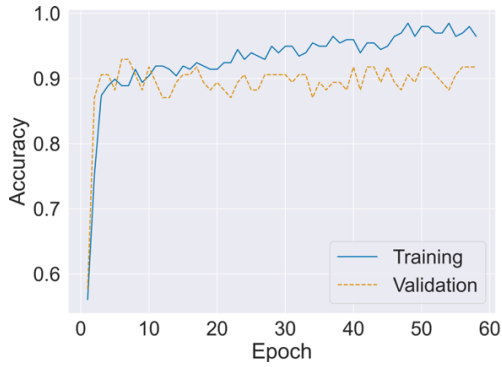

Figure 13. The Plot Depicting The Accuracy Values Attained for Both The Training and Validation Sets During The Training Phase of the Proposed Novel 1D CNN Based on Used Gold Standard Dataset

As evident from Fig. 14, depicting the accuracy values obtained when evaluating the employed ML models on the test set, the proposed novel 1D CNN outperformed all other models with an accuracy as high as 95,8%. For a more comprehensive classification comparison, the comparison of the employed ML models in terms of all evaluation metrics is listed in Table 5.

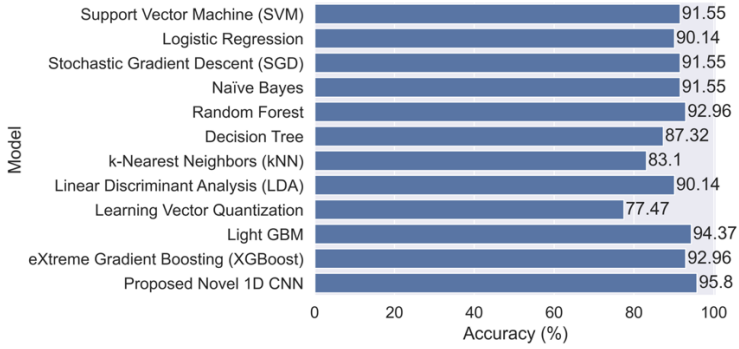

Figure 14. The Evaluation Results of the Employed ML Models Using The Dataset That Covers All Features

Table 5. The evaluation results of the employed ML models based on all metrics utilized, utilizing the dataset containing all features.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Support Vector Machine (SVM)	91,55	94,44	71,43	72,1
Logistic Regression	90,14	93,64	70,24	71,08
Stochastic Gradient Descent (SGD)	91,55	94,44	71,43	72,1
Naïve Bayes	91,55	84,37	75	77,06
Random Forest	92,96	95,21	76,19	79
Decision Tree	87,32	75,79	75,79	75,79
k-Nearest Neighbors (kNN)	83,1	70,11	68,39	69,01
Linear Discriminant Analysis (LDA)	90,14	93,64	70,24	71,08
Learning Vector Quantization	77,47	56,41	55,96	54,22
Light GBM	94,37	96,02	80,95	84,62
eXtreme Gradient Boosting (XGBoost)	92,96	87,64	80,03	82,33
Proposed novel 1D CNN	**95,8**	**95,62**	**95,59**	**95,78**

As listed in Table 6, this experimental result indicates that the proposed novel 1D CNN outperformed the state-of-the-art that used the same dataset. As the confusion matrix of the evaluation result of the proposed novel 1D CNN is presented in Fig. 15, 68 of 71 test samples were correctly classified. In other words, only 3 (1 demented and 2 converted) test samples were incorrectly classified. The recall of the target class "demented" was calculated as high as 100%, which indicates that all demented subjects were correctly identified.

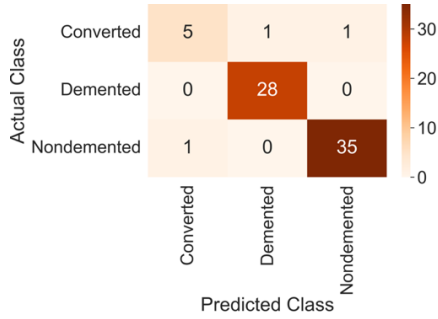

Figure 15. The Confusion Matrix of the Proposed Novel 1D CNN, Which Achieved The Best Classification Performance Among All The Employed ML Models

Table 6. A Quantitative Comparison of the Proposed 1D CNN With Studies in the Literature That Used The Same Dataset

