Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3114–3122

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Classification of Alzheimer's Disease from EEG Signal Using Robust-PCA Feature Extraction

Giorgio Biagetti[a], Paolo Crippa[a], Laura Falaschetti[a,*], Simona Luzzi[b], Claudio Turchetti[a]

[a]*DII - Department of Information Engineering, Università Politecnica delle Marche, via Brecce Bianche, 12, I-60131 Ancona, Italy*

[b]*Clinica Neurologica - Dipartimento di Medicina Sperimentale e Clinica, Università Politecnica delle Marche, via Conca 71, I-60020 Ancona, Italy*

Abstract

The encephalographic (EEG) signal is an electrical signal that measures the brain activity. Due to its noninvasive acquisition process, it is often used to investigate the presence of Alzheimer's disease (AD) or other common forms of neurodegerative disorders due to brain changes, that occur most frequently in older adults. Early detection of prodromal stages of AD, in which an individual has mild but measurable cognitive deficiencies with no significant effect on the functional activity of daily living, may help to reduce mortality and morbidity. This paper proposes an investigation of the classification of AD from EEG signal using robust-principal component analysis (R-PCA) feature extraction algorithm. Four widely used machine learning algorithms such as k-nearest neighbor (kNN), decision tree (DT), support vector machine (SVM), and naive Bayes have been implemented and compared by using a custom dataset composed of 13 subjects healthy or affected by AD in order to asses their classification performance.

Keywords: Machine learning; classification; robust-PCA; EEG; Alzheimer's disease.

1. Introduction

According to the *2019 Revision of World Population Prospects*, the total world population aged 60 years or over is estimated to reach 1.049748 billion in 2020, nearly doubling its size in the last fifteen years. Alzheimer's disease (AD) is the most common form of dementia due to brain changes, that occur most frequently in older adults and dramatically increases with age [24, 1]. The baby boom generation has already begun to reach age 65 and beyond, the age range of greatest risk of AD; in fact, the oldest members of the baby boom generation turn age 74 in 2020. Again, the number of elderly persons is expected to grow more rapidly in the next years: in particular this number

* Corresponding author. Tel.: +39-071-220-4129.
E-mail address: l.falaschetti@univpm.it

1877-0509
10.1016/j.procs.2021.09.084

is expected to double again by 2050 reaching nearly 2.08 billion [25]. The World Alzheimer Day 2016 estimated that the incidence of AD, 46.8 million people in 2015, will double in 2030. As the world population living with Alzheimer's dementia increases, the burden of caring for that population also increases. The impact on society caused by this people ageing is becoming more and more relevant worldwide. The Alzheimer's disease (along with the other age-related neurodegerative diseases, such as Parkinson's disease) and dementia, are often characterized by the progressive declining of functional and cognitive performances. Thus, the management of people affected by these diseases is very hard especially due the behavioural and psychiatric progressive symptoms that are often present in this kind of patients. As a result, caring for these persons is particularly stressful and the quality of life experienced by unpaid family caregivers is lower than the quality of life of caregivers for persons who do not have neurodegenerative diseases [16, 8]. Because it has been demonstrated that lower quality of life results in increasing absences from work and in reduced job productivity, therefore identifying systems and technologies able to early detection of patients affected of this kind of diseases will be of great value.

Mild cognitive impairment (MCI) is a condition in which an individual has mild but measurable cognitive deficiencies with no significant effect on the functional activity of daily living. Early detection of MCI, a prodromal stage of AD, may help to clarify the mechanism of brain changes and to reduce mortality and morbidity [2].

Analysis of electroencephalogram (EEG) signal, representing the electrical activity of the brain, could lead to the development of an indicator of AD and other presenile neurodegerative diseases. The EEG signals, due to its noninvasive acquisition process, is often used for the extraction of features to characterize the AD.

