Dementia detection based on speech acoustics using machine learning

Ahmad Tay, Mihai Andries, Christophe Lohr

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

(Rasmussen and Langerman, 2019).

1.2. Review of past research

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

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

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

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

1.3. Objectives of the current study

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

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

2. Materials and Methods

Figure 1 shows the workflow proposed in the current study. Module 1 corresponds to the data used in this study (Pitt Corpus (Becker et al., 1994)), downloaded from the Dementia TalkBank database. In module 2, we extract the acoustic features from the speech signals presented in module 1. Those features are subjected to analysis in order to find the significant 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.

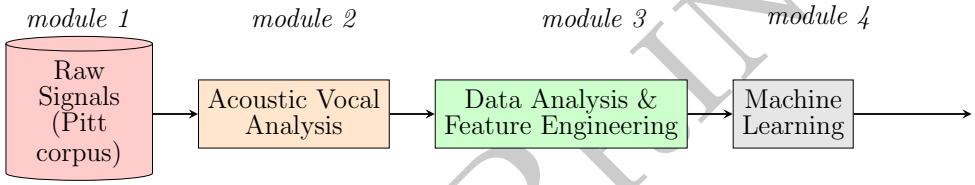


Figure 1: Flowchart of the proposed method

2.1. Data acquisition (module 1)

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

2.2. Acoustic vocal analysis (module 2)

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

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

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

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

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

$$jittr = \frac{jitta}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{2}$$

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

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

Jitter five-point Period Perturbation Quotient (*ppq*5): Represents the ratio of disturbance within five periods, i.e., the average absolute difference between a period and the average containing its four nearest neighbor periods, divided by average period.

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

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

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

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

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

Shimmer (relative): Represents average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

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

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

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

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

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

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

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

Shimmer (*dda*): Average absolute difference between consecutive differences between the amplitudes of consecutive periods. Its value is three times *apq*3.

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

2.3. Data Analysis and Feature Engineering (module 3)

2.3.1. Data Analysis

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

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 $\lambda_1, \ldots, \lambda_p$ and their corresponding eigenvectors $u_1, \ldots, u_p$. 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:

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

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

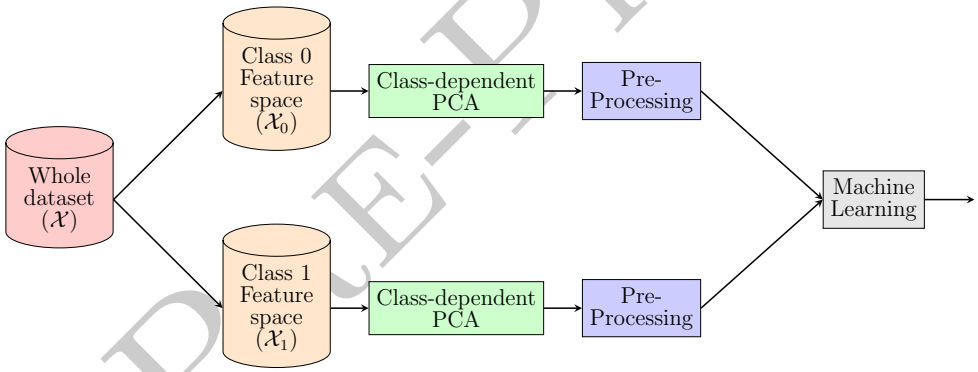


Figure 2: Framework of the proposed classifier (adapted from Sharma et al. (2006))

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 $\lambda_{1c}, \ldots, \lambda_{pc}$ and eigenvectors $u_{1c}, \ldots, u_{pc}$ 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:

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

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 and n_c, is the total number of samples in subset $\mathcal{X}_c$, respectively. The new pattern is therefore assigned to the class with minimal ξ_c. This is an important step to choose the class loadings (the correlations between the original variables and the unit-scaled components).