Related Work	**Employed Method**	**Accuracy (%)**
Amrutesh et al. (Amrutesh et al., 2022)	Traditional ML algorithms	92,13
Almubark et al. (Almubark et al., 2020)	Traditional ML algorithms	88,06
Chowdary et al. (Chowdary et al., 2021)	Traditional ML algorithms	89,4
AbdElminaam et al. (Abdelminaam et al., 2023)	Traditional ML algorithms	86,9
Vidushi et al. (Vidushi et al., 2020)	Traditional ML algorithms	84,21
Kavitha et al. (Kavitha et al., 2022)	Traditional ML algorithms	86,92
Jadhao et al. (Jadhao et al., 2023)	Traditional ML algorithms	85,71
Ozhan et al. (Ozhan et al., 2022)	Artificial Neural Network	94,7
Shetty et al. (Shetty et al., 2022)	Traditional ML algorithms	91,99
Arjaria et al. (Arjaria et al., 2022)	Traditional ML algorithms	90
This study	**1D CNN**	**95,8**

4.4. Known Limitations

The developed GUI misses the features that require input from medical experts. Also, we eliminated some questions of MMSE that require analysis of the performed actions. Even though the developed GUI can record these actions through a webcam, as of now, we are not able to programmatically interpret a given video recording. This limitation is regarded as a future work of this study.

6. CONCLUSION

AI plays a crucial role in healthcare, contributing to a diverse range of key applications. These include, but are not limited to, medical imaging and diagnostics, predictive analysis for patient outcomes, virtual healthcare assistants, robot-assisted surgery, and mental health monitoring. Similarly, AI is instrumental in facilitating early disease diagnosis. In this study, we introduce a GUI integrated with machine learning capabilities, enabling end-users to seamlessly assess the early diagnosis of AD, which is one of the most, if not the most, devastating neurodegenerative diseases. To achieve

this, we initially trained a variety of ML models using a gold standard dataset specifically designed for early AD diagnosis, consisting of 373 records from 150 subjects. Addressing the lack of publicly available GUIs for early AD diagnosis, we present a user-friendly interface powered by the proposed ML model. The proposed models were evaluated on this gold standard dataset to reveal their efficiencies in terms of classifying a given subject into one of the three target subject classes: (i) *non-demented*, (ii) *demented*, and (iii) *converted*. According to the experimental results, the proposed novel 1D CNN outperformed the state-of-the-art by obtaining an accuracy as high as 95,8%.

In future endeavors, our aim is to incorporate features that were initially excluded from consideration due to their reliance on input from medical experts. This entails finding a method to interpret video recordings effectively. This enhancement would enable us to (i) introduce an AI tool for the early diagnosis of AD that encompasses all the standard questions of the MMSE and (ii) potentially enhance the classification accuracy of the proposed model by leveraging the enriched feature set. Additionally, we intend to explore the utilization of other types of DNNs (e.g., RNNs) to further improve the obtained classification accuracy. Finally, we would like to employ data augmentation techniques to achieve a balanced dataset.

Authors Contribution

Abdullah Talha Kabakus contributed for Methodology, Software, Investigation, Writing – original draft, Reviewing, Editing. *Pakize Erdogmus* contributed for Methodology, Conceptualization, Investigation, Writing – original draft, Reviewing, Editing.

Competing Interest

The authors declare that they have no competing interests related to the content of this manuscript.

Ethical and Informed Consent for Data Used

Ethical approval and informed consent for data use were not required for this research, as the study did not involve human subjects, personal data, or sensitive information. The data utilized were obtained from publicly available and anonymized sources, and all aspects of the research adhered to ethical standards and legal requirements.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016)*, 265–283.

