End-to-End Parkinson's Disease Detection Using a Deep Convolutional Recurrent Network

Cristian David Rios-Urrego[0000−0003−0174−1452] 1, Santiago Andres Moreno-Acevedo[0000−0001−7300−8562] 1, Elmar Nöth[0000−0002−3396−555X] 2, and Juan Rafael Orozco-Arroyave[0000−0002−8507−0782] 1,2

Faculty of Engineering, University of Antioquia UdeA, Medellín, Colombia
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
Corresponding author: cdavid.rios@udea.edu.co

Abstract. Deep Learning (DL) has enabled the development of accurate computational models to evaluate and monitor the neurological state of different disorders including Parkinson's Disease (PD). Although researchers have used different DL architectures including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) units, fully connected networks, combinations of them, and others, but few works have correctly analyzed and optimized the input size of the network and how the network processes the information. This study proposes the classification of patients suffering from PD vs. healthy subjects using a 1D CNN followed by an LSTM. We show how the network behaves when its input and the kernel size in different layers are modified. In addition, we evaluate how the network discriminates between PD patients and healthy controls based on several speech tasks. The fusion of tasks yielded the best results in the classification experiments and showed promising results when classifying patients in different stages of the disease, which suggests the introduced approach is suitable to monitor the disease progression.

Keywords: Parkinson's Disease · Speech Processing · Convolutional Neural Networks · Long Short-Term Memory.

1 Introduction

The automatic evaluation of pathological speech has captured the attention of the research community for many years. Among the benefits of using speech signals to diagnose and monitor different diseases are that it is non-invasive, and can be captured remotely at a very low cost. In the context of Parkinson's Disease (PD), different biomarkers have been studied for the development of computer-aided tools to support the diagnosis and monitoring of patients [11]. PD is a neurological disease characterized by resting tremor, rigidity, bradykinesia, postural instability, and other symptoms [9]. The disease is caused by a progressive loss of dopaminergic neurons in the substantia nigra of the brain [6]. Most PD patients develop speech deficits which are grouped and called hypokinetic dysarthria where the speech is characterized by monotone intensity, low pitch variability, and poor prosody that tends to fade at the end of the utterance [12,14].

Several speech tasks are typically performed with the aim to model different pathologies. The most common tasks are read text, isolated words, the rapid repetition

Petr Sojka, Aleš Horák, Ivan Kopeček and Karel Pala (Eds.): TSD 2022, LNAI 13502, pp. 308–319, 2022.
This is preprint prepared by Proceedings editor for Springer International Publishing Switzerland.

of diadochokinetic (DDK) tasks, sustained vowels, modulated vowels, and others. Each task brings different information that enables a better understanding of the pathology. Although there is significant progress in modeling pathological speech signals through classical methods mainly based on digital signal processing techniques, nowadays, Deep Learning (DL) has enabled different methodologies to speech traits processing in PD [17,3,1]. The main limitation of DL approaches is that it is typically considered as as *black-box* because their interpretability is very limited or null, therefore it is not possible to know what happens inside the model.

In the last years, DL techniques have been implemented to classify PD patients vs. Healthy Control (HC) subjects, achieving promising results in the automatic assessment of speech in PD patients. Different techniques and architectures have been used to analyze speech data including Convolutional Neural Networks (CNNs) [18,15,17], 1D convolutional layers [3,7], Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units [1,13], fully-connected networks [13,3,2], and combinations of them [8]. However, to the best of our knowledge, there are no studies about the interpretation and understanding of the configuration of the networks. Different DL models used to predict pathologies work with the raw data, therefore their hyper-parameters should have a meaning or interpretation depending on the studied phenomenon.

Motivated by the above mentioned, the main objective of this study is to present different experiments using several speech tasks to find the best network configuration that allows the discrimination of pathological speech traits. In order to address this objective, we have created an architecture composed of two 1D convolutional layers, 2 LSTM layers, and a fully-connected neural network. The architecture was trained to classify PD patients vs. HC subjects by varying the input size of the architecture. Once the best input size is found, the kernel size of the 1D convolutional layers is varied to find the best kernel configuration. Afterwards, given the best input and kernel sizes, the architecture was tested upon different tasks and combinations. In addition, we evaluated the model in a multi-class experiment where ranges of the modified Frenchay Dysarthria Assessment (m-FDA) scale [16] are considered as threshold to create different groups of speakers. This experiment allows to evaluate the suitability of the proposed approach to classify patients in different stages of the disease. Finally, with the aim to perform a more realistic evaluation of the proposed model, an independent test set with 20 PD patients and 20 HC subjects was considered.