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

2.3.2. Feature Engineering

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

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

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

Similarly, the same procedure is adopted for M_1 (309 Dementia samples) to get M_1'.

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

Finally, we concatenate M_0' and M_1' to get the matrix M,

$$
M = \begin{array}{c}
\begin{array}{cccccccc} X_1 & X_2 & \dots & X_{14} & XPC1 & XPC2 & Y \end{array} \\
\left[\begin{array}{cccccccc}
x_{1,1} & x_{1,2} & \dots & x_{1,14} & x_{1,15} & x_{1,16} & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\
x_{242,1} & \dots & \dots & x_{242,14} & x_{242,15} & x_{242,16} & 0 \\
x_{243,1} & x_{243,2} & \dots & x_{243,14} & x_{243,15} & x_{243,16} & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 1 \\
x_{551,1} & \dots & \dots & x_{551,14} & x_{551,15} & x_{551,16} & 1
\end{array}\right]
\end{array}
$$

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

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

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

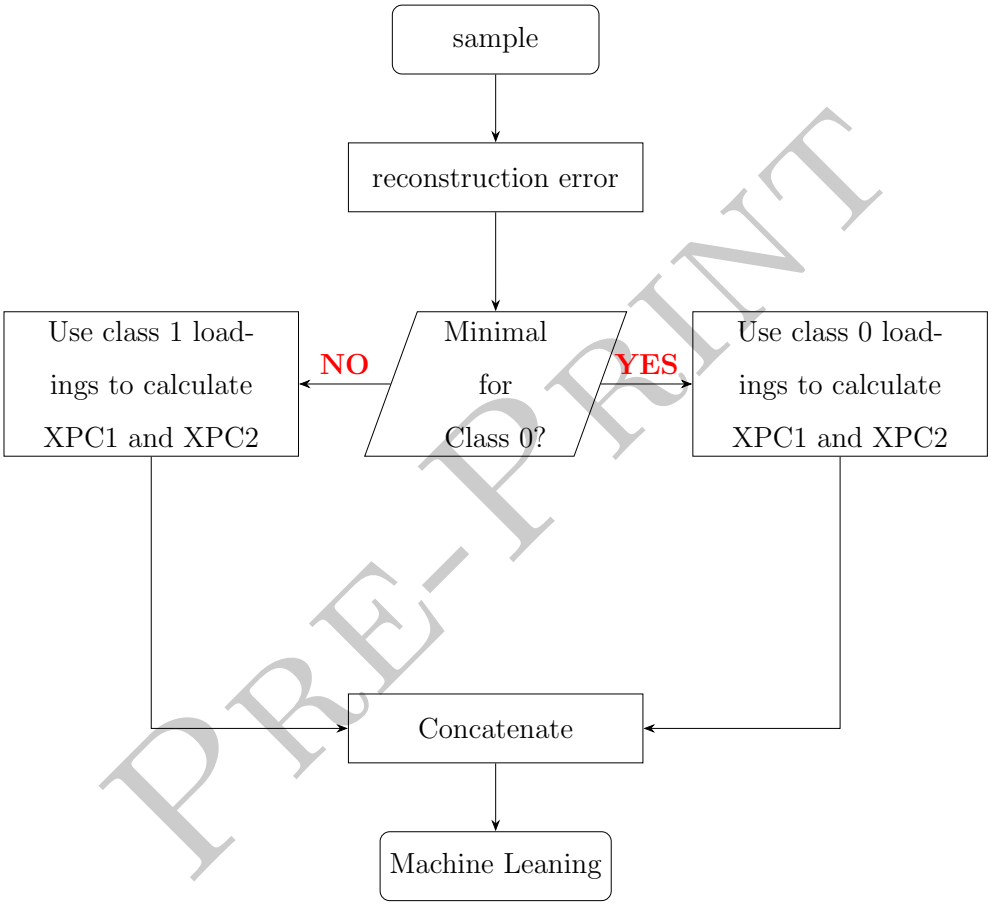


Figure 3: Feature Engineering procedure based on the reconstruction error

2.4.1. Logistic Regression

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

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

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

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

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

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

To estimate the coefficients θ we first need to calculate the maximum likelihood function of equation 11:

