



HAL
open science

Dementia detection based on speech acoustics using machine learning

Ahmad Tay, Mihai Andries, Christophe Lohr

► **To cite this version:**

Ahmad Tay, Mihai Andries, Christophe Lohr. Dementia detection based on speech acoustics using machine learning. 2024. hal-04612243

HAL Id: hal-04612243

<https://hal.science/hal-04612243v1>

Preprint submitted on 14 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Dementia detection based on speech acoustics using machine learning

Ahmad Tay^{a,b,*}, Mihai Andries^a, Christophe Lohr^a

^a*IMT Atlantique, Lab-STICC, 655 Avenue du Technopôle, 29280, Plouzané, France*

^b*Efrei Research Lab, 30 Avenue de la République, 94800, Villejuif, France*

Abstract

Background: The work presented here is part of a larger study to identify voice markers for early dementia detection and it focuses on evaluating the suitability of a new approach for early diagnosis by non-invasive methods.

Methods: In this context, we used class-dependent principal component analysis for feature engineering and three machine learning techniques, namely, logistic regression, support vector machines, and artificial neural networks for the automatic classification of the two classes (dementia and control subjects).

Findings: We developed a non-invasive, low cost, and side-effects free approach. Our method also comprises a small number of variables and does not require heavy computing power. The developed model showed that speech parameters constitute a promising biomarker for dementia detection.

Results: The obtained experimental results were satisfactory and promising when evaluated on the test set (accuracy=0.972, precision=0.983, recall=0.968, and F1-score=0.975), making the model reliable for early de-

*Corresponding author

Email address: `firstname.lastname@imt-atlantique.fr` (Ahmad Tay)

mentia detection.

Keywords: Dementia disease, artificial neural networks, class-dependent principal component analysis, speech analysis, classification

1. Introduction

1.1. Describing the problem

The World Health Organization (WHO) defines dementia as a syndrome that leads to deterioration in cognitive functions and is the seventh leading cause of death among all diseases (WHO, 2019). It is characterized by the decline in memory, thinking, and the ability to perform daily activities (Chen et al., 2022). There are currently around 55 million people diagnosed with dementia and the number is expected to reach 139 million in 2050 (WHO, 2019). Alzheimer’s disease is the leading cause of neurodegenerative dementia and is responsible for around two thirds of all its diagnoses (Rasmussen and Langerman, 2019). Unfortunately, no ultimate curative treatments for dementia currently exist (Yadav, 2019). Nevertheless, early detection can still help to slow down dementia’s progression by maintaining the patients’ quality of life and managing dementia symptoms (for example, medicines to control blood pressure and cholesterol can prevent additional damage to the brain due to vascular dementia) (WHO, 2019). According to Roeck et al. (2019), premature diagnosis is imperative to preserve good living conditions and promote early intervention, including counseling, psycho-education, cognitive training, and medication. This intervention can help in employing control measures to delay the onset of this disease. Hence, it is important to improve diagnosis tools so that people at high risk are identified early

22 ([Rasmussen and Langerman, 2019](#)).

23 *1.2. Review of past research*

24 Speech-based screening is among the techniques that have been widely
25 used for automated cognitive assessment ([Tóth et al., 2018](#)). Multiple
26 studies suggested studying semi-spontaneous and spontaneous speech by
27 selecting pathological phonetic, and lexico-semantic features, among others
28 ([Boschi et al., 2017](#)). In addition, the success of Machine Learning (ML)
29 techniques in the medical domain attracted many researchers to use ML
30 for dementia detection ([Tsang et al., 2020](#)). Combining speech analysis
31 technology with ML algorithms is an intrinsic opportunity to utilize speech
32 data for automatic screening of dementia, and finally translate speech-based
33 methods into clinical practice ([Chen et al., 2021](#)). [Meilán et al. \(2020\)](#)
34 studied more than 30 speech parameters related to mild cognitive impair-
35 ment (MCI) and other neurodegenerative processes. More than 30 variables
36 including duration, speech fluency and rhythm, fundamental frequency and
37 long-term average spectrum, intensity, and acoustic voice quality parameters
38 were explored. Statistical analysis showed that speech duration, and an
39 alteration in rhythm rate and intensity are the most significant parameters
40 to distinguish MCI diagnosed individuals having high probability to develop
41 dementia from those who won't develop it. [López-de-Ipiña et al. \(2015\)](#)
42 focused on analyzing spontaneous speech through extracting silence-related,
43 time-domain and frequency domain features from AZTIAHO dataset (a
44 multicultural and multilingual dataset they created for their study). They
45 also analyzed three families of features in speech (acoustic features like pitch,
46 voice quality features like shimmer and jitter, and duration features like the

47 degree of voice frames). In addition, they extracted emotional temperature
48 features — mainly prosodic and paralinguistic features. Afterwards, they
49 used these features to train several ML models. The best results were
50 obtained when combining spontaneous speech and emotional features along
51 with a multi-layer perceptron (MLP). The achieved accuracy was 92.24%
52 and 86.04% for their artificial neural network (ANN) and Support Vector
53 Machine (SVM) models, respectively. [Tóth et al. \(2018\)](#) developed a model
54 to distinguish between MCI and healthy patients. They extracted acoustic
55 parameters such as hesitation ratio, speech tempo, length and number of
56 silent and filled pauses (when the speaker hums, or produces other hesi-
57 tating sounds), length of utterance, from spontaneous speech in Hungarian
58 language. They showed that it is possible to separate the two-classes with
59 an accuracy of 75% and F1-score of 0.78. Likewise, [Qiao et al. \(2020\)](#)
60 focused on analyzing acoustic features from non-linguistic contents of the
61 silence/speech segments for healthy, MCI, and Alzheimer’s Disease (AD)
62 patients. They observed that all their parameters were significantly corre-
63 lated with cognitive performance, making it possible to detect pathologies
64 by analyzing the voice and detecting voice disorders. [Vizza et al. \(2019\)](#)
65 analyzed vocal signals and extracted relevant acoustic (F0, jitter, shimmer,
66 NHR) and vowel metric (tVSA, qVSA, FCR) parameters for neurological
67 disorder detection. Statistical tests reported significant differences in almost
68 all of the features in pathological (Parkinson and Multiple Sclerosis diseases)
69 and healthy voices. [Haulcy and Glass \(2021\)](#) used audio features (i-vectors
70 and x-vectors) and text features (word vectors, BERT embeddings, LIWC
71 features, and CLAN features) to automatically classify Alzheimer’s Diseases

72 and predict the Mini-Mental State Examination (MMSE) score for patients
73 from the ADReSS dataset. SVM and Random Forest (RF) classifiers
74 achieved 85.4% accuracy on the test set, making them the top-performing
75 classification models. It was concluded that it is feasible to use speech
76 analysis to classify AD and predict MMSE score. [Sadeghian et al. \(2017\)](#)
77 trained an MLP to classify patients with Alzheimer’s disease. They used 22
78 demographic, linguistic, and acoustic features including age, MMSE score,
79 race, number of pauses, total speech length, and others. Their approach
80 seemed promising, reaching 94% accuracy for diagnosing Alzheimer’s
81 disease. In a recent review study, [Vigo et al. \(2022\)](#) summarized the best
82 practices and most effective algorithms that were implemented between 2015
83 and 2020, as well as the most used datasets, including the dataset used in
84 this study ([Becker et al., 1994](#)). They focused on a wide variety of features
85 that are usually extracted from speech and used for dementia detection
86 and classification. They concluded that fundamental frequency, jitter, and
87 shimmer are among the characteristics that are able to differentiate between
88 healthy individuals and those with Alzheimer’s disease.

89
90 Although these studies have good results, they have potential draw-
91 backs. As explained above, the main disadvantage that can be marked is the
92 number of required variables to develop a good model. Hence, it is beneficial
93 to create a dementia detection model by relying on a small number of
94 variables. These variables should be easily extracted and be comprehensible
95 by health professionals.

96 *1.3. Objectives of the current study*

97 In this study, we analyze the speech of sick (dementia) and healthy
98 (control) individuals. We aim to assist doctors in monitoring patients
99 to aid in the early disclosure of perilous conditions for the development
100 of dementia. We are interested in developing a tool that can be lightly
101 embedded and used without the need for heavy computers. Therefore, we
102 present a study that requires a minimal number of features to create an
103 efficient and reliable tool for dementia detection. Another advantage of
104 the suggested system is that it can be embedded into an electronic device
105 (for example, a smartphone) and provide regular and frequent (e.g., daily)
106 classification results. This advantage is substantial because it helps health
107 professionals employ appropriate treatment protocols based on the given
108 prediction. We are also interested in developing a low-cost non-invasive
109 dementia detection method.

110

111 To the best of our knowledge, the system achieves state-of-the-art ac-
112 curacy for acoustic-based dementia classification when evaluated on the
113 **benchmark DementiaBank Pitt database**. The rest of the paper
114 is organized as follows. Section 2 presents the materials and methods,
115 including the original data, extracted features, and the proposed models.
116 Section 3 includes the statistical analysis and experimental results. Section
117 4 discusses the findings of this research. At the end, a general conclusion is
118 presented.

119 2. Materials and Methods

120 Figure 1 shows the workflow proposed in the current study. Module
121 1 corresponds to the data used in this study (Pitt Corpus ([Becker et al.,](#)
122 [1994](#))), downloaded from the [Dementia TalkBank database](#). In module 2,
123 we extract the acoustic features from the speech signals presented in module
124 1. Those features are subjected to analysis in order to find the significant
125 ones (in module 3). Finally, in module 4, the remaining variables are fed to
a machine learning algorithm to detect the class of the person.