The rest of the paper is as follows: Section 2 describes the corpora considered for this study. Section 3, presents the methods used in the study and the final architecture created to identify which task and which hyper-parameters settings yield better results. Section 4 shows the results of the study, and finally, Section 5 contains the conclusions and future work.

2 Data

2.1 PC-GITA

This corpus contains speech recordings of 50 PD patients and 50 HC subjects sampled at 44.1 kHz [10]. All participants are native speakers of Colombian Spanish and are balanced in age and gender. Each patient was in ON-state during the recording session, i.e., under the effect of their medication and was evaluated by an expert neurologist who labeled the patients according to the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS-III) scale [5]. Additionally, the dysarthria level of each participant (patients and healthy controls) was evaluated by three phoniatricians according to the m-FDA scale. The median value over the three labels was considered as the dysarhtria level of each participant. Further details can be found in [16]. The speech signals were down-sampled to 16 kHz to standardize the sampling rate with the independent test set presented below. Table 1 shows demographic and clinical information of the speakers. The subjects produced a total of 21 speech tasks, including: 10 sentences, a monologue, a read text, 24 isolated words, rapid repetition of 6 DDK tasks, sustained vowels, and modulated vowels.

Table 1. Demographic and clinical information of the participants. [F/M]: Female/Male. Time since diagnosis and age are given in years. Values are reported as mean ± standard deviation.

	PD patients	**HC subjects**	**Patients vs. Controls**
Gender [F/M]	25/25	25/25	$^{*}p$=1.00
Age [F/M]	60.7±7/61.3±11	61.4±7/60.5±12	$^{**}p$=0.98
Range of age [F/M]	49-75/33-81	49-76/31-86	
Time since diagnosis [F/M]	12.6±12/8.7±6		
MDS-UPDRS-III [F/M]	37.6±14/37.8±22		
Speech item (MDS-UPDRS-III) [F/M]	1.3±0.8/1.4±0.9		
m-FDA [F/M]	28.3±8.5/29.2±8.5	7.3±7/6.7±7.8	

$^{*}p$–value calculated through Chi–square test.
$^{**}p$–value calculated through t-test.

2.2 Independent test set

This corpus is formed with 20 PD patients and 20 HC subjects. The patients group consisted of 9 males and 11 females with ages between 29 an 83 years (mean = 61.3 ± 14.3). All of them were evaluated by a neurologist expert according to the MDS-UPDRS-III scale. The scores of such evaluations ranged between 9 and 106 (mean = 40.1 ± 22.7). The healthy group is formed with 11 males and 9 females with ages between 49 and 78 years (mean = 62.6 ± 10). None of the participants of this group had symptoms of neurological or movement disorders. All participants are also native speakers of Colombian Spanish and are independent of the PC-GITA database. Each patient was captured at a sampling frequency of 16 kHz and performed 3 tasks: (1) read text, (2) monologue, and (3) /pa-ta-ka/ DDK.

3 Methods

Figure 1 summarizes the architecture proposed in this work. It consists of two 1D convolutional layers followed by an LSTM network. Finally, a fully-connected layer is in charge of the classification between PD patients and HC subjects. In addition to this, max pooling layers were added to down-sample the information after each convolutional layer. Details of each stage are presented below.