Classification with supervised machine learning methods applied to EEG signal may support the diagnosis of AD [7, 9] and other neurological diseases, such as epilepsy [20, 23], depression [18] and alcoholism [11], helping to discriminate the healthy control subjects. In this context, principal components analysis (PCA), independent components analysis (ICA) and linear discriminant analysis (LDA), and other methods to reduce the dimension of data are commonly used [23].

This paper proposes an investigation of the classification of Alzheimer's disease from EEG signal using robust-PCA feature extraction. Several machine learning algorithms have been implemented in the last decade in order to perform classification, regression, and in general pattern recognition tasks. Among the most common algorithms we selected the k-nearest neighbor (kNN), the decision tree (DT) [3], the support vector machine (SVM) and the naive Bayes on the basis of their simple implementation and reduced computational time. We investigated their performance considering a custom dataset composed of 13 subjects healthy or affected by AD. Accuracy and computational time of the four algorithms have been compared in order to give useful insights for AD detection system designers in adopting the one that best fit the system they are implementing.

The rest of this paper is organized as follows: Section 2 is divided into three subsections that present the used data set, machine learning algorithms and feature extraction, respectively. Section 3 discusses the experimental results and related findings. Finally, Section 4 gives some conclusions.

2. Material and Methods

2.1. Dataset

The electroencephalographic signals used in this work are related in part to patients suffering from the Alzheimer's disease and in part to healthy people The signals were acquired through an anonymous source, in the form of "raw" data in the following formats: European data format (EDF) [14, 13], DAT, and GNT . The data have been then divided into two folders named in relation to the two specific classes, i.e.

- AD: containg the EEG signals acquired from subjects affected by AD.
- HS: containg the EEG signals acquired from healthy subjects.

The subjects whose data have been acquired are in total 13, of which 7 correspond to the AD class and the remaining 6 correspond to the HS class. For each subject a number (ranging from 21 to 23) of continuous EEG signals have been acquired through the electrodes, for a total of 283 observations. Details on the available number of observations for each class is reported in Table 1, each observation was adjusted (by truncation) to a common length of 148096 samples

(predictors) at 128 Hz. Among the various formats, with which the data were initially formatted, the EDF format was chosen with consequent conversion to MAT format. This procedure was performed for each subject belonging to both classes, obtaining 13 matrices of different sizes.

Table 1. Dataset consistency.

Class label	Class ID	Number of subjects	Number of observations
AD	1	7	154
HS	2	6	129
Total		13	283

The steps of data conversion, creation of data_matrix (data matrix) with application of feature extraction techniques and classification were carried out using the Matlab scientific computing environment (Matrix Laboratory). Through the generation of the matrices, for each subject the EEG traces were plotted, but in order to have a more accurate representation of the signals it was necessary to use specific software for files with the chosen extension, called edfbrowser[1]. It is a free, open source, cross-platform universal viewer and toolbox, designed for, but not limited to, timeseries such as EEG, EMG, ECG, BioImpedance and other file formats. The software therefore allowed the representation, in a single window, of 21-23 traces of each patient, lasting an average of about 24 minutes, with a minimum of about 19 minutes, and in which the scale time was compressed to the entire recording interval and the amplitude adapted to the size of the window. Two examples of the EEG signal extracted from the considered dataset have been displayed in Fig. 1 and Fig. 2 for an AD and an HS subject, respectively.

2.2. Classification

2.2.1. K-nearest neighbor (kNN)

This classifier is one of the most popular neighborhood classifiers in pattern recognition and machine learning because of its simplicity and efficiency. It categorizes each unlabelled test example using the label of the majority of examples among its k-nearest (most similar) neighbors in the training data set. The similarity depends on a specific distance metric, therefore, the performance of the classifier strictly depends on the distance metric used. However, it suffers of memory requirements and time complexity, because it is fully dependent on every example in the training set [26].