Abdelminaam, D. S., Madbouly, M. M., Farag, M. S., Gomaa, I. A., Abd-Elghany Zeid, M., & Abualigah, L. (2023). ML_Alzheimer: Alzheimer Disease Prediction Using Machine Learning. *Proceedings of the 3rd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC 2023)*, 409–414. https://doi.org/10.1109/MIUCC58832.2023.10278361

Almubark, I., Alsegehy, S., Jiang, X., & Chang, L. C. (2020). Early Detection of Mild Cognitive Impairment using Neuropsychological Data and Machine Learning Techniques. *Proceedings of the 2020 IEEE Conference on Big Data and Analytics (ICBDA 2020)*, 32–37. https://doi.org/10.1109/ICBDA50157.2020.9289741

Amrutesh, A., Gowtham Bhat, C. G., Amruthamsh, A., Asha Rani, K. P., & Gowrishankar, S. (2022). Alzheimer's Disease Prediction using Machine Learning and Transfer Learning Models. *Proceedings of the 6th IEEE International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS 2022)*, 1–6. https://doi.org/10.1109/CSITSS57437.2022.10026365

Arjaria, S. K., Rathore, A. S., Bisen, D., & Bhattacharyya, S. (2022). Performances of Machine Learning Models for Diagnosis of Alzheimer's Disease. *Annals of Data Science*, 1–29. https://doi.org/10.1007/s40745-022-00452-2

Becker, A. (2019). Artificial intelligence in medicine: What is it doing for us today? *Health Policy and Technology*, *8*(2), 198–205. https://doi.org/10.1016/J.HLPT.2019.03.004

Bellard, F. (2023). *FFmpeg*. Retrieved January 1, 2024 from https://ffmpeg.org

Berthel, E., Pujo-Menjouet, L., Le Reun, E., Sonzogni, L., Al-Choboq, J., Chekroun, A., Granzotto, A., Devic, C., Ferlazzo, M. L., Pereira, S., Bourguignon, M., & Foray, N. (2023). Toward an Early Diagnosis for Alzheimer's Disease Based on the Perinuclear Localization of the ATM Protein. *Cells*, *12*(1747), 1–21. https://doi.org/10.3390/cells12131747

Brown, J., Wiggins, J., Lansdall, C. J., Dawson, K., Rittman, T., & Rowe, J. B. (2019). Test Your Memory (TYM test): diagnostic evaluation of patients with non-Alzheimer dementias. *Journal of Neurology*, *266*(10), 2546–2553. https://doi.org/10.1007/s00415-019-09447-1

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.

Chowdary, B. V., Muppidi, S., Sruthi, B., Madhuri, K. S., & Sumanth, L. (2021). An Effective and Efficient Alzheimer Disease Prediction System Using Machine Learning Model. *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2021)*, 342–347. https://doi.org/10.1109/I-SMAC52330.2021.9641022

Durette, P. N. (2023). *gTTS*. Retrieved January 1, 2024 https://gtts.readthedocs.io/en/latest/

Early-Onset Dementia and Alzheimer's Rates Grow for Younger Americans. (2022). https://doi.org/10.9

Erdogmus, P., & Kabakus, A. T. (2023). The promise of convolutional neural networks for the early diagnosis of the Alzheimer's disease. *Engineering Applications of Artificial Intelligence*, *123*, 1–13. https://doi.org/10.1016/j.engappai.2023.106254

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Gustavsson, A., Norton, N., Fast, T., Frölich, L., Georges, J., Holzapfel, D., Kirabali, T., Krolak-Salmon, P., Rossini, P. M., Ferretti, M. T., Lanman, L., Chadha, A. S., & van der Flier, W. M. (2023). Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's and Dementia*, *19*(2). https://doi.org/10.1002/alz.12694

Hnilicova, P., Kantorova, E., Sutovsky, S., Grofik, M., Zelenak, K., Kurca, E., Zilka, N., Parvanovova, P., & Kolisek, M. (2023). Imaging Methods Applicable in the Diagnostics of Alzheimer's Disease, Considering the Involvement of Insulin Resistance. In *International Journal of Molecular Sciences* (Vol. 24, Issue 3325, pp. 1–31). https://doi.org/10.3390/ijms24043325

Hollingshead, A. (1975). Four factor index of social status. In *Yale Journal of Sociology* (Vol. 8).

