Neural Computing and Applications (2020) 32:15761–15769
https://doi.org/10.1007/s00521-018-3494-1

S.I. : ADVANCES IN BIO-INSPIRED INTELLIGENT SYSTEMS

On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment

K. López-de-Ipiña[1] · U. Martinez-de-Lizarduy[1] · P. M. Calvo[1] · B. Beitia[1] · J. García-Melero[1] · E. Fernández[1] · M. Ecay-Torres[2] · M. Faundez-Zanuy[3] · P. Sanz[4]

Received: 15 January 2018 / Accepted: 18 April 2018 / Published online: 8 May 2018

Abstract

Alzheimer's disease is characterized by a progressive and irreversible cognitive deterioration. In a previous stage, the so-called Mild Cognitive Impairment or cognitive loss appears. Nevertheless, this previous stage does not seem sufficiently severe to interfere in independent abilities of daily life, so it is usually diagnosed inappropriately. Thus, its detection is a crucial challenge to be addressed by medical specialists. This paper presents a novel proposal for such early diagnosis based on automatic analysis of speech and disfluencies, and Deep Learning methodologies. The proposed tools could be useful for supporting Mild Cognitive Impairment diagnosis. The Deep Learning approach includes Convolutional Neural Networks and nonlinear multifeature modeling. Additionally, an automatic hybrid methodology is used in order to select the most relevant features by means of nonparametric Mann–Whitney U test and Support Vector Machine Attribute evaluation.

Keywords Mild Cognitive Impairment · Automatic Speech Analysis · Deep Learning · Convolutional Neural Networks · Nonlinear features · Disfluencies

1 Introduction

Alzheimer's disease (AD) is characterized by a progressive and irreversible cognitive deterioration, which includes memory loss and impairments in emotion, language, and judgment, along with other cognitive deficits and symptoms in behavior. Its prevalence keeps increasing mainly among the elderly, and as highlighted by the last World Alzheimer Reports, AD is becoming epidemic as 900 million people can be regarded as the world's elderly population, living most of them in developed countries [1]. Therefore, an early and accurate diagnosis of AD helps patients and relatives to plan the future and offers the best possibilities that symptoms could be treated.

✉ K. López-de-Ipiña
karmele.ipina@ehu.eus

U. Martinez-de-Lizarduy
unai.martinezdelizarduy@ehu.eus

P. M. Calvo
pilarmaria.calvo@ehu.eus

B. Beitia
mariablanca.beitia@ehu.eus

J. García-Melero
jgarcia@ehu.eus

E. Fernández
elsa.fernandez@ehu.eus

M. Ecay-Torres
mecay@cita-alzheimer.org

M. Faundez-Zanuy
faundez@tecnocampus.cat

P. Sanz
msanz@csdm.cat

[1] Faculty of Engineering, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), 20018 Donostia-San Sebastian, Spain

[2] Fundación CITA Alzheimer, 20009 Donostia, Spain

[3] Tecnocampus Mataró, Escola Superior Politècnica de Mataró (UPF), 08302 Mataró, Spain

[4] Neurology Department, Mataro Hospital, 08302 Mataró, Spain

In an early stage, a previous cognitive loss or Mild Cognitive Impairment (MCI) appears. Nevertheless, it does not seem sufficiently severe to interfere in abilities of daily life; thus, it usually does not receive an appropriate diagnosis, and afterward, some patients develop AD. The detection of MCI is a challenge to be addressed by medical specialists and could help future AD patients [2]. Along with memory loss, one of the main problems of AD is the loss of social and language skills. This loss can be noted in difficulties speaking to and understanding people, which make even more complicated social interactions and the natural communication process. Other crucial abilities for communication are impaired as well, such as emotion and expression. This difficulty communicating appears in early stages of the disease due to language issues, and it leads people with AD to social exclusion, with a serious negative impact not only on the patients, but also on their families [3]. During communication, language resources that include pauses or disfluencies are used to maintain verbal fluency. In AD/MCI, verbal fluency clearly changes: Speech fluency is progressively substituted by more pauses and disfluencies. Therefore, disfluencies are interesting language elements that could be useful to properly diagnose MCI. Both disfluencies and speech silences have valuable information for understanding the meaning of the uttered message.

One of the main aims of this project is to develop an automatic analysis of standard assessment tests, such as *Categorical Verbal Fluency (CVF)*, by using speech therapy techniques which will allow to obtain quickly and reliably these specific analyses [4]. In the last years, several papers in the state of the art have addressed this issue. In the present work, we focus on the integration of more robust language-independent methodologies in order to detect AD in speech, using one of the classical tasks of CVF, the so-called animals naming task. Machine Learning and Deep Learning Paradigms will be used for modeling, as well as several feature sets based on linear and nonlinear approaches in order to develop a real-time and robust system.