2.2.2. Decision tree

The decision tree classifier partitions the input space into small segments, and labels these small segments with one of the various output categories. However, conventional decision tree only does the partitioning to the coordinate axes. With the growth of the tree, the input space can be partitioned into very small segments so as to recognize subtle patterns [19]. The main drawback is that overgrown trees could lead to overfitting.

2.2.3. Support Vector Machine (SVM)

Support vector machine is an automatic supervised learning algorithm for classification or regression problems, in which the algorithm build the classes from the dataset, in such a way as to classify any new data. Given a set of training data, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new data to one category or the other, making it a non-probabilistic binary linear classifier. SVM maps training data to points in space so as to maximise the width of the gap between the two categories. New data are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall [5].

[1] https://www.teuniz.net/edfbrowser/

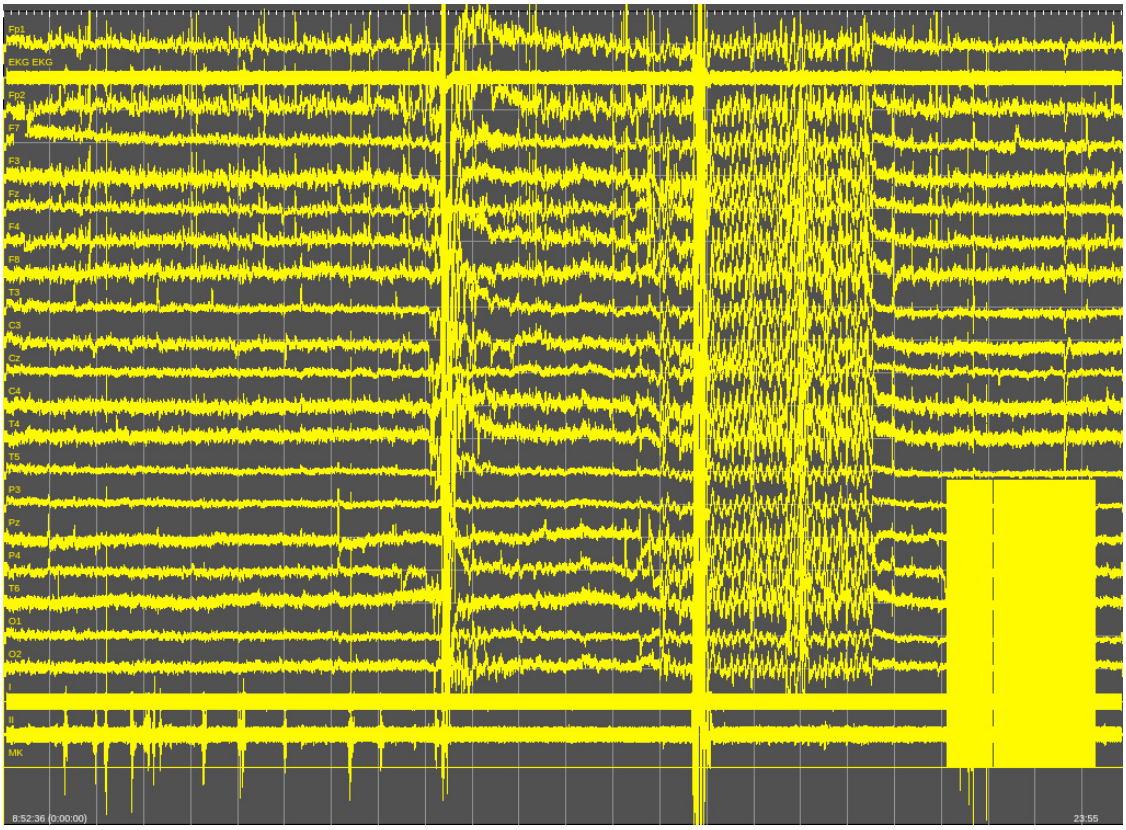

Fig. 1. Example of an EEG trace of the considered dataset - AD subject.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. In this work both quadratic kernels and Gaussian kernels, two commonly adopted choices, have been considered.