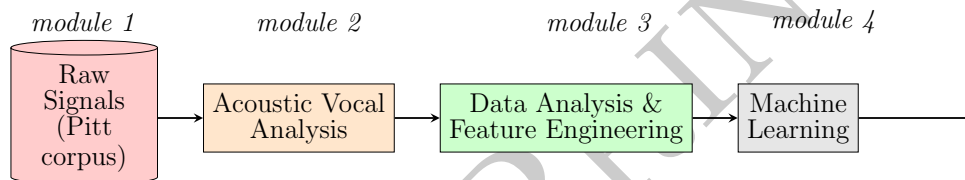


Figure 1: Flowchart of the proposed method

126

127 2.1. Data acquisition (module 1)

128 Data used in this study was downloaded from the Dementia Bank
129 Database ([Becker et al., 1994](#)). It corresponds to the English language
130 Pitt corpus that contains Dementia and control data for four language tasks
131 (cookie theft, verbal fluency, sentence construction, and story recall) from
132 a large longitudinal study. The audio files we used are the responses of
133 the Control group (242 samples) and Dementia group (309 samples) to the
134 Cookie Theft stimulus photo. These responses are of different durations.

135 2.2. Acoustic vocal analysis (module 2)

136 Several studies suggest that neurodegenerative disorders can be identified
137 by analyzing the acoustic parameters of the voice ([Meilán et al., 2020](#); [Tóth](#)

138 et al., 2018). The commonly used features in the literature for dementia
 139 detection are the mean and standard deviation of the fundamental frequency
 140 F_0 , harmonic-to-noise ratio (HNR), jitter, and shimmer (Teixeira et al., 2013;
 141 Farrús et al., 2007; Boersma, 1993).

142 Jitter (local, absolute, *jitta*): Represents the average absolute difference
 143 between two consecutive periods and is known as *jitta*.

$$jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (1)$$

144 Jitter (relative, *jitr*) : Represents the average absolute difference between
 145 two consecutive periods, divided by the average period.

$$jitr = \frac{jitta}{\frac{1}{N} \sum_{i=1}^N T_i} \times 100 \quad (2)$$

146 Jitter Relative Average Perturbation (*rap*): Represents ratio of disturbance
 147 within three periods, i.e, the average absolute difference between a period
 148 and the average of it and its two neighbours, divided by the average period.

$$rap = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - (T_{i-1} + T_i + T_{i+1})/3|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (3)$$

149 Jitter five-point Period Perturbation Quotient (*ppq5*): Represents the ratio of
 150 disturbance within five periods, i.e., the average absolute difference between
 151 a period and the average containing its four nearest neighbor periods, divided
 152 by average period.

$$ppq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |T_i - (T_{i-2} + T_{i-1} + T_i + T_{i+1} + T_{i+2})/5|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (4)$$

153 Jitter *ddp*: Average absolute difference between consecutive differences be-
 154 tween consecutive periods, divided by the average period. Its value is three
 155 times *rap*.

156 The shimmer measurements are very similar to those of the jitter, except
 157 that the period of the signal is replaced by the amplitude.

158 Shimmer (db): Is the average absolute base-10 logarithm of the difference
 159 between the amplitudes of consecutive periods, multiplied by 20.

$$shimdb = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \times \log \left(\frac{A_{i+1}}{A_i} \right) \right| \quad (5)$$

160 Shimmer (relative): Represents average absolute difference between the am-
 161plitudes of consecutive periods, divided by the average amplitude.

$$shimmr = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \times 100 \quad (6)$$

162 Shimmer three-point Amplitude Perturbation Quotient (*apq3*): The average
 163 absolute difference between the amplitude of a period and the average of the
 164 amplitudes of its neighbours, divided by the average amplitude.

$$apq3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - (A_{i-1} + A_i + A_{i+1})/3|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (7)$$

165 Shimmer five-point Amplitude Perturbation Quotient (*apq5*): The average
 166 absolute difference between the amplitude of a period and the average of
 167 the amplitudes of it and its four closest neighbours, divided by the average
 168 amplitude.

$$apq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |A_i - (A_{i-2} + A_{i-1} + A_i + A_{i+1} + A_{i+2})/5|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (8)$$

169 Shimmer eleven-point Amplitude Perturbation Quotient (*apq11*): Represents
 170 the ratio of disturbance within eleven periods.

$$apq11 = \frac{\frac{1}{N-1} \sum_{i=5}^{N-5} |A_i - (\frac{1}{11} \sum_{k=i-5}^{i+5} A_k)|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (9)$$

171 Shimmer (*dda*): Average absolute difference between consecutive differences
172 between the amplitudes of consecutive periods. Its value is three times *apq3*.

173 The Praat voice analysis software with the default parameters was used
174 to extract the acoustic features (Boersma and Weenik, 2022).

175 2.3. Data Analysis and Feature Engineering (module 3)

176 2.3.1. Data Analysis

177 Data analysis is performed through a Principal Component Analysis
178 (PCA) (Jolliffe, 2002), a dimension reduction technique often used to reduce
179 the size of a correlated data set without losing much information. It can be
180 also used to determine the correlation between two or more variables and
181 for feature engineering.

182

Let \mathcal{X} be an $n \times p$ matrix, where n denotes the number of samples and p denotes the number of features. We center and scale \mathcal{X} to avoid the loss of information that might be caused by disparate scales or units of variables. We then calculate the covariance matrix V associated to \mathcal{X} and consequently compute the eigenvalues $\lambda_1, \dots, \lambda_p$ and their corresponding eigenvectors u_1, \dots, u_p . The calculated eigenvectors are known as the principal components, where each *PC* is a linear combination of the original features. We usually select $d < p$ components to reduce the feature space dimension, most of the time by retaining the ones whose Inertia (I) covers a certain percentage of the explained variance such that:

$$I_j = \frac{\lambda_j}{\sum_{j=1}^p \lambda_j} \times 100.$$

We can also calculate the correlation between the variables and the principal components such that:

$$corr_j = \sqrt{\lambda_j} u_j.$$

183 This class-independent fashion of PCA is used for dimension reduction and
 184 feature extraction (Fukunaga, 1990). There exists a PCA-derived method
 185 called class-dependent PCA (c-PCA) or class-specific PCA, which focuses on
 186 conducting a PCA for each class separately (Sharma et al., 2006; Pan et al.,
 187 2020). c-PCA is utilized to find a linear transform for each class using the
 188 training patterns for that class in the feature space. This method is mainly
 189 used for classification problems and can be also used for feature extraction
 190 (Sharma et al., 2006).

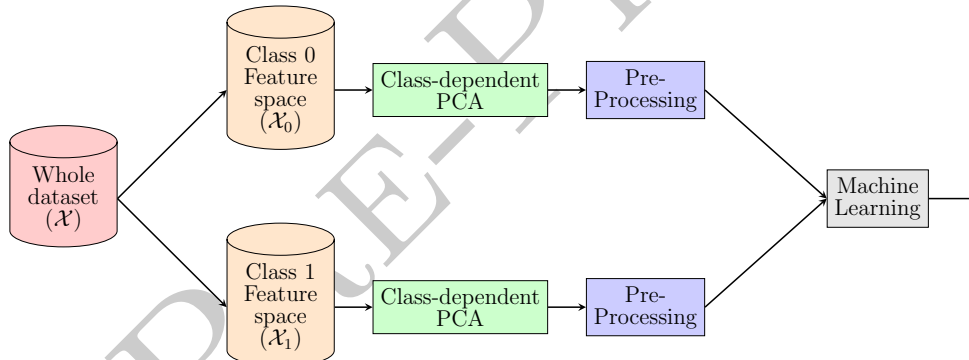


Figure 2: Framework of the proposed classifier (adapted from Sharma et al. (2006))

The c-PCA fashion consists of partitioning the initial dataset \mathcal{X} in a classification problem into C subsets, where C indicates the total number of classes in the data. In our case, since we have a binary classification problem, $c = 0, 1$. \mathcal{X} is divided into 2 subsets \mathcal{X}_0 and \mathcal{X}_1 because we have 2 classes (Control: 0, Dementia: 1) as shown in Figure 2. For each \mathcal{X}_c , we calculate

the covariance matrix V_c and its corresponding eigenvalues $\lambda_{1c}, \dots, \lambda_{pc}$ and eigenvectors u_{1c}, \dots, u_{pc} for $c = 0, 1$. We can then define the orthonormal transformation matrix Φ_c of dimension $p \times d$ where $d < p$ is the number of retained components. Φ_c contains the eigenvectors for each class c . To be able to generalize the class-dependent PCA for any new patterns x , i.e, assign the class loadings (green box) before pre-processing, we calculate the reconstruction error (distance) ξ between the original values and the reconstructed ones such that:

$$\xi_c = \|x - \hat{x}\|_2,$$

191 where $\hat{x} = \mu_c + \Phi_c \Phi_c^T (x - \mu_c)$ and $\mu_c = \frac{1}{n_c} \sum_{x \in \mathcal{X}_c} x$ for $c = 0, 1$ is the mean
 192 and n_c , is the total number of samples in subset \mathcal{X}_c , respectively. The
 193 new pattern is therefore assigned to the class with minimal ξ_c . This is an
 194 important step to choose the class loadings (the correlations between the
 195 original variables and the unit-scaled components).