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

and its logarithmic form:

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

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

2.4.2. Kernel-Support Vector Machines

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

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

and the decision function is nothing but *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:

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

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

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

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

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

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

Finally, the classification function becomes:

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

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

2.4.3. Artificial Neural Networks

ANNs are mathematical models used to approximate non-linear relationships between an input space and an output space. The multi-layer perceptron (MLP) is the most common topology when the input data is numerical. An MLP usually consists of an input layer, a hidden layer (universal approximator), and an output layer (Negnevitsky, 2005) as in Figure 4. The output of the k^{th} hidden neuron can be represented by the following equation:

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

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

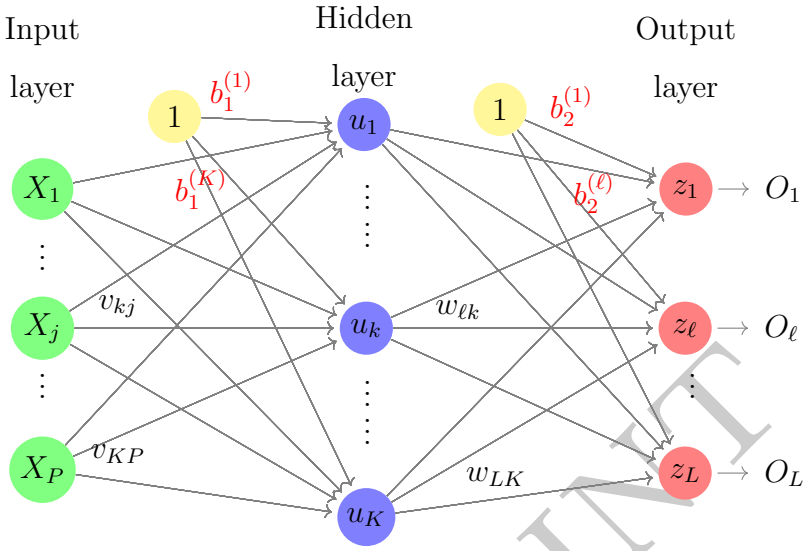


Figure 4: Typical architecture of a multi-layer network with a hidden layer

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

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

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

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

where y_i is the label and $p(y_i)$ its predicted probability. We initialized the weights following Glorot's algorithm (Glorot and Bengio, 2010) to break symmetry during backpropagation.

2.5. Evaluation metrics

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

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

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

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

were able to predict correctly with our model.

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

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

$$Precision = \frac{TP}{TP + FP} \tag{22}$$

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

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{23}$$

- Specificity (True negative rate) is the ratio of true negatives with respect to all negative outcomes. It represents the percentage of negative samples that got the correct label.

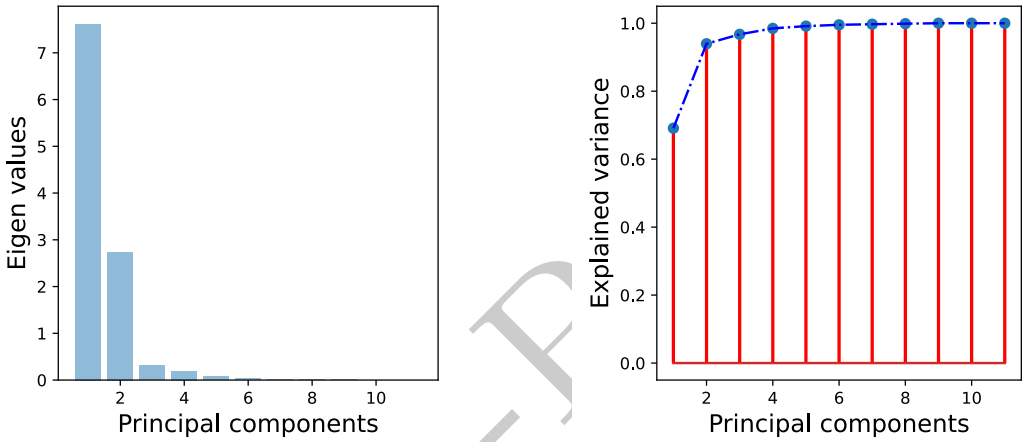
$$Specificty = \frac{TN}{TN + FP} \tag{24}$$