2.2.4. *Naive Bayes*

The so called naive Bayes algorithm is a Bayesian classifier that assigns the most likely class to a given example described by its feature vector. The learning of such classifier is greatly simplified by the assumption that the features are independent given class, that is,

$$\mathcal{P}(\mathbf{x}|\Omega) = \prod_{i=1}^{N} \mathcal{P}(x_i|\Omega) \,, \tag{1}$$

where $\mathbf{x} = (x_1, \cdots x_N)$ is a feature vector and Ω is a class. Despite this unrealistic assumption, the naive Bayes classifier has been demonstrated to be very successful in many cases, often competing with much more sophisticated techniques [22, 6, 15]. In particular, naive Bayes has proven effective in a number of practical applications such as medical diagnosis and systems performance management [17]. The success of naive Bayes in the presence of feature dependencies can be explained as follows: optimality in terms of zero-one loss(classification error) is not necessarily related to the quality of the fit to a probability distribution (i.e., the appropriateness of the hypothesis of independence). Rather, an optimal classifier is built as long as both the actual and estimated distributions agree on the most-probable class [6].

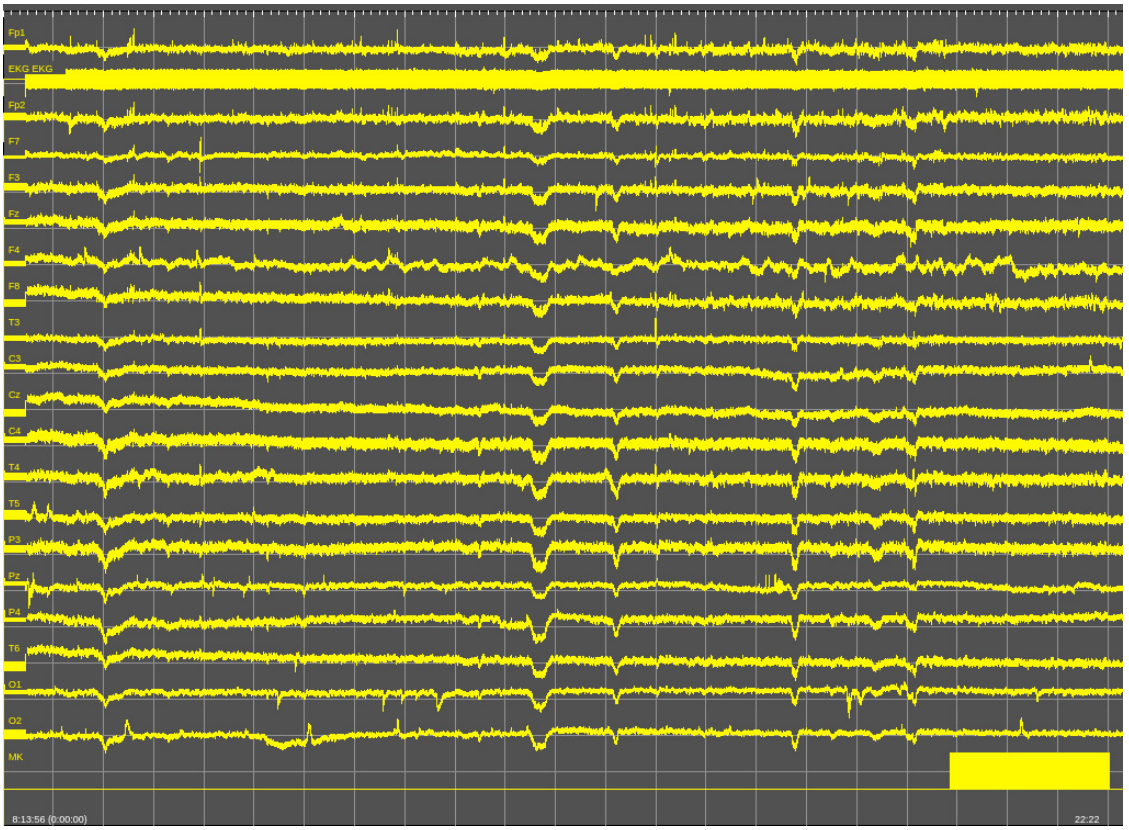

Fig. 2. Example of an EEG trace of the considered dataset - HS subject.