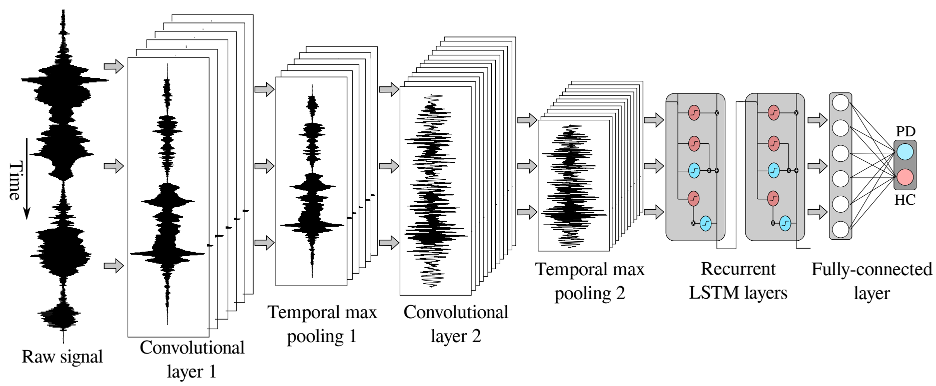

Fig. 1. Architecture proposed in this work to classify between PD patients and HC subjects.

3.1 1D Convolutional Layer

1D convolutional layer configurations are typically used to model sequential data, the main idea is to extract different representations based on the temporal domain. The layer consists of 2 main elements, number of channels and kernel. The kernel works as a filter upon the input data, it slides through the signal to extract information according to its weights and size. The layer has as many kernels as channels. The output data will have different representations from the same signal, as many as the number of channels. The weights of the kernel are learned during the training process and the size is given as a hyper-parameter. The output value of a convolutional layer with input size (N, C_{in}, L) and output (N, C_{out}, L) can be precisely described as Equation 1[1].

$$\text{out}(N_i, C_{out_j}, L) = \text{bias}(C_{out_j}, L) + \sum_{k=0}^{C_{in}-1} \text{weight}(C_{out_j}, k, L) \star \text{input}(N_i, k, L) \quad (1)$$

where $\star$ is the valid cross-correlation operator, N is a batch size, C denotes a number of channels, and L is the length of the signal sequence. Notice that in this type of layers, the kernel represents a temporal sliding window. Therefore, when a signal is represented by 16000 samples per second, a kernel size of 40 corresponds to an analysis window of 2.5 ms.

[1] https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html

3.2 Temporal max pooling

Temporal max pooling is a down-sampling technique used to reduce the temporal size of the data. The pooling depends on the kernel size, which indicates how many samples of the input vector should be reduced. The max pooling procedure creates a new vector obtained from the input data with less samples. The kernel slides through the original data taking as many samples as its size and calculating the maximum value of the samples. Such a maximum will be the number of samples in the new vector. The sliding step is equal to the length of the kernel. In this way, if a temporal max pooling with a kernel size of 2 is applied to a 1 sec signal with 16000 samples, the new vector will be temporally down-sampled representing the same 1 second of information with 8000 samples, where each sample is the maximum of a pair of consecutive values in the original data.

3.3 Recurrent LSTM layer

RNNs have been proposed to model sequential data, inside its architecture, it has a hidden state h_t that contains the information of the samples that have already passed through the network. These networks were improved when the storing process was introduced, resulting in the well-known LSTM networks. The LSTM is a cell that tries to remember sequential information for a longer time than the RNNs. The LSTM includes a status state that stores the long-term information, and also includes the following three new concepts: the input gate $\boldsymbol{I}$ aims to determine what new information should be added to the network status state. The forget gate $\boldsymbol{F}$ decides what information to keep for a long term and what information to forget from the status state. And finally, the output gate $\boldsymbol{S}$ decides the new hidden state as a combination of the previous hidden state, the new input, and the status state.

3.4 Network's topology

Figure 1 shows the architecture implemented in this work. It is composed of two 1D convolutional layers with 16 and 32 channels, respectively. Each layer is followed by a temporal max pooling with a kernel size of 2. Then, the characterization performed by the convolutional layers is the input to an LSTM that is responsible of performing the temporal analysis of the network. The recurrent network is composed of 2 LSTMs layers with 64 cells each. Finally, the output of the LSTMs feeds a fully-connected network to make the final decision. ReLu activations are considered in the convolutional layers, and a Softmax activation function is used at the output. For the training of the network, we used Pytorch with a cross-entropy loss function and an Adam optimizer. Batch normalization, dropout, and L2-regularization techniques are also used.