196

197 Due to the size of the dataset used in this study, we used all the samples
 198 of each class for c-PCA. The hypothesis was that we wanted the principal
 199 components to capture as much diversity of data as possible because there
 200 is no training phase at this stage. Consequently, the reconstruction errors
 201 of the control group are minimal for class 0, and those of the dementia class
 202 are minimal for class 1. Once the most probable class is known, we can use
 203 its corresponding *loadings* for feature engineering.

204 2.3.2. Feature Engineering

205 The projections of the data used for c-PCA on the first two factorial
 206 axes (we retained the first 2 axes) were added to the features space in order

207 to create new meaningful variables. The features space now contains 16
 208 variables (the variables in Section 2.2 in addition to the projection of the jitter
 209 and shimmer variables on the factorial axes). For simplicity, we present the
 210 following demonstration. For instance, if M_0 is the matrix of control samples
 211 (242 samples), then its corresponding c-PCA is PCA_0 . Knowing that the
 212 reconstruction errors are smaller for class 0, then we use class 0 loadings
 213 (which are highly associated with the data in M_0) for feature engineering.
 214 The projections of the jitter and shimmer measurements on the first two
 215 axes, i.e, the principal components, XPC1 and XPC2 are then added to the
 original dataset, resulting in M'_0 .

$$M_0 = \begin{bmatrix} X_1 & X_2 & \dots & X_{14} & Y \\ x_{1,1} & x_{1,2} & \dots & x_{1,14} & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 \\ x_{242,1} & \dots & \dots & x_{242,14} & 0 \end{bmatrix} \xrightarrow{PCA_0} \begin{bmatrix} X_1 & X_2 & \dots & X_{14} & XPC1 & XPC2 & Y \\ x_{1,1} & x_{1,2} & \dots & x_{1,14} & x_{1,15} & x_{1,16} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ x_{242,1} & \dots & \dots & x_{242,14} & x_{242,15} & x_{242,16} & 0 \end{bmatrix} = M'_0$$

216

217 Similarly, the same procedure is adopted for M_1 (309 Dementia samples)

218 to get M'_1 .

$$M_1 = \begin{bmatrix} X_1 & X_2 & \dots & X_{14} & Y \\ x_{243,1} & x_{243,2} & \dots & x_{243,14} & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ x_{551,1} & \dots & \dots & x_{551,14} & 1 \end{bmatrix} \xrightarrow{PCA_1} \begin{bmatrix} X_1 & X_2 & \dots & X_{14} & XPC1 & XPC2 & Y \\ x_{243,1} & x_{243,2} & \dots & x_{243,14} & x_{243,15} & x_{243,16} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 1 \\ x_{551,1} & \dots & \dots & x_{551,14} & x_{551,15} & x_{551,16} & 1 \end{bmatrix} = M'_1$$

219 Finally, we concatenate M'_0 and M'_1 to get the matrix M ,

$$M = \begin{matrix}
& & X_1 & X_2 & \dots & X_{14} & XPC1 & XPC2 & Y \\
220 & & x_{1,1} & x_{1,2} & \dots & x_{1,14} & x_{1,15} & x_{1,16} & 0 \\
& & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\
& & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\
& & x_{242,1} & \dots & \dots & x_{242,14} & x_{242,15} & x_{242,16} & 0 \\
& & x_{243,1} & x_{243,2} & \dots & x_{243,14} & x_{243,15} & x_{243,16} & 1 \\
& & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 1 \\
& & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 1 \\
& & x_{551,1} & \dots & \dots & x_{551,14} & x_{551,15} & x_{551,16} & 1
\end{matrix}$$

221 Figure 3 shows a scheme of the feature engineering procedure described
222 above. For a given sample, we calculate the reconstruction error ζ per class.
223 If ζ is minimal for class 0, then we use PCA_0 loadings to calculate XPC1 and
224 XPC2 (M'_0). If not, we use those of class 1 (M'_1). Finally, we concatenate
225 the outputs (M) to obtain a dataset to train a machine learning model.

226 2.4. Learning the Classifier: Theoretical Background (module 4)

227 Three machine learning algorithms were studied in this paper. The idea
228 is to find the best classification model that fits the data. For example, Lo-
229 gistic regression is practical when data is linear and when we are willing to
230 predict a probability. Kernel-based support vector machines are accurate
231 and efficient when it comes to non-linear data. Artificial neural networks can
232 separate non-linear representations by learning complex relationships in the
233 data. This section will present the theoretical backbones of the classifiers
234 studied in this paper. Although the underlying mathematics are familiar to
235 many, it's always worth reminding readers.

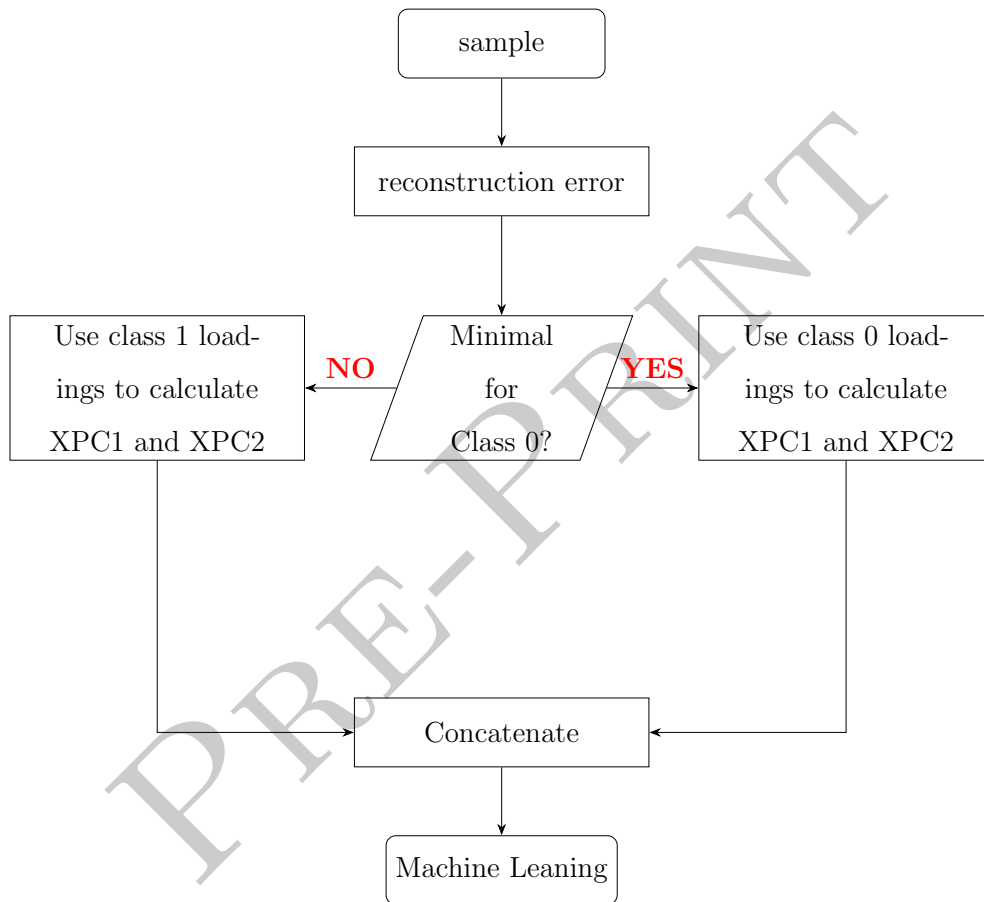


Figure 3: Feature Engineering procedure based on the reconstruction error

236 *2.4.1. Logistic Regression*

237 Logistic regression (LR) (Hastie et al., 2009) is a discriminative supervised
 238 machine learning algorithm that aims at modeling the posterior probabilities
 239 of two or more classes via linear functions. It uses the logistic function σ to
 240 obtain the predictive probabilities, such that:

$$\begin{aligned}\sigma_{\theta}(X) &= \frac{\exp(\theta^T X)}{1 + \exp(\theta^T X)} \\ &= \frac{\exp(\theta_0 + \theta_1 X_1 + \dots + \theta_p X_p)}{1 + \exp(\theta_0 + \theta_1 X_1 + \dots + \theta_p X_p)}\end{aligned}\quad (10)$$

241 where X_1, \dots, X_p are the p variables of the model and $\theta = (\theta_0, \theta_1, \dots, \theta_p)$ is
 242 parameters' vector (weights of the variables in which θ_0 is the bias or intercept
 243 term). For simplicity, we'll consider a binary classification problem. The
 244 probabilities of the default class ($Y=1$) and the other one ($Y=0$) are then
 245 written as:

$$\begin{aligned}P(Y = 1/X, \theta) &= \sigma_{\theta}(X) \\ P(Y = 0/X, \theta) &= 1 - \sigma_{\theta}(X)\end{aligned}$$

246 These two equations can be then combined in a single one:

$$P(Y/X, \theta) = (\sigma_{\theta}(X))^Y (1 - \sigma_{\theta}(X))^{(1-Y)} \quad (11)$$

247 To estimate the coefficients θ we first need to calculate the maximum likeli-
 248 hood function of equation 11:

$$L(\theta) = \prod_{i=1}^n (\sigma_{\theta}(X^{(i)})^{Y^{(i)}} (1 - \sigma_{\theta}(X^{(i)}))^{(1-Y^{(i)})} \quad (12)$$

249 and its logarithmic form:

$$\text{Log}(L(\theta)) = \sum_{i=1}^n Y^{(i)} \log \sigma_{\theta}(X^{(i)}) + (1 - Y^{(i)}) \log(1 - \sigma_{\theta}(X^{(i)})) \quad (13)$$

250 where n is the number of independent training samples. An optimization
 251 algorithm (gradient descent and its variants (Zhang, 2019), quasi-newton
 252 (Hennig and Kiefel, 2012), etc.) is then used to minimize equation 13 and
 253 accordingly return back the best values of the coefficients.