2.3. Feature Extraction

2.3.1. Principal Component Analysis (PCA)

Principal component analysis is a widely adopted technique to reduce the number of predictors [28, 12], and is based on the singular value decomposition (SVD) of the data matrix. The idea is that singular vectors associated to small singular values can be omitted from the representation of the data, thus reducing the number of projections needed for its reconstruction.

Mathematically, given a row-oriented data matrix D, it can be written by means of the SVD as:

$$D = U \Lambda V^T \,, \tag{2}$$

where Λ is a diagonal matrix. A generic signal s can then be expressed as the superposition of the columns $V^{(i)}$ of V:

$$s \simeq \sum_{i=1}^{L} p_i V^{(i)T} \,, \tag{3}$$

where $p = s\,V$ and L can usually be much lower than the rank of the original data matrix.

2.3.2. Robust PCA

Robust PCA [30, 4, 29] is a variation on the basic PCA methodology, which essentially repeatedly performs a PCA on the data matrix and rebuilds it after excluding the components whose singular value is below a predetermined threshold. This way, a low-rank matrix L can be determined, so that the original data matrix D can be written as $D = L + S$, with the constraint that S must be a sparse matrix.

PCA is a very popular dimension reduction technique that is widely used as a first step in the analysis of high-dimensional data. However, the classical approach that is based on the mean and the sample covariance matrix of the data is very sensitive to outliers. Moreover, classification methods based on this covariance matrix do not give good results in the presence of outlying measurements. In literature, the common approach to solve the problem of outlier detection is to apply the Robust PCA algorithm, so that inliers lie in a low dimensional subspace [27, 10, 21].

3. Experimental Results

To perform a set of experiments to determine the best performing classification algorithm on the data at hand, the available data was split into a training set, comprising 4 healthy subjects and 5 subjects with the Alzheimer disease, totaling 195 observations, and a test set, comprising 2 subjects for each categories, and totaling 88 observations. Table 2 summarizes this separation. The number of predictors, i.e., the number of samples in the signals, have been uniformized by truncating those longer than the reported value.

Table 2. Dataset consistency for training/testing.

	Number of subjects	Class ID	Number of observations	Number of predictors
Train	9	[5 AD subj, 4 HS subj]	195	148096
Test	4	[2 AD subj, 2 HS subj]	88	148096
Total	13	[7 AD subj, 6 HS subj]	283	148096

Before applying the R-PCA and trying all the above mentioned classification algorithms on the dataset, two preliminary experiments have been made. As a first try, to establish a baseline, the classification algorithm were applied to the raw data without any specific pre-processing, using all the samples as predictors.

The results are reported in Table 3. As could be expected, given the huge number of predictors, none of the classifiers was able to classify the data in the test set with any reliability.

Table 3. Classification accuracy of the original data matrix - Split: train = 9 subjects, test = 4 subjects.

	kNN [%]	Decision Tree [%]	SVM quadratic [%]	SVM Gaussian [%]	Naive Bayes [%]
Accuracy	52.27%	34.09%	38.63%	52.27%	42.04%
Training Time	1.800 s	7.066 s	5.578 s	11.255 s	462.186 s
Testing Time	2.097 s	0.466 s	0.836 s	3.450 s	5.838 s

It is apparent that the number of predictors must be reduced before training the classifiers. To this end, the R-PCA was performed on the data matrix to remove outliers, where the singular vectors of the resulting low-rank matrix have been used to project the data set and obtain a reduced number of predictors to use in the classification problem. As a reference, the eigenvalue spectrum, which can be useful to select the optimum number of principal components to retain, is reported in Figure 3.