Section 2 describes the materials. Section 3 presents the used methods. Section 4 includes the results and discussion, and finally in Sect. 5, conclusions are drawn.

2 Materials

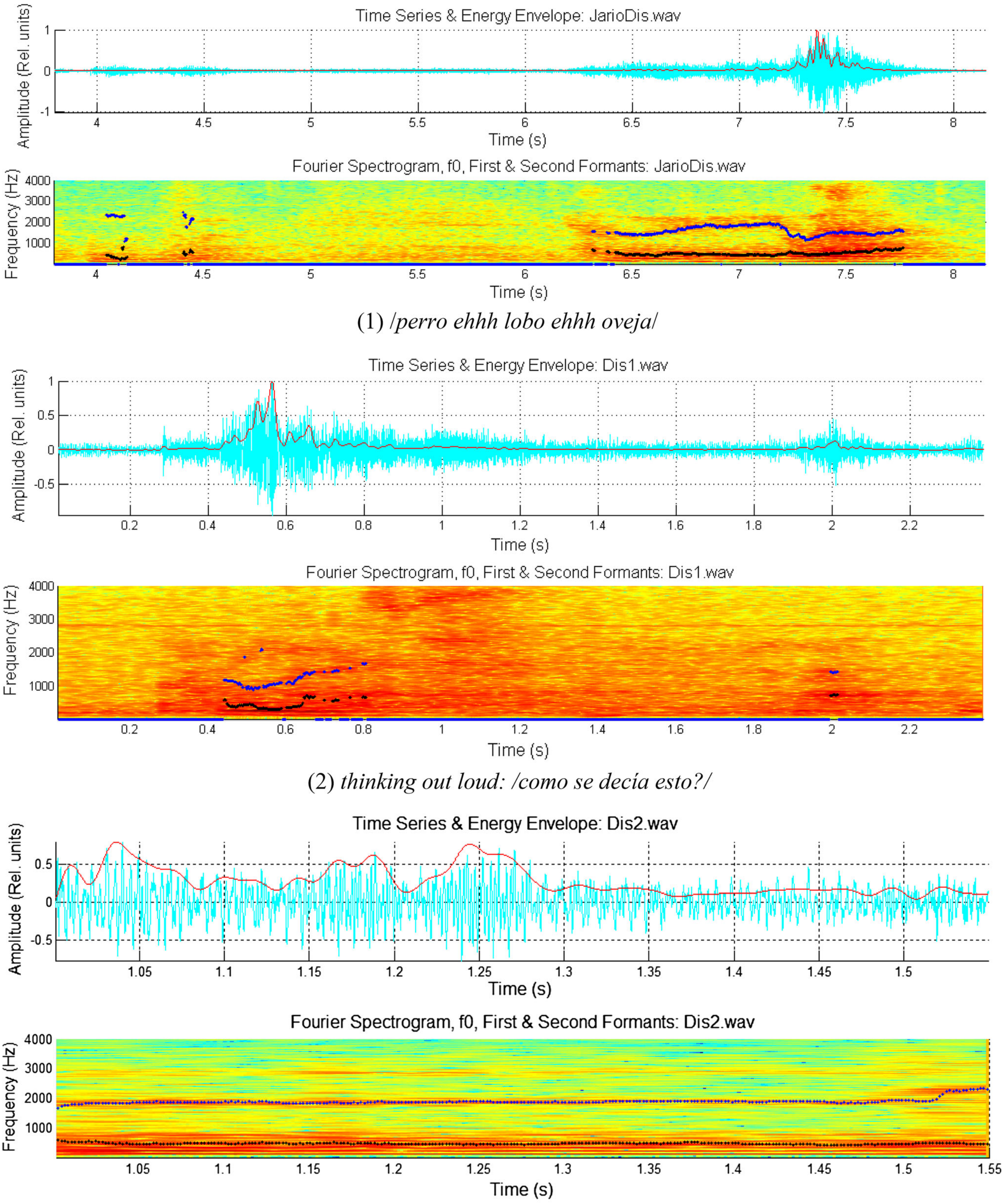
Recent studies highlight the relevance of non-speech elements such as disfluencies in verbal communication to identify MCI and AD. In [5, 6], it is suggested that more pauses and shorter recording times reflect that AD patients require a greater effort to produce speech than healthy people: AD patients speak with longer pauses, more slowly, with shorter speech segments, and they spend more time trying to find the correct word, leading to speech disfluencies or broken messages. Speech disfluencies are any irregularity, break, or non-language element that occurs during the period of fluent speech, and they can start, complement, or interrupt it. These include elements such as: false starts, restarted or repeated phrases, extended or repeated syllables, thinking out loud, grunts or non-lexical utterances such as repaired utterances and fillers, and speakers correcting mispronunciations or their own slips of the tongue [7]. If these disfluencies increase, it could be a clear sign of cognitive impairment. In AD patients, sometimes the verbal utterance reflects their internal cognitive process or inner dialogue when they think out loud: “What is that?”, “How was this…”, “/uhm/ I cańt remember,” “What was the name?”. If the number of silences and disfluencies increases, it may indicate that there is a worsening of the disease, which could lead to a deficit in effective and clear communication.