3. Results on acoustic voice-based dementia classification

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

3.1. First method: classical PCA

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



(a) Eigenvalues

(b) Accumulated explained variance (in %)

Figure 5: Eigenvalues and explained variance for each principal component

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

Table 1: Metrics for the ANN model with 16 variables

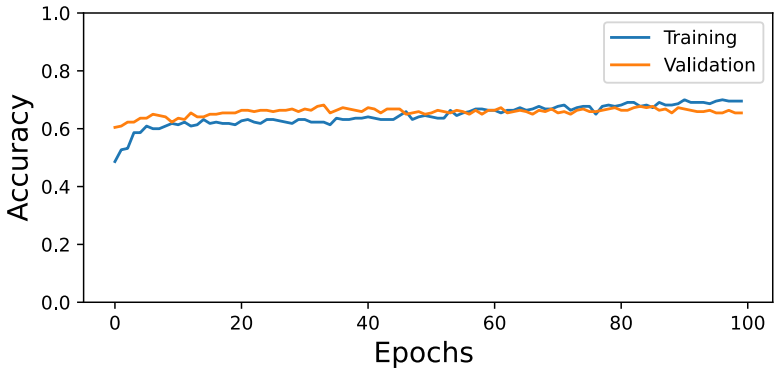
	Total	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-score
Training	440	197	122	71	50	0.725	0.73	0.797	0.76
Testing	111	42	27	22	20	0.62	0.65	0.68	0.7

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

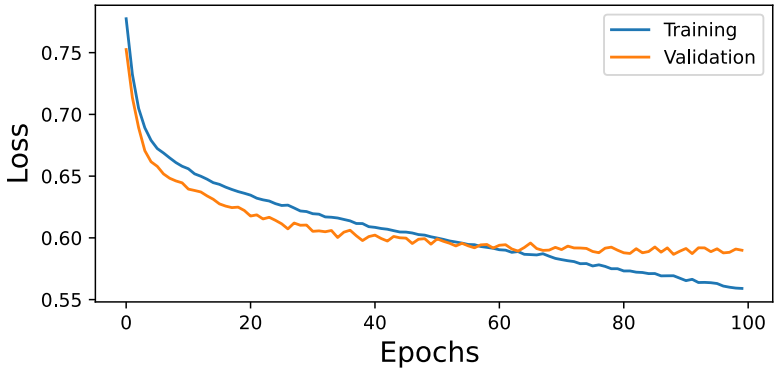
The classification results and metrics evaluation are presented in Table 1. The metrics show that the model works fine on the training data, with 79.7% of actual dementia samples being correctly classified (recall), and that among those classified as dementia, 73% are real dementia patients (precision).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