With these preliminary tests performed, the classifiers were then tested using as data the projections onto the first components of the R-PCA, varying the number of said components to highlight the optimum number of predictors. Results are shown in Table 4.

As can be seen, only the Gaussian SVM and the naive Bayes methods achieve significant, and similar, results. Among these two, whose best performance is also reported in Table 5 for easier reference, SVM has a significantly lower computational cost during testing, and can also achieve a higher overall accuracy, but only when using a very

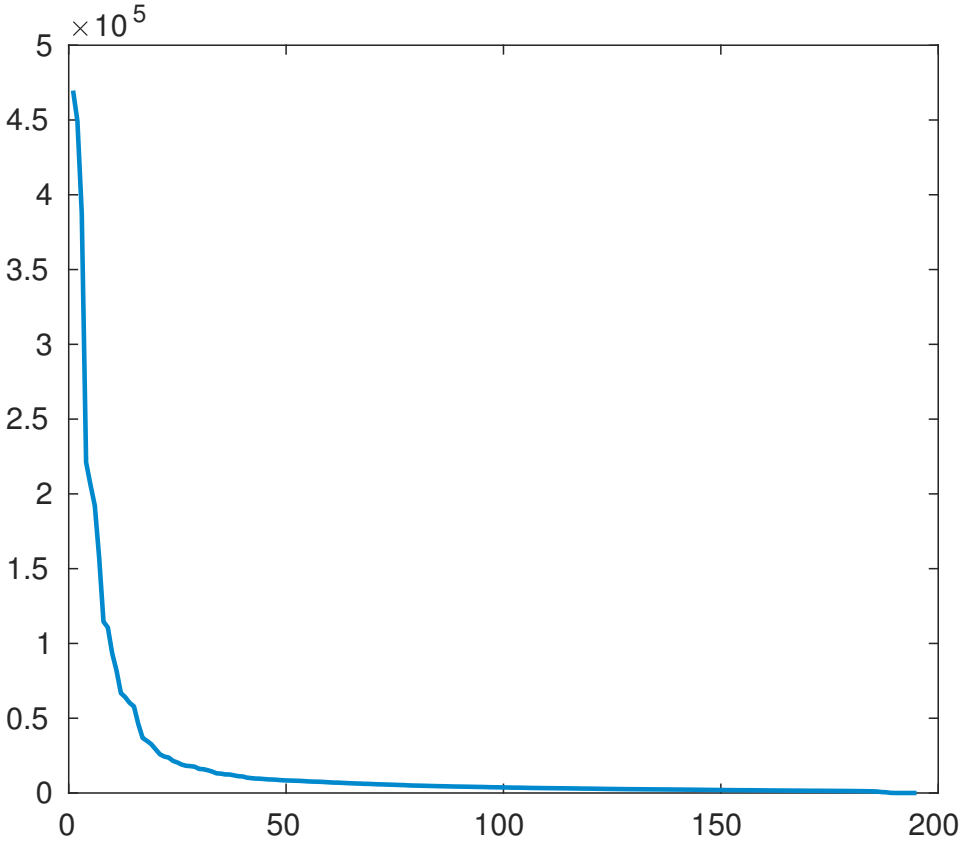

Fig. 3. Eigenvalues from R-PCA.

large number of components. From the eigenvalue spectrum, it is indeed apparent that about 30–50 components should suffice to obtain a detailed description of the signal, but the SVM classifier cannot work in this reduced-space case. On the other hand, although the naive Bayes classifier achieves only a slightly lower overall accuracy, it does so using much fewer components, and its performance at 40 components (86.36%) is also remarkable and a sign that this classifier works well combined with R-PCA. Unfortunately the Bayes approach needs to estimate a probability density in the feature space, and the difficulty of this task increases dramatically with the number of dimensions, so its classification performance starts to get worse if the number of components is increased too much.