As a consequence, disfluencies play an important role in verbal communication and they are a direct reflection of the cognitive process that takes part in communication and convey an unquestionable characteristic for the diagnosis of these disorders, when fluent speech starts to disappear or is replaced by some disfluencies, Fig. 1. Although AD is mainly a cognitive disease, it may have articulation and phonation biomechanical alterations.

The Categorical Verbal Fluency task (CVF), animal naming (AN), or animal fluency task, is a test used in neurodegenerative diseases, which measures and quantifies the progression of cognitive impairment [8]. It is commonly used to assess language skills, executive functions, and semantic memory [9]. The used sample includes 187 healthy individuals and 38 MCI patients that belong to the cohort of Gipuzkoa-Alzheimer Project (PGA) of the CITA-Alzheimer Foundation [4, 10], Table (1). For the experiments, a balanced subset PGA-OREKA has been selected.

3 Methods

Recent state-of-the art approaches include modeling by means of linear and nonlinear speech features [14, 15]. The proposed approach is based on the integration of several types of optimum features to model speech and disfluencies, using both linear and nonlinear ones. Furthermore, this proposal is based on the description of speech pathologies [5, 12] with regard to articulation, phonation, quality of the speech, human perception, and the complex dynamics of the system. In this paper, some of the most used speech features (linear and nonlinear) will be taken into account for differentiation between pathological and healthy and speech [4, 5, 12–16], and discrimination



(1) */perro ehhh lobo ehhh oveja/*

(2) *thinking out loud: /como se decía esto?/*

(3) */ehhh/*

Fig. 1 Details of several utterances with disfluencies in the *Animals Naming, Categorical Verbal Fluency (CVF)* task, for an individual of the MCI group, **a** signal (top image), **b** spectrogram and formants (bottom image), by *BioMetroLing* software: first formant displayed in black, second formant displayed in blue [11]

Table 1 Demographic data of the subsets selected for the experiments: PGA-OREKA, AN task subset (CR/MCI: Control Group/MCI Group)

	Female	Male	Range of age	Age-mean	Age-SD
AN (CR/MCI)	36/21	26/17	39–74/42–79	56.73/57.15	7.8/8.9

through human perception. Most of them are well known in the field of pathological speech characterization, and therefore for each parameter, a reference is given where further information and a deeper description can be found. All features are calculated by means of software developed within our research group [4, 5], *SPSS* [17], *MATLAB* [18], *Praat* [19], and *WEKA* [20].

3.1 Automatic segmentation of disfluencies

The speech recording has been automatically segmented in disfluencies and speech signal by a *VAD* (voice activity detection) algorithm [6].

3.2 Extraction of features

After the segmentation, the following features will be extracted:

- Classical features (CF)

  1. Spectral domain features: jitter, pitch, shimmer, noise to harmonic ratio (NHR), harmonic to noise ratio (HNR), harmonicity, spectrum centroid, APQ (Amplitude Perturbation Quotient), and formants and its variants (min, max, mean, median, mode, std) [4, 5].
  2. Time-domain features: breaks, voiced/unvoiced segments, ZCR (Zero-Crossing Rate) [4, 5], and its variants (min, max, mean, median, mode, std).
  3. Energy, short time energy, intensity, and spectrum centroid. These features are sometimes extended by their first- and second-order regression coefficients (Δ and $\Delta\Delta$, respectively) [8, 12–15].

- Perceptual features (PF)

  1. Mel Frequency Cepstrum Coefficients (MFCC): These coefficients try to approach human perception. Human ears behave as some filters, and they only concentrate on some frequency components with different levels. These filters are not equispaced on the frequency axis: At low frequencies, there are more filters, and at high frequencies, there are fewer filters with different bandwidths. This type of performance is simulated by Mel Frequency analysis, and particularly Mel Frequency Cepstrum Coefficients (MFCC). Its variants are also calculated (min, max, mean, median, mode, std).
  2. These features are sometimes extended by their first- and second-order regression coefficients (Δ and $\Delta\Delta$, respectively) [4, 5, 12, 14].