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

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

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

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

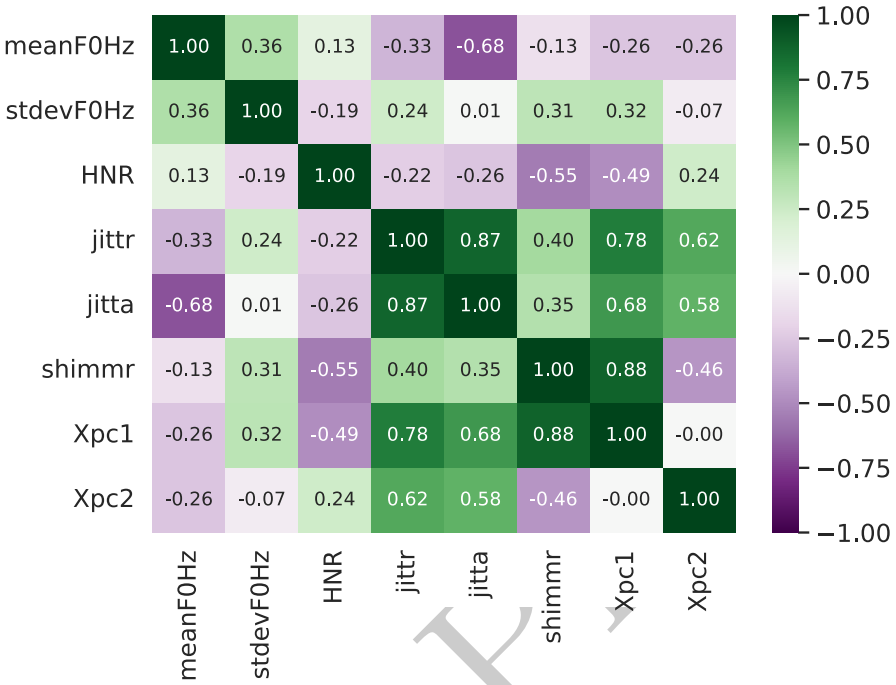


Figure 7: Linear correlation coefficients among variables

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

3.2. Second method: class-dependent PCA

This approach consists of applying PCA on the data corresponding to each class separately, as shown below. This separation ensures that we are capturing class-specific variations independently. Based on the reconstruction error criterion, we choose the *loadings* to compute the principal components of each class. Studying the linear correlation between the variables showed a high association among the shimmer variables and among the jitter 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

eliminated the highly correlated variables $\rho > 0.9$. Hence, the number of features decreased from 16 to 8. We then studied four scenarios based on the reduced feature space (see Table 3). In the first scenario S1, we included all the variables, whereas we excluded XPC1 and XPC2 in S2 and S3, respectively. The last scenario, S4, did not include any of the added features. We scaled and standardized the data in order to avoid scaling invariant issues that often lead to underfitting. Data was initially divided into 80% for training data and 20% for test data.

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	✓	✓		

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

accuracy metric. For each of the 4 scenarios, the logistic regression model 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%

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

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

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

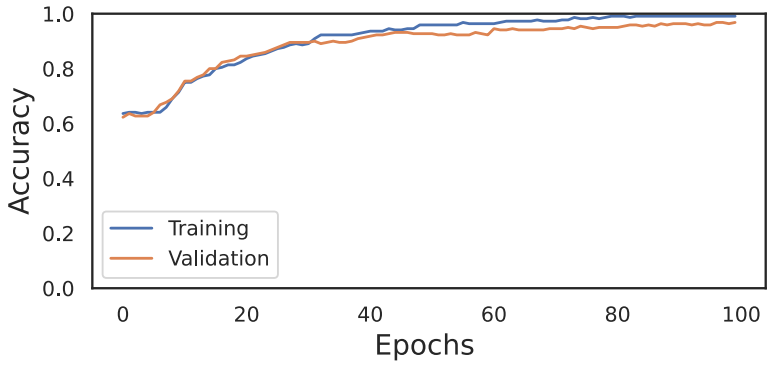
Table 5: 10 folds Cross-validation accuracy