Table 4. Classification accuracy of the original data matrix with R-PCA applied - Split: train = 9 subjects, test = 4 subjects.

R-PCA components	kNN [%]	Decision Tree [%]	SVM quadratic [%]	SVM Gaussian [%]	Naive Bayes [%]
1	52.27	48.86	52.27	46.59	47.72
2	51.13	53.40	48.86	47.72	51.13
3	44.31	46.59	48.86	46.59	50.00
4	52.27	50.00	47.72	47.72	50.00
5	53.40	43.18	50.00	47.72	50.00
6	53.40	52.27	51.13	47.72	48.86
7	62.50	55.68	51.13	52.27	50.00
8	56.81	52.27	51.13	54.54	51.13
9	54.54	59.09	51.13	56.81	51.13
10	44.31	59.09	53.40	48.86	51.13
20	51.13	40.90	54.54	47.72	46.59
30	45.45	61.36	46.59	44.31	80.68
40	53.40	61.36	52.27	53.40	86.36
50	63.63	61.36	52.27	53.40	84.09
60	56.81	61.36	51.13	77.27	81.81
70	48.86	61.36	47.72	70.00	82.95
80	54.54	61.36	46.59	84.09	87.50
90	52.27	61.36	44.31	87.50	**90.91**
100	59.09	61.36	43.18	88.63	90.91
110	48.86	61.36	45.45	92.04	89.77
120	51.13	61.36	45.45	92.04	87.50
130	50.00	61.36	45.45	92.04	71.59
140	45.45	61.36	45.45	92.04	61.36
150	52.27	61.36	45.45	**93.18**	61.36

Table 5. Performance of the original data matrix versus R-PCA data matrix, in terms of accuracy and computation time.

Data matrix	N. predictors	Algorithm	Accuracy	Training time	Testing time
Original	148096	kNN	52.27%	1.847 s	2.299 s
R-PCA	150	SVM Gaussian	93.18%	1.589 s	0.034 s
R-PCA	90	Naive Bayes	90.91%	1.560 s	0.121 s

4. Conclusion

In this paper four widely used machine learning algorithms, such as k-nearest neighbor, decision tree, support vector machine and naive Bayes have been implemented in order to investigate their ability in the detection of Alzheimer disease from EEG signal. In order to reduce the problem dimensionality and remove outliers, before the training the robust PCA was performed on the data matrix obtained from a custom dataset composed of 13 subjects healthy or affected by AD. As a result of this investigation the Gaussian SVM and the naive Bayes methods achieve good similar, results taking into account that SVM has a much lower computational cost during testing, besides showing a slightly higher accuracy at the cost of needing way more predictors to reach its best results.

Acknowledgements

This research was supported by a Università Politecnica delle Marche Research Grant.