- Advanced features (AF)

  1. Coefficients that provide detailed information are linked to voice quality, perception, adaptation, and amplitude modulation: PLP (Perceptual Linear Predictive coefficients), MSC (Modulation Spectra Coefficients), ICC (Inferior Colliculus Coefficients), ACW (Adaptive Component Weighted coefficients), LPCT (Linear Predictive Cosine Transform coefficients), LPCC (Linear Predictive Cepstral Coefficients), and their variants are also calculated (min, max, mean, median, mode, std). These features are sometimes extended by their first- and second-order regression coefficients (Δ and $\Delta\Delta$, respectively) [12, 14–16].

- Nonlinear features (NLF)

  1. Fractal features: Fractal dimension and its variants are also calculated (min, max, mean, median, mode, std) [4, 5, 14–16].
  2. Entropy features: Shannon entropy, multiscale permutation entropy, and their variants are also calculated (min, max, mean, median, mode, std) [4, 5, 14–16].

3.3 Automatic selection of features by Kolmogorov–Smirnov and Mann–Whitney U test

In this step, the best features are automatically selected taking into account medical criteria with regard to common significance level.

1. In a first step, the normality of the distributions is analyzed by means of the nonparametric Kolmogorov–Smirnov test [17].
2. The automatic feature selection is performed by means of Mann–Whitney U test because the distributions are

not normal distributions, being p value < 0.1 in order to obtain a larger set for the second phase of feature selection [18].

3.4 Automatic feature selection by WEKA

Afterward, a new selection phase is carried out in WEKA [20]:

1. In a first step, the feature selection algorithm *SVMAttributeEval* is used. This provides a selection by analyzing the integration of features in the group.
2. Then, for the experimentation, several feature sets with different feature numbers are created in order to develop a real-time system.

3.5 Normalization of features by WEKA

Moreover, during data preprocessing, all the features will be normalized by means of WEKA algorithms.

3.6 Automatic classification

In order to model the system, four classifiers will be used:

1. Support Vector Machines (SVM).

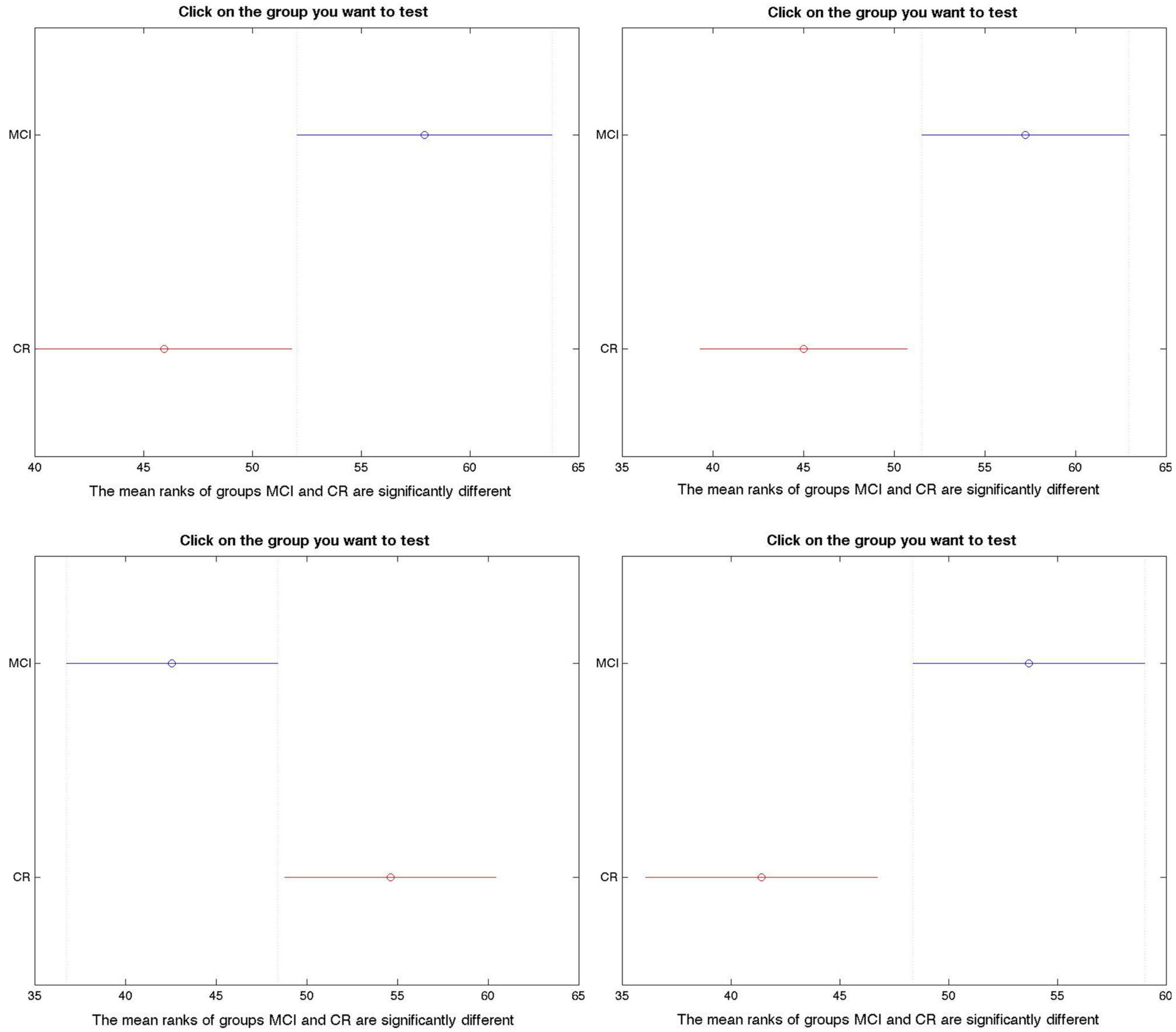


Fig. 2 Details of several results of the *Mann–Whitney U test* for the disfluencies and the unvoiced segments, in the *Animal Naming, Categorical Verbal Fluency (CVF)* task, for the MCI group and the CR group: (top-left) mean of unvoiced segments, (top-right) longest unvoiced segments, (bottom-left) Jitter ddp for disfluencies, (bottom-right) standard deviation for permutation entropy

2. k-nearest neighbors (*k*-NN)
3. Multilayer Perceptron (MLP) with *L* layers of *N* neurons, *Number of Neurons in Hidden Layers (*NNHL*).*
4. Convolutional Neural Network (CNN) with *L* layers of *N* neurons, a convolution mask of *cxc*, and a pool mask of *pxp*. We have used the WEKA software suite [16] in order to perform the experimentation.

3.7 System evaluation

For the evaluation of the system, three criteria will be used:

1. Classification Error Rate (CER in %) has been used in order to evaluate the results. We have used k-fold cross-validation with k = 10 for both training and validation [20, 21].
2. False Positive Rate has been used in order to avoid false diagnosis in the CR group where healthy people could undergo medical treatment unnecessarily, which would lead to misuse of medication.
3. Time Model: Time to build the model, oriented to real-time systems.

4 Results and discussion

4.1 Creation of feature sets

In the experiments, the used materials are about 60 speech samples for the control group (CR), and 40 speech samples for the MCI group, both belonging to the aforementioned PGA-OREKA (Table 1).

1. The signal is segmented into speech and disfluencies by a VAD algorithm with a minimum signal level.
2. Initially, the number of features obtained by the methodology described in Sect. 3 is about 920 (473 for speech and 447 for disfluencies) for a 22,000 kHz sampling frequency. The proposed set of features includes features from all the kinds of features described in Sect. 3.2 for speech and disfluencies.
3. Afterward, after a normalization test, an automatic feature selection is carried out using a nonparametric Mann–Whitney U test with p value < 0.1, and about 150 features are selected (Fig. 2).
4. In the second step of optimization, the attribute selection algorithm *SVMAttributeEval* of WEKA provides about 80 features.
5. Finally, several feature sets are created with the best 5, 10, 25, and 50 features, named P5, P10, P25, and P50, respectively.

4.2 Classifiers configuration

Several classifiers have been created using the criteria in Sect. 3.5. Table 2 shows the used configurations.