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

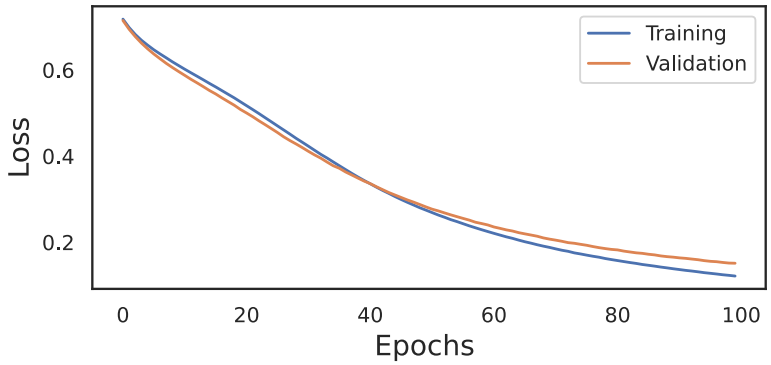
doctors to detect dementia effectively, we seek a reliable model that provides the highest accuracy. Therefore, in the following sections and paragraphs, we will focus on the model that showed the best performance, i.e, the neural network model.

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

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



(a) Training and Validation Accuracy

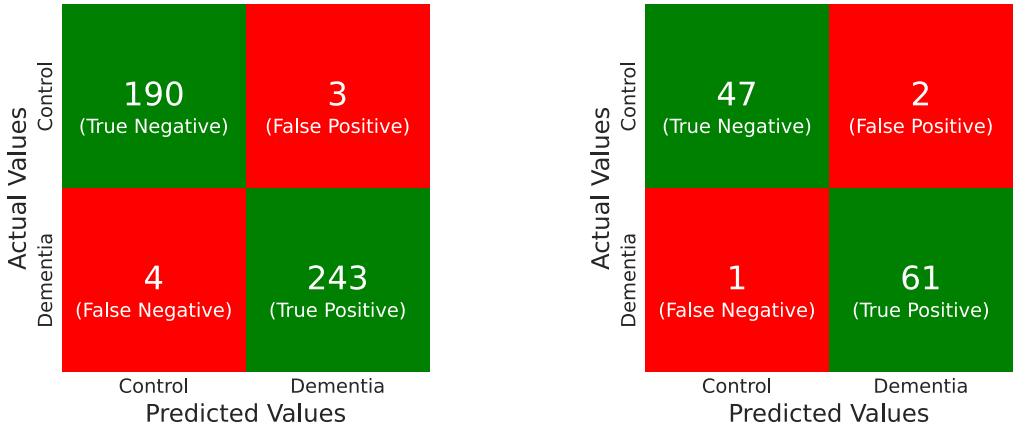


(b) Training and Validation Loss

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

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

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



(a) Training data

(b) Test data

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

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

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

4. Discussion

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

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

Table 7 demonstrates a comparison between the results we obtained in this study and those existing in the literature. For convenience, we compared with studies that used the **Dementia bank dataset only.** Comparison with other methods can be found in (Javeed et al., 2023), but were not listed here because they are not directly comparable. The system achieves an accuracy 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**

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

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

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

5. Conclusion

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

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

Data availability

The data analyzed in this study is subject to the following licenses/restrictions: in order to gain access to the datasets used in the paper, researchers must become a member of DementiaBank. Requests to access these datasets should be directed to https://dementia.talkbank.org/.

Declaration Statement - Conflict of Interest

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

Authors' contributions

Literature review by AT, MA, CL. Methodology and theoretical developments by AT. Experiment design and implementation by AT. Analysis of the experimental results by AT. Document writing and illustrations by AT, MA, CL. All authors contributed to manuscript revision, read, and approved the submitted version.

Acknowledgments

This research was led within the context of the Mementop project, co-funded by the French Agence Nationale de la Recherche (ANR) through the Plan "France Relance" and the Smart Macadam company.
We also thank the creators of the Dementia Bank database (Pitt corpus) Becker et al. (1994) and the grant support for the Pitt corpus – NIA AG03705 and AG05133.

License

References

Akhila, J., Markose, C., Aneesh, R.P., 2017. Feature extraction and classification of dementia with neural network. 2017 International Conference on Intelligent Computing, Instrumentation and Control Technolo-

gies (ICICICT) , 1446–1450URL: https://api.semanticscholar.org/CorpusID:5038545.