254 2.4.2. Kernel-Support Vector Machines

255 Support Vector Machine (SVM) (James et al., 2021) is an extension of
 256 the support vector classifier that was originally designed to perform linear
 257 classification. SVM is a margin-maximizing technique based on the idea
 258 of constructing a multidimensional hyperplane to discriminate between two
 259 classes. The linear solution of SVM can be written as:

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle \quad (14)$$

and the decision function is nothing but $\text{sign}(f(x))$. However, because data in
 real-world problems is not always linearly separable, a *kernel transformation*
 K is applied to handle such a case. This trick allows us to work in high
 dimensional vector spaces (Hilbert \mathcal{H}). If we consider a mapping function
 $h : \mathcal{X} \rightarrow \mathcal{H}$, then $\forall x, x' \in \mathcal{X}$, the inner product becomes:

$$K(x, x') = \langle h(x), h(x') \rangle_{\mathcal{H}}$$

260 where $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ returns the similarity of two inputs from \mathcal{X} in the
 261 feature space \mathcal{H} . The radial basis function (RBF) kernel is a popular choice
 262 for K such that:

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2)$$

263 where γ is a positive constant and x_i and $x_{i'}$ are two data points. The optimal
 264 solution is then obtained by maximizing the Lagrangian dual problem (L_D)
 265 such that:

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,i'=1}^n \alpha_i \alpha_{i'} y_i y_{i'} K(x_i, x_{i'}) \quad (15)$$

$$\text{subject to } \alpha_i \geq 0, i = 1, \dots, n, \quad \text{and } \sum_{i=1}^n \alpha_i y_i = 0. \quad (16)$$

266 Finally, the classification function becomes:

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i y_i K(x, x_i) \quad (17)$$

267 for $0 \leq \alpha_i \leq C$, where C is the regularization parameter that aims at achiev-
 268 ing a perfect margin separation. Once α_i is given, β_0 can be easily estimated
 269 for all x_i .

270 2.4.3. Artificial Neural Networks

271 ANNs are mathematical models used to approximate non-linear relation-
 272 ships between an input space and an output space. The multi-layer percep-
 273 tron (MLP) is the most common topology when the input data is numerical.
 274 An MLP usually consists of an input layer, a hidden layer (universal approx-
 275 imator), and an output layer (Negnevitsky, 2005) as in Figure 4. The output
 276 of the k^{th} hidden neuron can be represented by the following equation:

$$\begin{aligned} u_k &= \left(b_1^k + \sum_{j=1}^P v_{kj} X_j \right) \\ H_k &= \varphi_1(u_k) \end{aligned} \quad (18)$$

277 where P is the number of variables, b_1^k is the bias of the k^{th} neuron of the
 278 hidden layer, φ_1 is the activation function of the hidden layer, v_{kj} are the

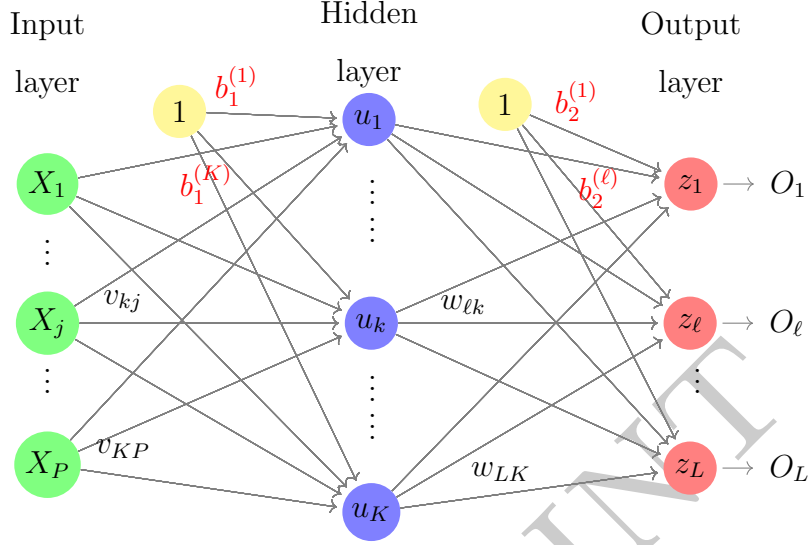


Figure 4: Typical architecture of a multi-layer network with a hidden layer

279 weights between the k^{th} hidden neuron and j^{th} input variable X_j . Likewise,
 280 the output of the ℓ^{th} neuron of the output layer is linear combination of the
 281 outputs and weights connecting all the neurons of the previous layer (hidden)
 282 with the actual neuron. It is expressed by the following equation:

$$\begin{aligned}
 z_\ell &= \left(b_2^\ell + \sum_{k=1}^K \omega_{\ell k} H_k \right) \\
 O_\ell &= \varphi_2(z_\ell)
 \end{aligned} \tag{19}$$

283 in which φ_2 stands for the activation function of the output layer, $\omega_{\ell k}$ is the
 284 weight between the ℓ^{th} output neuron and k^{th} , b_2^ℓ is the bias of the ℓ^{th} output
 285 neuron. In this study, we have chosen the rectified linear unit for φ_1 and the
 286 Sigmoid function for φ_2 . Accordingly,

$$\begin{aligned}
 \varphi_1(u_k) &= \text{ReLU}(u_k) = \max(0, u_k) \\
 \varphi_2(z_\ell) &= \text{Sigmoid}(z_\ell) = \frac{1}{1 + e^{-z_\ell}}
 \end{aligned}$$

We used *Adam* optimization algorithm (Zhang, 2019) and employed the binary cross-entropy loss \mathcal{L} to train the network,

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))]$$

287 where y_i is the label and $p(y_i)$ its predicted probability. We initialized the
288 weights following Glorot’s algorithm (Glorot and Bengio, 2010) to break sym-
289 metry during backpropagation.

290 2.5. Evaluation metrics

291 The confusion matrix (Pedregosa et al., 2011) is a common way to show
292 the prediction results obtained by a classifier. The elements of a confusion
293 matrix are true positives (TP), true negatives (TN), false positives (FP)
294 and false negatives (FN). To evaluate the quality of the developed methods,
295 we will compute the accuracy, recall, precision, and F1-score. We will also
296 use the specificity metric to compare the obtained results with those in the
297 literature.

- 298 • Accuracy is the most common metric used when the classification prob-
299 lem is balanced. It measures how many observations (both Dementia
300 and control) were correctly classified. In other words, it is the ratio of
301 correctly classified patients.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (20)$$

- 302 • Recall, or sensitivity, is the ratio of correctly predicting dementia with
303 respect to the sum of predicted true positive and false negative obser-
304 vations. This metric explains how many of the actual positive cases we

305 were able to predict correctly with our model.

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

306 • Precision explains how many of the correctly predicted cases actually
307 turned out to be positive. It's the ratio of true positives divided by the
308 number of predicted positives.

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

309 • F1-score is defined as the harmonic mean between precision and recall.
310 A high F1-score is associated with high precision and recall scores. This
311 metric is popular for imbalanced classification because it maintains a
312 balance between the precision and recall.

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (23)$$

313 • Specificity (True negative rate) is the ratio of true negatives with re-
314 spect to all negative outcomes. It represents the percentage of negative
315 samples that got the correct label.

$$Specificity = \frac{TN}{TN + FP} \quad (24)$$

316 3. Results on acoustic voice-based dementia classification

317 In this section, we present the results obtained after applying PCA and
318 class-specific PCA. We also show the classification results obtained by each
319 of the aforementioned models, with more attention to the neural network
320 model.

321 3.1. First method: classical PCA

322 Figure 5a shows that the first two principal components have the highest
323 eigenvalues, capturing around 94% of the explained variance (Figure 5b).
324 Hence, we chose these axes for feature engineering. The projection of the
325 data on the obtained axes were added to the features space as XPC1 and
XPC2.