IPinfo. (2023). Retrieved January 1, 2024 https://ipinfo.io

Jadhao, P., Palsodkar, P., Raut, R., Chaube, K., Rathod, D., & Palsodkar, P. (2023). Prediction of Early Stage Alzheimer's using Machine Learning Algorithm. *2023 4th International Conference for Emerging Technology, INCET 2023*, 1–5. https://doi.org/10.1109/INCET57972.2023.10170583

Jiang, T., Yu, J.-T., Tian, Y., & Tan, L. (2013). Epidemiology and Etiology of Alzheimer's disease: From Genetic to Non-Genetic Factors. *Current Alzheimer Research*, *10*(8), 852–867. https://doi.org/10.2174/15672050113109990155

Joshi, S., Shenoy, P. D., Venugopal, K. R., & Patnaik, L. M. (2009). Evaluation of Different Stages of Dementia Employing Neuropsychological and Machine Learning Techniques. *Proceedings of the 2009 1st International Conference on Advanced Computing (ICAC 2009)*, 154–160. https://doi.org/10.1109/ICADVC.2009.5378199

Karande, S., & Kulkarni, V. (2023). Automated Prognosis of Alzheimer's Disease using Machine Learning Classifiers on Spontaneous Speech Features. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(2), 245–251.

Kato, Y., Narumoto, J., Matsuoka, T., Okamura, A., Koumi, H., Kishikawa, Y., Terashima, S., & Fukui, K. (2013). Diagnostic performance of a combination of Mini-Mental State Examination and Clock Drawing Test in detecting Alzheimer's disease. *Neuropsychiatric Disease and Treatment*, *9*, 581–586. https://doi.org/10.2147/NDT.S42209

Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., & Tavera Romero, C. A. (2022). Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. *Frontiers in Public Health*, *10*, 1–13. https://doi.org/10.3389/fpubh.2022.853294

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*(NIPS 2017), 3149–3157.

Kingma, D. P., & Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, 1–15.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, 1–21. https://doi.org/10.1016/j.ymssp.2020.107398

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018). Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *Journal of Machine Learning Research*, *18*(1), 6765–6816.

Light Gradient Boosting Machine. (2024). Microsoft. Retrieved January 1, 2024 https://lightgbm.readthedocs.io

Lins, A. J. C. C., Muniz, M. T. C., & Bastos-Filho, C. J. A. (2019). Comparing Machine Learning Techniques for Dementia Diagnosis. *Proceedings of the 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI 2018)*, 1–6. https://doi.org/10.1109/LA-CCI.2018.8625209

Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2010). Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*, *22*(12), 2677–2684. https://doi.org/10.1162/jocn.2009.21407

Morar, U., Martin, H., Izquierdo, W., Forouzannezhad, P., Zarafshan, E., Curiel, R. E., Roselli, M., Loewenstein, D., Duara, R., Unger, E., & Adjouadi, M. (2020). A Deep-Learning Approach for the Prediction of Mini-Mental State Examination Scores in a Multimodal Longitudinal Study. *Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence (CSCI 2020)*, 761–766. https://doi.org/10.1109/CSCI51800.2020.00144

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*, 105–125. https://doi.org/10.1016/S2468-2667(21)00249-8

O'Malley, T., Bursztein, E., Long, J., & Chollet, F. (2019). *KerasTuner*. Keras. Retrieved January 1, 2024 https://github.com/keras-team/keras-tuner

Ozhan, O., Kucukakcali, Z., & Balikci Cicek, I. (2022). Risk Prediction Model for Dementia by Deep Learning Using Clinical Data. *The Journal of Cognitive Systems*, *7*(2), 1–4. https://doi.org/10.52876/jcs

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust Speech Recognition via Large-Scale Weak Supervision. *ArXiv*, *2212.04356*, 1–28. https://doi.org/10.48550/arXiv.2212.04356

Rasmussen, J., & Langerman, H. (2019). Alzheimer's Disease – Why We Need Early Diagnosis. *Degenerative Neurological and Neuromuscular Disease*, *9*, 123–130. https://doi.org/10.2147/dnnd.s228939

Reisberg, B., Ferris, S. H., De Leon, M. J., & Crook, T. (1982). The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, *139*, 1136–1139. https://doi.org/10.1176/ajp.139.9.1136

Ringman, J. M., Liang, L. J., Zhou, Y., Vangala, S., Teng, E., Kremen, S., Wharton, D., Goate, A., Marcus, D. S., Farlow, M., Ghetti, B., McDade, E., Masters, C. L., Mayeux, R. P., Rossor, M., Salloway, S., Schofield, P. R., Cummings, J. L., Buckles, V., … Morris, J. C. (2015). Early behavioural changes in familial Alzheimer's disease in the Dominantly Inherited Alzheimer Network. *Brain*, *138*(4), 1036–1045. https://doi.org/10.1093/brain/awv004