4.3 Experimentation

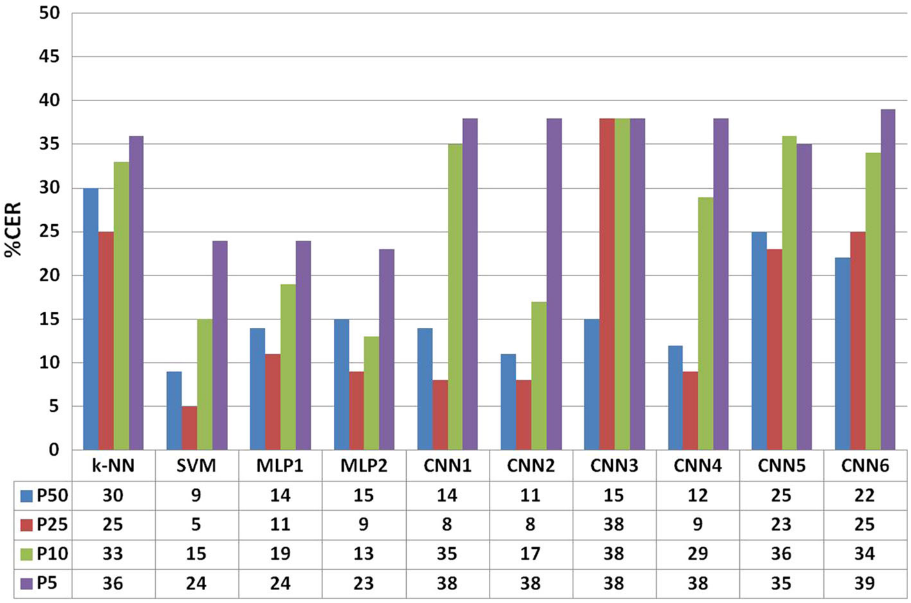
The models described in Table 2 have been evaluated by means of the 3.7 criteria. The results are stable, hopeful, good, and equilibrated for all of them.

1. Figure 3 shows the global CER results of the automatic classification for both the control group (CR) and the MCI group. CER (%) is evaluated for all the classifiers in Table 2. The new approach that integrates disfluency analysis outperforms previous works [4] for most of the classifiers. With the developed methodology, the results are in general very satisfactory for this simple task.
2. The best results are achieved with the 25 feature set, P25, and according to the rates, SVM can be considered the optimal solution. MLP2 and CNN for configurations 1, 2, and 4 obtain hopeful results with less computational load than classical MLP. In those cases, an average of 95 and 92% is achieved. As it can be seen in the selected parameters, this is due to the evaluation of features related to disfluencies and because the models are tested with important data that were not taken into consideration in the previous experiments; for example, the conversations of patients with themselves.
3. As it can be seen in Fig. 3, the results obtained in the task (AN) for MCI are very good, especially taking

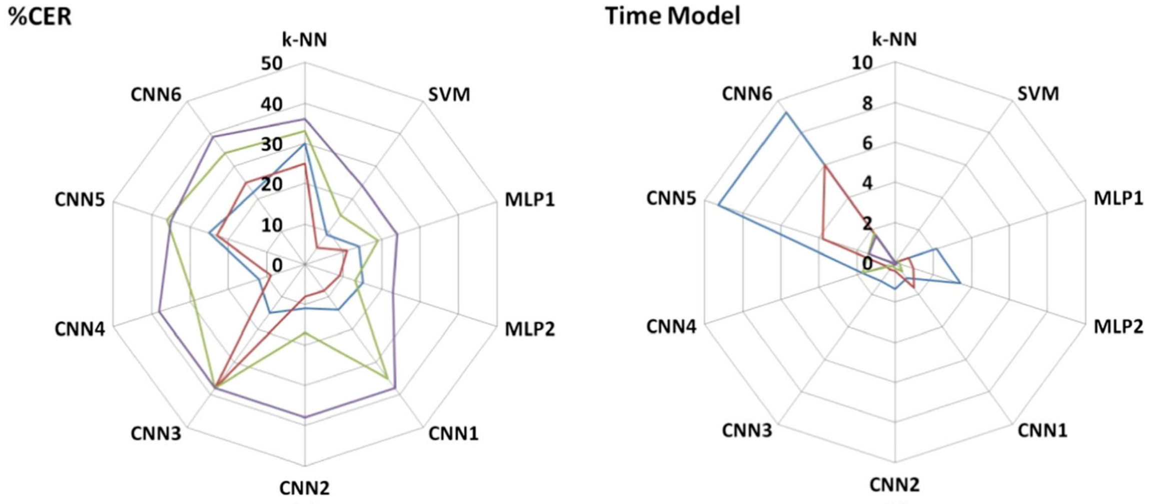
Table 2 Configuration of the proposed classifiers: kernel type; Number of Neurons in Hidden Layers (NNHL), /a/=(number of features+classes number)/2, /a, a/=2 layers with a NNHL; convolution mask (cxc); pooling mask (pxp); ID (Initial Dropout); HD (Hidden Dropout)

Model	Kernel	NL-NNHL	cxc	pxp	ID	HD
k-NN						
SVM	Polynomial					
MLP1		/a/				
MLP2		/a,a/				
CNN1		/a/			0.2	0.5
CNN2		/a/			0.2	0.2
CNN3		/a,a/			0.2	0.5
CNN4		/a,a/			0.2	0.2
CNN5		/a/	2 × 2	2 × 2	0.2	0.5
CNN6		/a/	2 × 2	2 × 2	0.2	0.2

Fig. 3 CER (%) for different feature sets and selected classifiers (Table 2): k-nearest neighbors (k-NN), Support Vector Machines (SVM), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN)



	k-NN	SVM	MLP1	MLP2	CNN1	CNN2	CNN3	CNN4	CNN5	CNN6
P50	30	9	14	15	14	11	15	12	25	22
P25	25	5	11	9	8	8	38	9	23	25
P10	33	15	19	13	35	17	38	29	36	34
P5	36	24	24	23	38	38	38	38	35	39



(a) (b)

Fig. 4 %CER (**a**) vs. time to build the model (**b**) for classifiers in Table 2

into account that they are modeled by 25 characteristics.