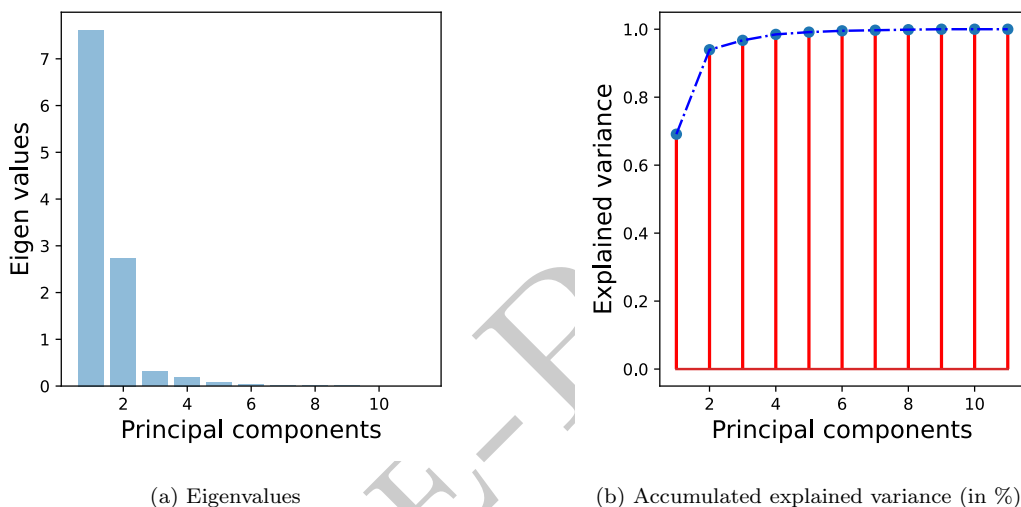


Figure 5: Eigenvalues and explained variance for each principal component

326

327 Using the trial and error method, we tested many architectures to find the
328 one that gives best results. The networks with more than one hidden layer
329 or too many neurons resulted in overfitting. Based on [Negnevitsky \(2005\)](#)
330 recommendation, the best architecture comprised one hidden layer with $2p+1$
331 neurons, following Kolmogorov's theorem ([Kolmogorov, 1957](#)), where p is the
332 number of input variables. At first, we considered all the input variables in
333 addition to XPC1 and XPC2, having a total of $p=16$ features. The model was
334 trained for 100 epochs without adding any early stopping criterion, because

Table 1: Metrics for the ANN model with 16 variables

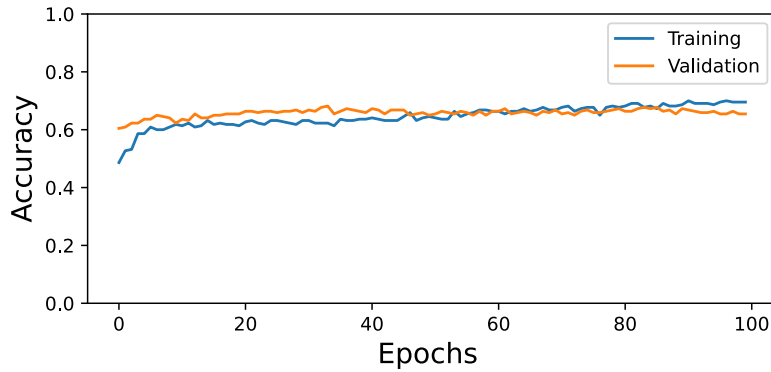
	Total	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-score
Training	440	197	122	71	50	0.725	0.73	0.797	0.76
Testing	111	42	27	22	20	0.62	0.65	0.68	0.7

335 we wanted to observe the evolution during training, and whether the current
 336 architecture will overfit or not. Figure 6 shows the history of the training
 337 and validation loss and accuracy. Figure 6a depicts that both accuracies were
 338 close to each other at each epoch, and their values seemed almost stable after
 339 epoch 15. However, we can observe that starting from epoch 90, the training
 340 accuracy started shifting up from the validation accuracy which could be a
 341 sign of overfitting. The loss curves shown in Figure 6b show that the training
 342 loss decreased gradually, whereas the validation one has decreased rapidly,
 343 and then stabilized.

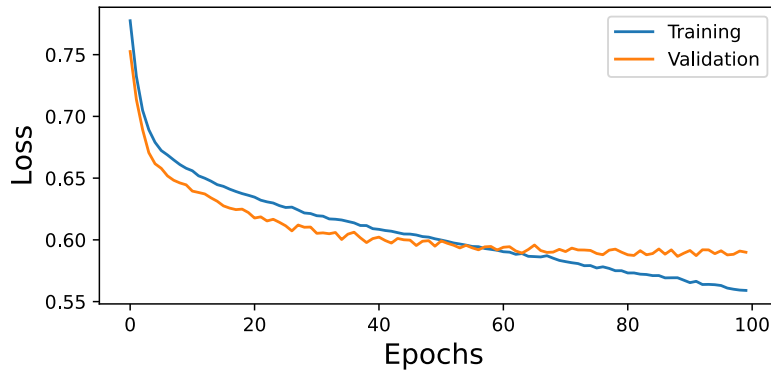
344 The classification results and metrics evaluation are presented in Table 1.
 345 The metrics show that the model works fine on the training data, with 79.7%
 346 of actual dementia samples being correctly classified (recall), and that among
 347 those classified as dementia, 73% are real dementia patients (precision). On
 348 the other side, the accuracy on the test data was 62%, compared to 72.5%
 349 on the training data.

350 Intuitively, XPC1 and XPC2 are correlated with the jitter and shimmer
 351 features. After removing the variables whose correlation ρ is greater than
 352 0.9, from the 16 initial variables we were left with 8 (MeanF0, StdF0, HNR,
 353 jittr, jitta, shimmr, XPC1, and XPC2) as shown in Figure 7.

354 The classification results and metric evaluation of the neural network



(a) Training and Validation Accuracy



(b) Training and Validation Loss

Figure 6: Loss and accuracy curves for the classical PCA-ANN model with 16 variables

355 model with the 8 variables above is shown in Table 2. Compared to Table 1,
 356 we notice that the model is less performant than the model with 16 variables,
 357 where the accuracy dropped from 72.5% to 66% on the training data, and
 358 from 62% to 54% on the test data.

359 Considering the low accuracy, the model is incapable of approximating
 360 the relationships between the descriptors and the classes. We also conclude
 361 that the existing features are insufficient for distinguishing between dementia

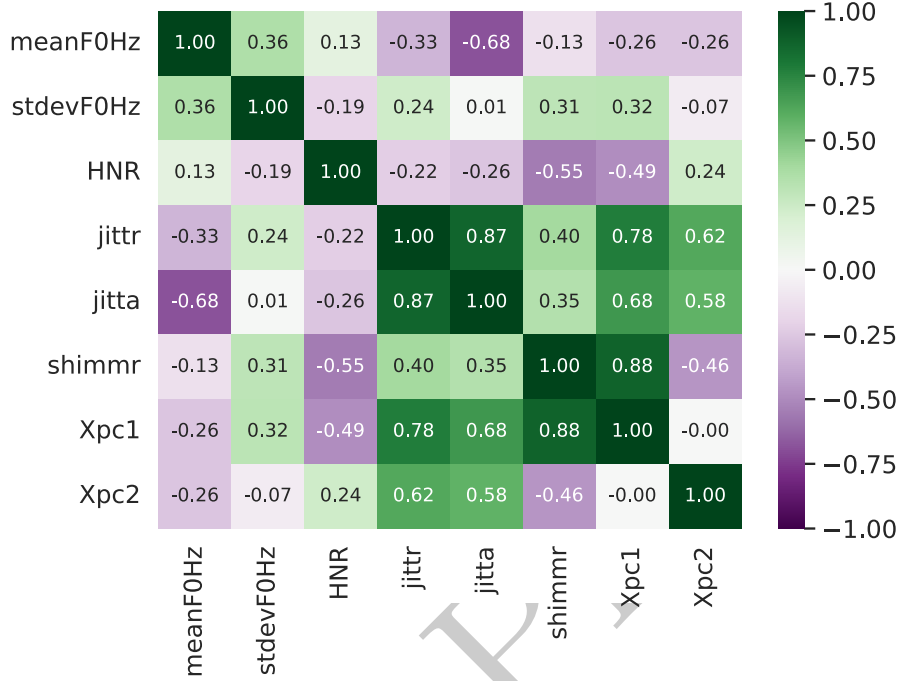


Figure 7: Linear correlation coefficients among variables

362 and healthy patients. Therefore, we propose a second method to overcome
 363 this issue.

364 3.2. Second method: class-dependent PCA

365 This approach consists of applying PCA on the data corresponding to
 366 each class separately, as shown below. This separation ensures that we are
 367 capturing class-specific variations independently. Based on the reconstruc-
 368 tion error criterion, we choose the *loadings* to compute the principal com-
 369 ponents of each class. Studying the linear correlation between the variables
 370 showed a high association among the shimmer variables and among the jitter
 371 variables, forming two blocks. For efficiency and less computation time, we

Table 2: Metrics for the ANN model with 8 variables

	Total	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-score
Training	440	181	111	82	66	0.66	0.69	0.73	0.71
Testing	111	36	24	25	26	0.54	0.59	0.58	0.59

372 eliminated the highly correlated variables $\rho > 0.9$. Hence, the number of
 373 features decreased from 16 to 8. We then studied four scenarios based on
 374 the reduced feature space (see Table 3). In the first scenario S1, we included
 375 all the variables, whereas we excluded XPC1 and XPC2 in S2 and S3, re-
 376 spectively. The last scenario, S4, did not include any of the added features.
 We scaled and standardized the data in order to avoid scaling invariant is-

Table 3: Variables in each of the proposed scenarios

Scenario	S1	S2	S3	S4
(1) MeanF0	✓	✓	✓	✓
(2) StdF0	✓	✓	✓	✓
(3) HNR	✓	✓	✓	✓
(4) jittr	✓	✓	✓	✓
(5) jitta	✓	✓	✓	✓
(6) shimmr	✓	✓	✓	✓
(7) XPC1	✓		✓	
(8) XPC2	✓	✓		

377
 378 sues that often lead to underfitting. Data was initially divided into 80% for
 379 training data and 20% for test data.

380 The best models were identified by optimizing the hyper-parameters using
 381 cross-validation. Knowing that the two binary classes are balanced, we'll
 382 only focus on the metrics of the best model, i.e, the one chosen based on the

383 accuracy metric. For each of the 4 scenarios, the logistic regression model
 384 performed worse than SVM and ANN (Table 4).