References

[1] Alzheimer's Association and others, 2018. Alzheimer's Association report, 2018 Alzheimer's disease facts and figures. Alzheimer's & Dementia 14, 367–429.
[2] Biagetti, G., Crippa, P., Falaschetti, L., Luzzi, S., Santarelli, R., Turchetti, C., 2019. Classification of Alzheimer's disease from structural magnetic resonance imaging using particle-Bernstein polynomials algorithm, in: Czarnowski, I., Howlett, R.J., Jain, L.C. (Eds.), Intelligent Decision Technologies 2019, Springer Singapore, Singapore. pp. 49–62.
[3] Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth & BrooksCole Advanced Books & Software., Monterey, CA.
[4] Candès, E.J., Li, X., Ma, Y., Wright, J., 2011. Robust principal component analysis? J. ACM 58.
[5] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.
[6] Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian classifier under zero-one loss. Machine learning 29, 103–130.
[7] Fiscon, G., Weitschek, E., Felici, G., Bertolazzi, P., De Salvo, S., Bramanti, P., De Cola, M.C., 2014. Alzheimer's disease patients classification through EEG signals processing, in: 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp. 105–112.
[8] Hazzan, A.A., Ploeg, J., Shannon, H., Raina, P., Oremus, M., 2013. Association between caregiver quality of life and the care provided to persons with Alzheimer's disease: protocol for a systematic review. Systematic Reviews 2, 17.
[9] Houmani, N., Vialatte, F., Gallego-Jutglà, E., Dreyfus, G., Nguyen-Michel, V.H., Mariani, J., Kinugawa, K., 2018. Diagnosis of Alzheimer's disease with electroencephalography in a differential framework. PloS one 13, e0193607.
[10] Hubert, M., Engelen, S., 2004. Robust PCA and classification in biosciences. Bioinformatics 20, 1728–1736.
[11] Jiajie, L., Narasimhan, K., Elamaran, V., Arunkumar, N., Solarte, M., Ramirez-Gonzalez, G., 2018. Clinical decision support system for alcoholism detection using the analysis of EEG signals. IEEE Access 6, 61457–61461.
[12] Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 20150202.
[13] Kemp, B., 2002. European data format (EDF): current availability and additional applications. Journal of Sleep Research, Supplement 11.
[14] Kemp, B., Värri, A., Rosa, A.C., Nielsen, K.D., Gade, J., 1992. A simple format for exchange of digitized polygraphic recordings. Electroencephalography and clinical neurophysiology 82, 391–393.
[15] Langley, P., Iba, W., Thompson, K., et al., 1992. An analysis of Bayesian classifiers, in: Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92), Citeseer. AAAI Press, San Jose, CA, USA. pp. 223–228.
[16] McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia 7, 263 – 269.
[17] Mitchell, T.M., et al., 1997. Machine learning. McGraw-hill New York.
[18] Peng, H., Xia, C., Wang, Z., Zhu, J., Zhang, X., Sun, S., Li, J., Huo, X., Li, X., 2019. Multivariate pattern analysis of EEG-based functional connectivity: A study on the identification of depression. IEEE Access 7, 92630–92641.
[19] Quinlan, J.R., 1986. Induction of decision trees. Machine Learning 1, 81–106.
[20] Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., 2020. Review of methods for EEG signal classification and development of new fuzzy classification-based approach. IEEE Access 8, 189720–189734.
[21] Rahmani, M., Li, P., 2019. Outlier detection and robust PCA using a convex measure of innovation., in: NeurIPS, pp. 14200–14210.
[22] Rish, I., et al., 2001. An empirical study of the naive bayes classifier, in: IJCAI 2001 workshop on empirical methods in artificial intelligence, pp. 41–46.
[23] Subasi, A., Ismail Gursoy, M., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications 37, 8659–8666.
[24] Todd, S., Barr, S., Roberts, M., Passmore, A.P., 2013. Survival in dementia and predictors of mortality: A review. International journal of geriatric psychiatry 28, 1109–1124.
[25] United Nations, . World Population Prospects – Population Division. [Online]. Available: https://population.un.org/wpp/. Accessed: 15-Nov-2020.
[26] Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10, 207–244.
[27] Xu, H., Caramanis, C., Mannor, S., 2013. Outlier-robust PCA: The high-dimensional case. IEEE Transactions on Information Theory 59, 546–572.
[28] Ye, J., Janardan, R., Li, Q., 2004. Two-dimensional linear discriminant analysis. Advances in neural information processing systems 17, 1569–1576.
[29] Yi, X., Park, D., Chen, Y., Caramanis, C., 2016. Fast algorithms for robust pca via gradient descent, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA. p. 4159–4167.
[30] Zhang, T., Yang, Y., 2018. Robust PCA by manifold optimization. The Journal of Machine Learning Research 19, 3101–3139.