4. There is a clear improvement with regard to previous works due mainly to the improvement of the automatic feature selection and the integration of disfluency information.
5. Additionally, note that the specific weight of Thick Data could be important in this case by introducing the most noteworthy features into the algorithm, thus enriching information. This hybrid strategy of using both Thick Data and CNN could be a hopeful option for future real systems, even with small data sets.
6. The average between %CER and the time needed to build the models is shown in Fig. 4. SVM achieves good results in all tasks, especially in the significant 5% of CER. The fact that the data are well characterized is very helpful, and with CNN convolutional networks, an 8% CER is achieved. The Time Models are also optimum for these solutions.
7. Figure 5 shows the False Positive Rates for both groups, CR and MCI. These criteria are crucial in real health systems as medical criteria. In this case, SVM, MLP2, CNN1, and CNN2 appear as the best options.
8. Finally, the Objective Function (OF) with different weights for the system parameters is shown in Eq. 1 by medical criteria.

$$\mathrm{OF} = w1 * \mathrm{CER} - w2 * \mathrm{TM} - w3 * \mathrm{FP} \tag{1}$$

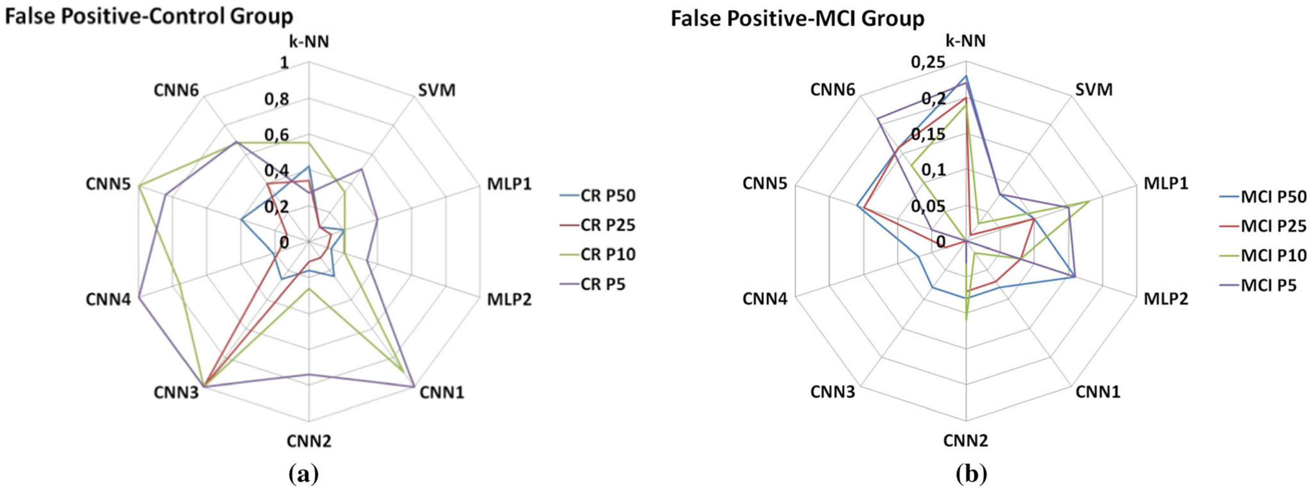


Fig. 5 False Positive (FP) for the control group (**a**) versus False Positive (FP) for the MCI group (**b**) for classifiers in Table 2

5 Conclusions

This paper presents a novel approach for the development of a real-time support system for the diagnosis of MCI. The system is based on automatic analysis of speech and disfluencies and Deep Learning modeling. Following the trend of Thick Data, the multifeature modeling is based on both automatic selection of the most relevant features by medical criteria and automatic selection of attributes over speech and disfluencies: Mann–Whitney U test, Support Vector Machine Attribute (SVM) evaluation, and Deep Learning approaches. The best approaches include deep learning by means of Convolutional Neural Networks (CNN) and SVM. The results are hopeful and lead to a new research line for the development of real-time health systems.