Table 4: Accuracy metric on the training and test data for each scenario

	S1		S2		S3		S4	
	Train	Test	Train	Test	Train	Test	Train	Test
LR	71%	65%	65%	56%	71%	67%	65%	57%
SVM	94%	91%	93%	90%	75%	63%	71%	63%
ANN	98%	97%	98%	98%	74%	63%	69%	63%

384

385 The accuracy scores indicate that the LR model was unable to discover
 386 complex patterns neither in the training nor in the test datasets, hence lead-
 387 ing to underfitting. This is due to the fact that this kind of algorithm con-
 388 structs linear boundaries, whereas our data are non-linearly separable. The
 389 SVM model shows a noticeably good performance, scoring an accuracy of
 390 94% and 93% on training data as well as 91% and 91% on testing data, for
 391 S1 and S2 respectively. As for the ANN, the results were outstanding. The
 392 accuracy score was 98% on training data for both scenarios, and 97% and
 393 98% on the test data, for S1 and S2 respectively. This demonstrates the
 394 superiority of S1 and S2 compared to S3 and S4, especially for SVM and
 395 ANN.

396 To examine the models' ability to generalize the results, we performed a
 397 10-fold cross-validation for SVM and ANN only (Table 5). The LR model
 398 was excluded because its training accuracy was not high enough.

399 The obtained accuracy scores show that most samples in the training
 400 and test datasets were correctly classified. Since this study aims at helping

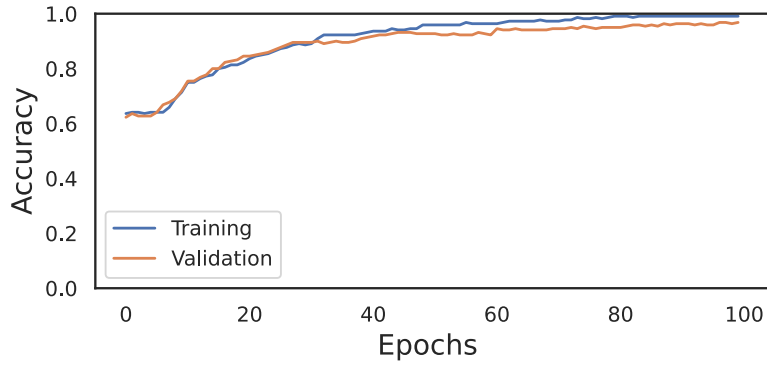
Table 5: 10 folds Cross-validation accuracy

	S1	S2	S3	S4
LR	–	–	–	–
SVM	94.7%	92.7%	77%	65%
ANN	97.2%	96%	76%	63%

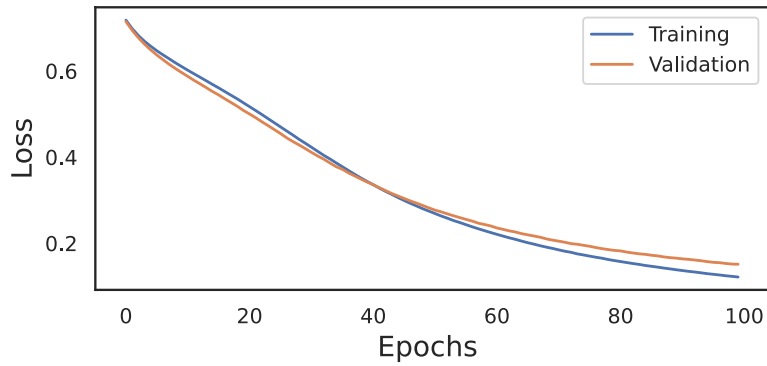
401 doctors to detect dementia effectively, we seek a reliable model that provides
 402 the highest accuracy. Therefore, in the following sections and paragraphs,
 403 we will focus on the model that showed the best performance, i.e, the neural
 404 network model.

405 The ANN architecture that showed the best results is composed of 1
 406 hidden layer with $2p+1=17$ neurons. The activation function used in the
 407 hidden layer is the ReLU whereas we used Sigmoid activation function for the
 408 output layer. A constant learning rate of 0.001 was chosen. We used Adam
 409 optimizer (Kingma and Ba, 2014) to minimize the binary cross-entropy loss.
 410 Empirically, a batch-size of 16 samples was best suited for our data.

411 Figure 8 shows the learning curves (accuracy and loss) on the training
 412 and validation data. We selected the model with the lowest validation loss
 413 to plot this figure. Figure 8a shows the evolution of the accuracy for the
 414 training and validation phase, in which the validation accuracy has been
 415 concordantly increasing with the training one, till the last epoch. As the
 416 loss curves (Figure 8b) decrease smoothly and continuously, we deduce that
 417 the ANN model is optimizing in the right direction. In addition, this ANN
 418 was trained for 100 epochs, in conformity with previous models. Figure 9
 419 shows the confusion matrices (classification results) of the neural network on



(a) Training and Validation Accuracy



(b) Training and Validation Loss

Figure 8: Learning curves of ANN on the training and validation data

420 the training and test data. We observe that the model misclassified only 7
 421 individuals out of 440 (1.59%) in the training data and only 3 out of 111
 422 (2.7%) in the test data.

423 The figure also allows us extract the TP, TN, FP, and FN values (Table 6).
 424 The precision metric reveals that among the predicted positives, 98.3% of
 425 them are truly positive in the training data and 98.7% in the test data. As
 426 for the recall, we conclude that the model correctly predicted 98.3% of the

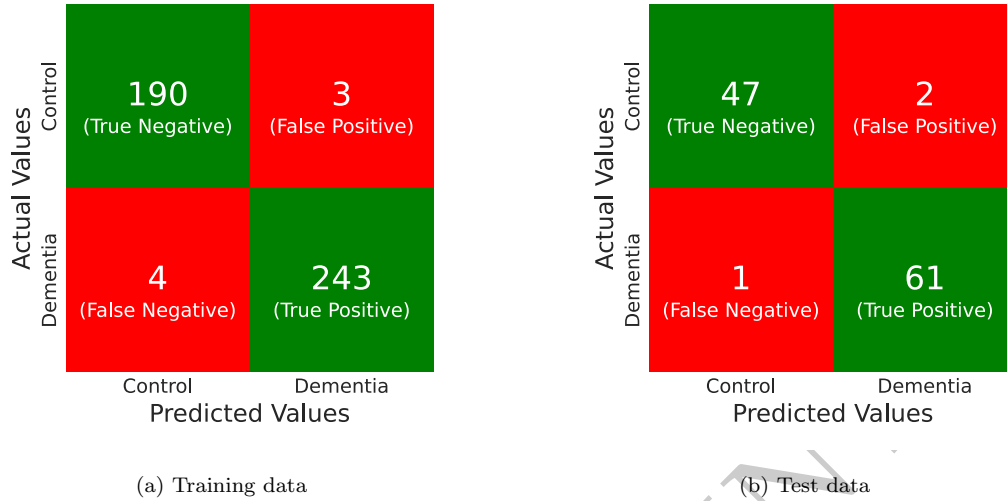


Figure 9: Confusion matrix results of the ANN model

Table 6: Metrics for the ANN model

	Total	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-score
Training	440	242	190	3	4	0.984	0.987	0.983	0.984
Testing	111	61	47	2	1	0.972	0.983	0.968	0.975

427 actual positive cases in the training set and 96.8% in the testing set. The
 428 F1-score is another indicator of the model’s ability to distinguish all positive
 429 cases and be accurate with the captured cases at 97.5% for test data. Finally,
 430 the specificity is equal to 98.4% on the training data and 95.9% on the test
 431 data.

432 All the obtained results show a good performance of the neural network
 433 model, confirming its efficiency for dementia detection.

434 4. Discussion

435 This study contributes to the early detection of dementia, making possible
436 a reliable classification of patients. The developed model relies on few
437 data automatically extracted from the vocal signal. Hence, it is easy-to-use
438 by health experts. In a recent study, [Javeed et al. \(2023\)](#) published a systematic
439 review on existing machine learning models for dementia detection. In
440 their article, the authors listed 61 existing datasets covering three modalities:
441 clinical-variables, images, and vocal signals. Among the image-based detection
442 models, the neural network proposed by [Akhila et al. \(2017\)](#) has been the
443 best performing with an accuracy, specificity, and sensitivity (recall) of 97.5%
444 each, when tested on the OASIS-image dataset ([Marcus et al., 2007](#)). There
445 are also other models that achieved 100% of accuracy, but the recall and
446 specificity metrics are either missing or inferior. Among the clinical-variable
447 based models, [Bansal et al. \(2018\)](#)'s decision tree (J48) scored 98.6% accuracy
448 on OASIS cross sectional and longitudinal data. However, the sensitivity and
449 specificity metrics were not as high as the accuracy. Concerning the voice-
450 based models, [Javeed et al. \(2023\)](#) showed that the best performing model
451 is that of [Syed et al. \(2021\)](#). It is based on multi-modal identification of
452 linguistic and paralinguistic traits of dementia using an automated screening
453 tool. By using bag-of-deep-features for feature selection, the authors built
454 an ensemble model for classification on the [ADReSS dataset](#). Their best results
455 were obtained when considering text data only (transcription of audio
456 signals) with accuracy=95.3%, recall=96.3% and specificity=94.4%. However,
457 their best audio based model was not as efficient with accuracy=86.1%,
458 recall=87%, and specificity=85.2%. Among all the studies presented in the

459 comparative study, only two of them used the Dementia Bank dataset. Ori-
 460 maye et al. (2014) extracted syntactic (number of predicates, reduced sen-
 461 tences, etc) and lexical features (word count, utterances, morphemes, and
 462 many others) from the Dementia bank (Becker et al., 1994) samples to build
 463 diagnostic models. The best model that distinguished dement patients from
 464 healthy patients was an SVM with accuracy=74% and recall=74%.