Sahu, H. K., Kumar, S., Alsamhi, S. H., Chaube, M. K., & Curry, E. (2022). Novel Framework for Alzheimer Early Diagnosis using Inductive Transfer Learning Techniques. *Proceedings of the 2022 2nd International Conference on Emerging Smart Technologies and Applications (ESmarTA 2022)*, 1–7. https://doi.org/10.1109/eSmarTA56775.2022.9935379

Saxton, J., Lopez, O. L., Ratcliff, G., Dulberg, C., Fried, L. P., Carlson, M. C., Newman, A. B., & Kuller, L. (2004). Preclinical Alzheimer disease: Neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology*, *63*(12), 2341–2347. https://doi.org/10.1212/01.WNL.0000147470.58328.50

Schimansky, T. (2024). *CustomTkinter*. Retrieved January 1, 2024 https://customtkinter.tomschimansky.com

Sharma, R., Goel, T., Tanveer, M., Lin, C. T., & Murugan, R. (2023). Deep-Learning-Based Diagnosis and Prognosis of Alzheimer's Disease: A Comprehensive Review. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1123–1138. https://doi.org/10.1109/TCDS.2023.3254209

Shetty, M., Deekshitha, Bhat, M., & Devadiga, M. (2022). Detection of Alzheimer's Disease Using Machine Learning. *International Conference on Artificial Intelligence and Data Engineering, AIDE 2022*, 117–120. https://doi.org/10.1109/AIDE57180.2022.10060433

The pandas development team. (2020). *pandas: Python Data Analysis Library*. Retrieved January 1, 2024 https://pandas.pydata.org

Therriault, J., Servaes, S., Tissot, C., Rahmouni, N., Ashton, N. J., Benedet, A. L., Karikari, T. K., Macedo, A. C., Lussier, F. Z., Stevenson, J., Wang, Y. T., Fernandez-Arias, J., Stevenson, A., Socualaya, K. Q., Haeger, A., Nazneen, T., Aumont, É., Hosseini, A., Rej, S., … Rosa-Neto, P. (2023). Equivalence of plasma p-tau217 with cerebrospinal fluid in the diagnosis of Alzheimer's disease. *Alzheimer's and Dementia*, *19*(11), 4967–4977. https://doi.org/10.1002/alz.13026

Theune, C. (2023). *pycountry: A Python library to access ISO country, subdivision, language, currency and script definitions and their translations*. Retrieved January 1, 2024 https://github.com/flyingcircusio/pycountry

van Veen, R., Biehl, M., & de Vries, G. J. (2021). sklvq: Scikit Learning Vector Quantization. *Journal of Machine Learning Research*, *22*(231), 1–6.

Vidushi, M., Akash, R., & Shrivastava, A. K. (2020). Diagnosis of Alzheimer Disease using Machine Learning Approaches. *International Journal of Advanced Science and Technology*, *29*(04), 7062–7073.

Wimo, A., Seeher, K., Cataldi, R., Cyhlarova, E., Dielemann, J. L., Frisell, O., Guerchet, M., Jönsson, L., Malaha, A. K., Nichols, E., Pedroza, P., Prince, M., Knapp, M., & Dua, T. (2023). The worldwide costs of dementia in 2019. *Alzheimer's and Dementia*, *19*(7). https://doi.org/10.1002/alz.12901

XGBoost Documentation. (2024). Retrieved January 1, 2024 https://xgboost.readthedocs.io

Xu, X., Lin, L., Sun, S., & Wu, S. (2023). A review of the application of three-dimensional convolutional neural networks for the diagnosis of Alzheimer's disease using neuroimaging. *Reviews in the Neurosciences*, *34*(6), 649–670. https://doi.org/10.1515/revneuro-2022-0122

Zhang, X., Chen, X., Yao, L., Ge, C., & Dong, M. (2019). Deep Neural Network Hyperparameter Optimization with Orthogonal Array Tuning. *International Conference on Neural Information Processing (ICONIP 2019)*, 287–295. https://doi.org/10.1007/978-3-030-36808-1_31