Acknowledgements This work has been supported by FEDER and MICINN, TEC2016-77,791-C4-2-R, and UPV/EHU-Basque Research Groups IT11156 and Basque Country EleKin Research Group.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest

References

1. World Alzheimer Report 2015. Available online: 2015-12-16, www.alz.co.uk/research/world-report-2015 (accessed on Mar 13rd 2018)
2. Laske C, Sohrabi HR, Frost SM, López-de-Ipiña K, Garrard P, Buscema M, Dauwels J, Soekadar SR, Mueller S, Linnemann C, Bridenbaugh SA, Kanagasingam Y, Martins RN, O'Bryant SE (2015) Innovative diagnostic tools for early detection of Alzheimer's disease. Alzheimer Dement 11(5):561–578. https://doi.org/10.1016/j.jalz.2014.06.004
3. Klimova B, Maresova P, Valis M, Hort J, Kuca K (2015) Alzheimer's disease and language impairments: social intervention and medical treatment. Clin Interv Aging Clin Interv Aging 2015(10):1401–1408. https://doi.org/10.2147/CIA.S89714
4. Lopez-de-Ipina K, Martinez-de-Lizarduy U, Barroso N, Ecay-Torres M, Martinez-Lage P, Torres F, Faundez-Zanuy M (2015) Automatic analysis of categorical verbal fluency for Mild Cognitive Impartment detection: A non-linear language independent approach. In: Bioinspired Intelligence (IWOBI), 2015 4th international work conference, pp 101–104, Donostia (Spain)
5. Lopez-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, Beitia B, Garcia-Melero J, Ecay-Torres M, Estanga A, Faundez-Zanuy M (2017) Analysis of disfluencies for automatic detection of Mild Cognitive Impartment: a deep learning approach. 1–4. https://doi.org/10.1109/iwobi.2017.7985526
6. Lopez-de-Ipina K, Alonso J-B, Manuel Travieso C, Sole-Casals J, Egiraun H, Faundez-Zanuy M et al (2013) On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. Sensors 2013(13):6730–6745
7. Dingemanse Mark, Torreira Francisco, Enfield NJ (2013) Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. PLoS ONE 8(11):e78273. https://doi.org/10.1371/journal.pone.0078273
8. Lezak MD, Howieson DB, Bigler ED, Tranel D (2012) Neuropsychological assessment, 5th edn. Oxford University Press, Oxford
9. Ruff RM, Light RH, Parker SB, Levin HS (1997) The psychological construct of word fluency. Brain Lang 57:394–405
10. CITA-Alzheimer Foundation, PGA project: http://www.cita-alzheimer.org/investigacion/proyectos, (accessed on June 10th 2016)
11. Gomez-Vilda P, Rodellar-Biarge V, Nieto-Lluis V, Munoz-Mulas C, Mazaira- Fernandez L, Martinez-Olalla R, Alvarez-Marquina A, Ramirez-Calvo C, Fernandez-Fernandez M (2013) Characterizing neurological disease from voice quality biomechanical analysis. Cogn Comput 5(4):399–425

12. Mekyska J et al (2015) Robust and complex approach of pathological speech signal analysis. Neurocomputing 167(2015):94–111
13. Gómez-Vilda P et al (2015) Phonation biomechanic analysis of Alzheimer's disease cases. Neurocomputing 167(1):83–93
14. Lopez-de-Ipina K, Martinez-de-Lizarduy U, Calvo PM, Mekyska J, Beitia B, Barroso N, Estanga A, Tainta M, Ecay-Torres M (2018) Advances on automatic speech analysis for early detection of Alzheimer disease: a non-linear multi-task approach. Curr Alzheimer Res 15(2):139–148
15. Meilan JJG, Martinez-Sanchez F, Carro J, Carcavilla N, Ivanova O (2018) Voice markers of lexical access in mild cognitive impairment and Alzheimer's disease. Curr Alzheimer Res 15(2):111–119
16. Lopez-de-Ipiña K, Satue-Villar A, Faundez-Zanuy M, Arreola V, Ortega O, Clavé P, Sanz P, Mekyska J, Calvo P (2016) Advances in a multimodal approach for dysphagia analysis based on automatic voice analysis. 54. 201–211. https://doi.org/10.1007/978-3-319-33747-0_20
17. MATLAB. www.mathworks.com, (accessed on Mar 13rd 2018)
18. SPSS. www.ibm.com, (accessed on Mar 13rd 2018)
19. Praat: doing Phonetics by Computer. www.praat.org/, (accessed on Mar 13rd 2018)
20. WEKA. http://www.cs.waikato.ac.nz/ml/weka, (accessed on Mar 13rd 2018)
21. Picard R, Cook D (1984) Cross-validation of regression models. J Am Stat Assoc 79(387):575–583