Bansal, D., Chhikara, R., Khanna, K., Gupta, P., 2018. Comparative analysis of various machine learning algorithms for detecting dementia. Procedia Computer Science 132, 1497–1502. doi:https://doi.org/10.1016/j.procs.2018.05.102. international Conference on Computational Intelligence and Data Science.

Becker, J.T., Boiler, F., Lopez, O.L., Saxton, J., McGonigle, K.L., 1994. The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. Archives of neurology 51, 585–594.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, in: IFA Proceedings, pp. 97–110.

Boersma, P., Weenik, D., 2022. Praat: doing phonetics by computer [computer program]. https://www.fon.hum.uva.nl/praat/.

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., Cappa, S.F., 2017. Connected speech in neurodegenerative language disorders: A review. Frontiers in Psychology 8. doi:10.3389/fpsyg.2017.00269.

Chen, J., Ye, J., Tang, F., Zhou, J., 2021. Automatic detection of alzheimer's disease using spontaneous speech only, in: Proc. Interspeech 2021. doi:10.21437/Interspeech.2021-2002.

Chen, T., Su, P., Shen, Y., Chen, L., Mahmud, M., Zhao, Y., Antoniou, G., 2022. A dominant set-informedinterpretable fuzzy system forautomated

diagnosis ofdementia. Frontiers in Neuroscience doi:10.3389/fnins.2022.867664.

Farrús, M., Hernando, J., Ejarque, P., 2007. Jitter and shimmer measurements for speaker recognition, in: Proc. Interspeech 2007, pp. 778–781. doi:10.21437/Interspeech.2007-147.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press.

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

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer series in statistics, Springer. URL: https://books.google.fr/books?id=eBSgoAEACAAJ.

Haulcy, R., Glass, J., 2021. Classifying alzheimer's disease using audio and text-based representations of speech. Frontiers in Psychology 11. doi:10.3389/fpsyg.2020.624137.

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

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics, Springer US. URL: https://books.google.fr/books?id=5dQ6EAAAQBAJ.

Javeed, A., Dallora, A.L., Berglund, J.S., Ali, A., Ali, L., Anderberg, P., 2023. Machine learning for dementia prediction: A systematic review and future research directions. Journal of Medical Systems 47. doi:10.1007/s10916-023-01906-7.

Jolliffe, I.T., 2002. Principal component analysis for special types of data. Springer.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. CoRR abs/1412.6980. URL: https://api.semanticscholar.org/CorpusID:6628106.

Kolmogorov, A., 1957. On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition. Doklady Akademii. Nauk USSR 114, 679–681.

López-de-Ipiña, K., Alonso, J.B., Solé-Casals, J., Barroso, N., Henriquez, P., Faundez-Zanuy, M., Travieso, C.M., Ecay-Torres, M., Martínez-Lage, P., Eguiraun, H., 2015. On automatic diagnosis of alzheimer's disease based on spontaneous speech analysis and emotional temperature. Cognitive Computation 7, 44–55. doi:10.1007/s12559-013-9229-9.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Journal of Cognitive Neuroscience 19, 1498–1507. URL: https://doi.org/10.1162/jocn.2007.19.9.1498, doi:10.1162/jocn.2007.19.9.1498.

Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E., Carro, J., 2020. Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia. Behavioural Neurology, Hindawi 2020. doi:10.1155/2020/4683573.

Negnevitsky, M., 2005. Artificial intelligence: a guide to intelligent systems. 2 ed., Addison-Wesley, New York.

Orimaye, S.O., Wong, J.S.M., Golden, K.J., 2014. Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances, in: Resnik, P., Resnik, R., Mitchell, M. (Eds.), Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Association for Computational Linguistics, Baltimore, Maryland, USA. pp. 78–87. URL: https://aclanthology.org/W14-3210, doi:10.3115/v1/W14-3210.