465 Table 7 demonstrates a comparison between the results we obtained in
 466 this study and those existing in the literature. For convenience, we compared
 467 with studies that used the **Dementia bank dataset only**. Comparison with
 468 other methods can be found in (Javeed et al., 2023), but were not listed here
 469 because they are not directly comparable. The system achieves an accuracy
 470 of 97.2%, surpassing the state-of-the-art in acoustic dementia detection.

Table 7: Performance evaluation of voice-modality based ML models for dementia, using the Dementia Bank dataset

Authors	Feature selection	Model	Accuracy (%)	Recall (%)	Specificity(%)
Orimaye et al. (2014)	Information Gain	SVM	74	74	75
Tóth et al. (2018)	ASR	RF	75	81.3	66.7
Santander-Cruz et al. (2022)	SBERT	SVM	77	80	80
Sarawgi et al. (2020)	–	ANN	88	82	
López-de-Ipiña et al. (2015)	–	ANN	93	NA	NA
Tay et al. (Ours)	c-PCA+Correlation	ANN	97.2	96.8	95.9

471 Regarding the c-PCA stage, we recall that because the objective is ba-
 472 sically to perform feature engineering based on the reconstruction error, we
 473 preferred to use all the samples associated to each class to capture as much
 474 diversity of information as possible. Thus, for any new observation (sample),
 475 we need to compare the reconstruction error based on the transformation

476 matrix Φ_c , and then decide which class loadings to choose in order to cal-
477 culate the new feature XPC1 and XPC2. Once the choice is done, the data
478 vector is then introduced to the model to predict whether the person is likely
479 to have dementia or not.

480 Finally, it is worth mentioning that for most samples, the audio quality
481 was poor. Despite this, the approach we studied was able to distinguish
482 between control and dementia subjects. As part of a further contribution,
483 we would like to improve signal quality by using appropriate transformations
484 and applying adequate filtering algorithms.

485 5. Conclusion

486 In this study, we proposed a machine learning approach for dementia
487 disease classification. The proposed model classifies patients as healthy or
488 sick after analyzing their speech and extracting pertinent features related to
489 fundamental frequency, harmonic to noise ratio, jitter, and shimmer. We
490 compared three models: logistic regression, radial basis function (rbf) kernel
491 support vector machines, and artificial neural networks. The results demon-
492 strated the superiority of ANN (accuracy=0.972), confirming it to be a reli-
493 able model for dementia detection. Two of the main contributions this study
494 brings are the development of a computationally low-cost and methodical
495 model that relies on a small number of features, that could be employed for
496 regular and frequent diagnosis, helping keep track of the mental health of pa-
497 tients with suspicion of dementia. The results also revealed the importance
498 of the proposed methodology in avoiding overfitting and obtaining excellent
499 classification results. This is due to the class-dependent PCA step which

500 captures as much information as possible from the data before engineering
501 new features. Our method achieves state-of-the-art test accuracy, precision,
502 recall, and F1-score for dementia classification on the DementiaBank Pitt
503 database. Our future work will focus on testing the proposed method on
504 other audio datasets (for example, [Address dataset](#)) and for other diseases.
505 We also aim to extend this work by adding other features related to silence,
506 speech rate, pauses, etc., and then proposing a method for multi-modal de-
507 mentia detection.

508 **Data availability**

509 The data analyzed in this study is subject to the following li-
510 censes/restrictions: in order to gain access to the datasets used in the paper,
511 researchers must become a member of DementiaBank. Requests to access
512 these datasets should be directed to <https://dementia.talkbank.org/>.

513 **Declaration Statement - Conflict of Interest**

514 Ahmad TAY's work was co-funded by the French Agence Nationale de
515 la Recherche (ANR) and the Smart Macadam company through the Plan
516 "France Relance". Smart Macadam was not involved in the study design,
517 collection, analysis, interpretation of data, the writing of this article or the
518 decision to submit it for publication. The remaining authors declare that
519 the research was conducted in the absence of any commercial or financial
520 relationships that could be construed as a potential conflict of interest.

521 **Authors' contributions**

522 Literature review by AT, MA, CL. Methodology and theoretical devel-
523 opments by AT. Experiment design and implementation by AT. Analysis of
524 the experimental results by AT. Document writing and illustrations by AT,
525 MA, CL. All authors contributed to manuscript revision, read, and approved
526 the submitted version.

527 **Acknowledgments**

528 This research was led within the context of the Mementop project, co-
529 funded by the French Agence Nationale de la Recherche (ANR) through the
530 Plan "France Relance" and the Smart Macadam company.

531 We also thank the creators of the Dementia Bank database (Pitt corpus)
532 [Becker et al. \(1994\)](#) and the grant support for the Pitt corpus – NIA AG03705
533 and AG05133.

534 **License**

535 In accordance with our funding institution's rules regarding open access
536 to results of publicly funded scientific research, the current and all subsequent
537 versions of this article will be published under [CC-BY 4.0 license](#).

538 **References**

539 Akhila, J., Markose, C., Aneesh, R.P., 2017. Feature extraction and clas-
540 sification of dementia with neural network. 2017 International Confer-
541 ence on Intelligent Computing, Instrumentation and Control Technolo-

- 542 gies (ICICICT) , 1446–1450URL: [https://api.semanticscholar.org/](https://api.semanticscholar.org/CorpusID:5038545)
543 [CorpusID:5038545](https://api.semanticscholar.org/CorpusID:5038545).
- 544 Bansal, D., Chhikara, R., Khanna, K., Gupta, P., 2018. Comparative anal-
545 ysis of various machine learning algorithms for detecting dementia. Pro-
546 ceedia Computer Science 132, 1497–1502. doi:[https://doi.org/10.1016/](https://doi.org/10.1016/j.procs.2018.05.102)
547 [j.procs.2018.05.102](https://doi.org/10.1016/j.procs.2018.05.102). international Conference on Computational Intel-
548 ligence and Data Science.
- 549 Becker, J.T., Boiler, F., Lopez, O.L., Saxton, J., McGonigle, K.L., 1994.
550 The natural history of alzheimer’s disease: description of study cohort and
551 accuracy of diagnosis. Archives of neurology 51, 585–594.
- 552 Boersma, P., 1993. Accurate short-term analysis of the fundamental fre-
553 quency and the harmonics-to-noise ratio of a sampled sound, in: IFA Pro-
554 ceedings, pp. 97–110.
- 555 Boersma, P., Weenik, D., 2022. Praat: doing phonetics by computer [com-
556 puter program]. <https://www.fon.hum.uva.nl/praat/>.
- 557 Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., Cappa, S.F.,
558 2017. Connected speech in neurodegenerative language disorders: A re-
559 view. Frontiers in Psychology 8. doi:[10.3389/fpsyg.2017.00269](https://doi.org/10.3389/fpsyg.2017.00269).
- 560 Chen, J., Ye, J., Tang, F., Zhou, J., 2021. Automatic detection of alzheimer’s
561 disease using spontaneous speech only, in: Proc. Interspeech 2021. doi:[10.](https://doi.org/10.21437/Interspeech.2021-2002)
562 [21437/Interspeech.2021-2002](https://doi.org/10.21437/Interspeech.2021-2002).
- 563 Chen, T., Su, P., Shen, Y., Chen, L., Mahmud, M., Zhao, Y., Antoniou, G.,
564 2022. A dominant set-informed interpretable fuzzy system for automated

565 diagnosis of dementia. *Frontiers in Neuroscience* doi:[10.3389/fnins.2022.](https://doi.org/10.3389/fnins.2022.867664)
566 [867664](https://doi.org/10.3389/fnins.2022.867664).

567 Farrús, M., Hernando, J., Ejarque, P., 2007. Jitter and shimmer measure-
568 ments for speaker recognition, in: *Proc. Interspeech 2007*, pp. 778–781.
569 doi:[10.21437/Interspeech.2007-147](https://doi.org/10.21437/Interspeech.2007-147).

570 Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*. Aca-
571 demic Press.

572 Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep
573 feedforward neural networks. *Journal of Machine Learning Research* 9,
574 249–256.

575 Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statis-*
576 *tical Learning: Data Mining, Inference, and Prediction*. Springer se-
577 ries in statistics, Springer. URL: [https://books.google.fr/books?id=](https://books.google.fr/books?id=eBSgoAEACAAJ)
578 [eBSgoAEACAAJ](https://books.google.fr/books?id=eBSgoAEACAAJ).

579 Haulcy, R., Glass, J., 2021. Classifying alzheimer’s disease using audio and
580 text-based representations of speech. *Frontiers in Psychology* 11. doi:[10.](https://doi.org/10.3389/fpsyg.2020.624137)
581 [3389/fpsyg.2020.624137](https://doi.org/10.3389/fpsyg.2020.624137).