Pan, F., Zhang, Z.X., Liu, B.D., Xie, J.J., 2020. Class-specific sparse principal component analysis for visual classification. IEEE Access 8, 110033–110047. doi:10.1109/ACCESS.2020.3001202.

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, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Qiao, Y., Xie, X.Y., Lin, G.Z., Y.Zou, Chen, S.D., Ren, R.J., Wang, G.,

2020. Computer-assisted speech analysis in mild cognitive impairment and alzheimer's disease: A pilot study from shanghai, china. Journal of Alzheimer's Disease 75, 211–221. doi:10.3233/JAD-191056.

Rasmussen, J., Langerman, H., 2019. Alzheimer's disease – why we need early diagnosis. Degenerative Neurological and Neuromuscular Disease 24, 123–130. doi:10.2147/DNND.S228939.

Roeck, E.E.D., Deyn, P.P.D., Dierckx, E., Engelborghs, S., 2019. Brief cognitive screening instruments for early detection of alzheimer's disease: a systematic review. Alzheimer's Research & Therapy 11. doi:10.1186/s13195-019-0474-3.

Sadeghian, R., Schaffer, J.D., Zahorian, S.A., 2017. Speech processing approach for diagnosing dementia in an early stage, in: Proc. Interspeech 2017, pp. 2705–2709. doi:10.21437/Interspeech.2017-1712.

Santander-Cruz, Y., Salazar-Colores, S., Paredes-García, W.J., Guendulain-Arenas, H., Tovar-Arriaga, S., 2022. Semantic feature extraction using sbert for dementia detection. Brain Sciences 12. doi:10.3390/brainsci12020270.

Sarawgi, U., Zulfikar, W., Soliman, N., Maes, P., 2020. Multimodal inductive transfer learning for detection of alzheimer's dementia and its severity. arXiv:2009.00700.

Sharma, A., Paliwal, K.K., Onwubolu, G.C., 2006. Class-dependent pca, mdc and lda: A combined classifier for pattern classification. Pattern Recognition 39, 1215–1229. URL: https://www.sciencedirect.

com/science/article/pii/S0031320306000410, doi:https://doi.org/10.1016/j.patcog.2006.02.001.

Syed, Z.S., Syed, M.S.S., Lech, M., Pirogova, E., 2021. Automated recognition of alzheimer's dementia using bag-of-deep-features and model ensembling. IEEE Access 9, 88377–88390. doi:10.1109/ACCESS.2021.3090321.

Teixeira, J.P., Oliveira, C., Lopes, C., 2013. Vocal acoustic analysis - jitter, shimmer and hnr parameters. Procedia Technology 9, 1112–1122. doi:10.1016/j.protcy.2013.12.124.

Tsang, G., Xie, X., Zhou, S.M., 2020. Harnessing the power of machine learning in dementia informatics research: Issues, opportunities, and challenges. EEE Reviews in Biomedical Engineering 13.

Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákáski, M., Kálmán, J., 2018. A speech recognition-based solution for the automatic detection of mildcognitive impairment from spontaneous speech. Current Alzheimer Research 15, 130–138. doi:10.2174/1567205014666171121114930.

Vigo, I., Coelho, L., Reis, S., 2022. Speech- and language-based classification of alzheimer'sdisease: A systematic review. Bioengineering, MDPI 9. doi:10.3390/bioengineering9010027.

Vizza, P., Tradigo, G., Mirarchi, D., Bossio, R.B., Lombardo, N., Arabia, G., Quattrone, A., Veltri, P., 2019. Methodologies of speech analysis for neurodegenerative diseases evaluation. International Journal of Medical Informatics , 45–54doi:10.1016/j.ijmedinf.2018.11.008.

WHO, 2019. Dementia. URL: https://www.who.int/news-room/fact-sheets/detail/dementia.

Yadav, V.G., 2019. The hunt for a cure for alzheimer's disease receives a timely boost. Science Translational Medicine 11. doi:10.1126/scitranslmed.aaz0311.

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