582 Hennig, P., Kiefel, M., 2012. Quasi-newton methods: A new direction. *J.*
583 *Mach. Learn. Res.* 14, 843–865.

584 James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. *An Introduc-*
585 *tion to Statistical Learning: with Applications in R*. Springer Texts
586 in Statistics, Springer US. URL: [https://books.google.fr/books?id=](https://books.google.fr/books?id=5dQ6EAAAQBAJ)
587 [5dQ6EAAAQBAJ](https://books.google.fr/books?id=5dQ6EAAAQBAJ).

- 588 Javeed, A., Dallora, A.L., Berglund, J.S., Ali, A., Ali, L., Anderberg, P.,
589 2023. Machine learning for dementia prediction: A systematic review and
590 future research directions. *Journal of Medical Systems* 47. doi:[10.1007/
591 s10916-023-01906-7](https://doi.org/10.1007/s10916-023-01906-7).
- 592 Jolliffe, I.T., 2002. *Principal component analysis for special types of data*.
593 Springer.
- 594 Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimiza-
595 tion. CoRR abs/1412.6980. URL: [https://api.semanticscholar.org/
596 CorpusID:6628106](https://api.semanticscholar.org/CorpusID:6628106).
- 597 Kolmogorov, A., 1957. On the representation of continuous functions of
598 several variables by superposition of continuous functions of one variable
599 and addition. *Doklady Akademii. Nauk USSR* 114, 679–681.
- 600 López-de-Ipiña, K., Alonso, J.B., Solé-Casals, J., Barroso, N., Henriquez, P.,
601 Faundez-Zanuy, M., Travieso, C.M., Ecay-Torres, M., Martínez-Lage, P.,
602 Eguiraun, H., 2015. On automatic diagnosis of alzheimer’s disease based
603 on spontaneous speech analysis and emotional temperature. *Cognitive
604 Computation* 7, 44–55. doi:[10.1007/s12559-013-9229-9](https://doi.org/10.1007/s12559-013-9229-9).
- 605 Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C.,
606 Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS):
607 Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and
608 Demented Older Adults. *Journal of Cognitive Neuroscience* 19, 1498–
609 1507. URL: <https://doi.org/10.1162/jocn.2007.19.9.1498>, doi:[10.
610 1162/jocn.2007.19.9.1498](https://doi.org/10.1162/jocn.2007.19.9.1498).

- 611 Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E.,
612 Carro, J., 2020. Changes in the rhythm of speech difference between peo-
613 ple with nondegenerative mild cognitive impairment and with preclini-
614 cal dementia. *Behavioural Neurology*, Hindawi 2020. doi:[10.1155/2020/
615 4683573](https://doi.org/10.1155/2020/4683573).
- 616 Negnevitsky, M., 2005. *Artificial intelligence: a guide to intelligent systems*.
617 2 ed., Addison-Wesley, New York.
- 618 Orimaye, S.O., Wong, J.S.M., Golden, K.J., 2014. Learning predictive lin-
619 guistic features for Alzheimer’s disease and related dementias using ver-
620 bal utterances, in: Resnik, P., Resnik, R., Mitchell, M. (Eds.), *Proceed-*
621 *ings of the Workshop on Computational Linguistics and Clinical Psychol-*
622 *ogy: From Linguistic Signal to Clinical Reality*, Association for Com-
623 *putational Linguistics*, Baltimore, Maryland, USA. pp. 78–87. URL:
624 <https://aclanthology.org/W14-3210>, doi:[10.3115/v1/W14-3210](https://doi.org/10.3115/v1/W14-3210).
- 625 Pan, F., Zhang, Z.X., Liu, B.D., Xie, J.J., 2020. Class-specific sparse prin-
626 cipal component analysis for visual classification. *IEEE Access* 8, 110033–
627 110047. doi:[10.1109/ACCESS.2020.3001202](https://doi.org/10.1109/ACCESS.2020.3001202).
- 628 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel,
629 O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.,
630 Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011.
631 *Scikit-learn: Machine learning in Python*. *Journal of Machine Learning*
632 *Research* 12, 2825–2830.
- 633 Qiao, Y., Xie, X.Y., Lin, G.Z., Y.Zou, Chen, S.D., Ren, R.J., Wang, G.,

- 634 2020. Computer-assisted speech analysis in mild cognitive impairment
635 and alzheimer's disease: A pilot study from shanghai, china. *Journal of*
636 *Alzheimer's Disease* 75, 211–221. doi:[10.3233/JAD-191056](https://doi.org/10.3233/JAD-191056).
- 637 Rasmussen, J., Langerman, H., 2019. Alzheimer's disease – why we need
638 early diagnosis. *Degenerative Neurological and Neuromuscular Disease* 24,
639 123–130. doi:[10.2147/DNND.S228939](https://doi.org/10.2147/DNND.S228939).
- 640 Roeck, E.E.D., Deyn, P.P.D., Dierckx, E., Engelborghs, S., 2019. Brief cog-
641 nitive screening instruments for early detection of alzheimer's disease: a
642 systematic review. *Alzheimer's Research & Therapy* 11. doi:[10.1186/
643 s13195-019-0474-3](https://doi.org/10.1186/s13195-019-0474-3).
- 644 Sadeghian, R., Schaffer, J.D., Zahorian, S.A., 2017. Speech processing ap-
645 proach for diagnosing dementia in an early stage, in: *Proc. Interspeech*
646 2017, pp. 2705–2709. doi:[10.21437/Interspeech.2017-1712](https://doi.org/10.21437/Interspeech.2017-1712).
- 647 Santander-Cruz, Y., Salazar-Colores, S., Paredes-García, W.J., Guendulain-
648 Arenas, H., Tovar-Arriaga, S., 2022. Semantic feature extraction us-
649 ing sbert for dementia detection. *Brain Sciences* 12. doi:[10.3390/
650 brainsci12020270](https://doi.org/10.3390/brainsci12020270).
- 651 Sarawgi, U., Zulfikar, W., Soliman, N., Maes, P., 2020. Multimodal inductive
652 transfer learning for detection of alzheimer's dementia and its severity.
653 [arXiv:2009.00700](https://arxiv.org/abs/2009.00700).
- 654 Sharma, A., Paliwal, K.K., Onwubolu, G.C., 2006. Class-dependent pca,
655 mdc and lda: A combined classifier for pattern classification. *Pat-
656 tern Recognition* 39, 1215–1229. URL: [https://www.sciencedirect.](https://www.sciencedirect.com)

- 657 [com/science/article/pii/S0031320306000410](https://doi.org/10.1016/j.patcog.2006.02.001), doi:[https://doi.org/](https://doi.org/10.1016/j.patcog.2006.02.001)
658 [10.1016/j.patcog.2006.02.001](https://doi.org/10.1016/j.patcog.2006.02.001).
- 659 Syed, Z.S., Syed, M.S.S., Lech, M., Pirogova, E., 2021. Automated recogni-
660 tion of alzheimer's dementia using bag-of-deep-features and model ensem-
661 bling. *IEEE Access* 9, 88377–88390. doi:[10.1109/ACCESS.2021.3090321](https://doi.org/10.1109/ACCESS.2021.3090321).
- 662 Teixeira, J.P., Oliveira, C., Lopes, C., 2013. Vocal acoustic analysis - jitter,
663 shimmer and hnr parameters. *Procedia Technology* 9, 1112–1122. doi:[10.](https://doi.org/10.1016/j.protcy.2013.12.124)
664 [1016/j.protcy.2013.12.124](https://doi.org/10.1016/j.protcy.2013.12.124).
- 665 Tsang, G., Xie, X., Zhou, S.M., 2020. Harnessing the power of machine learn-
666 ing in dementia informatics research: Issues, opportunities, and challenges.
667 *IEEE Reviews in Biomedical Engineering* 13.
- 668 Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti,
669 Z., Pákási, M., Kálmán, J., 2018. A speech recognition-based solution
670 for the automatic detection of mildcognitive impairment from sponta-
671 neous speech. *Current Alzheimer Research* 15, 130–138. doi:[10.2174/](https://doi.org/10.2174/1567205014666171121114930)
672 [1567205014666171121114930](https://doi.org/10.2174/1567205014666171121114930).
- 673 Vigo, I., Coelho, L., Reis, S., 2022. Speech- and language-based classifica-
674 tion of alzheimer'sdisease: A systematic review. *Bioengineering*, MDPI 9.
675 doi:[10.3390/bioengineering9010027](https://doi.org/10.3390/bioengineering9010027).
- 676 Vizza, P., Tradigo, G., Mirarchi, D., Bossio, R.B., Lombardo, N., Arabia,
677 G., Quattrone, A., Veltri, P., 2019. Methodologies of speech analysis for
678 neurodegenerative diseases evaluation. *International Journal of Medical*
679 *Informatics* , 45–54doi:[10.1016/j.ijmedinf.2018.11.008](https://doi.org/10.1016/j.ijmedinf.2018.11.008).

680 WHO, 2019. Dementia. URL: [https://www.who.int/news-room/](https://www.who.int/news-room/fact-sheets/detail/dementia)
681 [fact-sheets/detail/dementia](https://www.who.int/news-room/fact-sheets/detail/dementia).

682 Yadav, V.G., 2019. The hunt for a cure for alzheimer's disease re-
683 ceives a timely boost. Science Translational Medicine 11. doi:[10.1126/](https://doi.org/10.1126/scitranslmed.aaz0311)
684 [scitranslmed.aaz0311](https://doi.org/10.1126/scitranslmed.aaz0311).

685 Zhang, J., 2019. Gradient descent based optimization algorithms for deep
686 learning models training. IFM LAB Tutorial series .

PRE-PRINT