4 Experiments and results

Motivated to know the best configuration of the network concerning the input and kernel sizes to classify PD patients vs. HC subjects, we performed two experiments:

(1) we segmented the raw input waveform into different window sizes: 125ms, 250ms, 500ms, 1sec, 2sec, and 4sec. The size that yielded the best result was considered for the next experiment. (2) Given the best input size, we changed the kernel size in the convolutional layers from 20 to 640 which corresponds to analysis windows from 2.5ms to 40ms. Once for the best input and kernel size configuration were found, we performed the classification experiments. In addition, we included a multi-class classification experiment to evaluate the dysarthria level of the speakers according to the m-FDA scale. Finally, we evaluated the model over an independent test set with 20 PD patients and 20 HC subjects. All experiments (except the independent test) are performed following a speaker-independent 10-fold cross-validation strategy. The results are reported in terms of mean and standard deviation computed along the folds.

4.1 Parameters optimization

Input size: The raw signals were segmented into different sizes to observe the behavior of the network at different lengths and to conclude which input size gives the best performance. Each sample was down-sampled to 16 kHz and pre-processed by removing its DC level and normalizing its amplitude. The results obtained are shown in Table 2. Notice that the best configuration is obtained with 1 sec windows. This input size yielded an accuracy of 89.1% and 88.5% of F1-score. It is important to highlight that this result is balanced in sensitivity (86%) and specificity (91%).

Table 2. Classification of PD patients vs. HC subjects at different input sizes in the proposed architecture. Values are reported as mean ± standard deviation.

Input size	Accuracy(%)	Sensitivity(%)	Specificity(%)	F1-score(%)
125ms	83.6 ± 10.3	70.3 ± 19.6	94.3 ± 13.1	82.8 ± 10.9
250ms	85.1 ± 9.6	80.0 ± 23.0	91.3 ± 13.1	84.4 ± 10.4
500ms	86.4 ± 9.8	74.0 ± 23.2	96.0 ± 12.7	85.4 ± 10.9
1sec	**89.1 ± 9.3**	**86.0 ± 21.1**	**91.0 ± 15.5**	**88.5 ± 10.2**
2sec	87.3 ± 10.6	81.0 ± 23.3	96.3 ± 14.1	86.6 ± 11.4
4sec	83.6 ± 9.4	74.0 ± 23.1	90.7 ± 19.1	82.4 ± 10.6

A visual comparison was performed to analyze the behavior of each input size in terms of accuracy and also considering the Receiver Operating Characteristic (ROC) curves obtain in each case. The Area Under the ROC Curve (AUC) was also reported for each configuration. From Figure 2, we conclude that the input size of 1 second is the best choice, with an AUC of 0.87 and the maximum accuracy. It is worth noting that this input size not only yielded the highest accuracy, but also provided the best balance between sensitivity and specificity.

Kernel size: After obtaining the best configuration regarding the input size, we decided to evaluate different kernel sizes in the first and second convolutional layers, this allows determining the window of information that will be characterized by the network as a

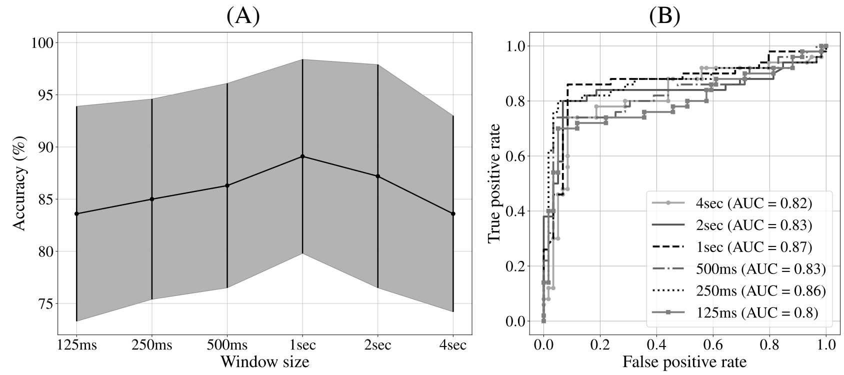

Fig. 2. (A). Mean accuracy and standard deviation (width of the gray stripe) with different input sizes. (B). ROC curves for different input sizes.

response to the input signal. In the first convolutional layer, we performed kernel length variations between 40 and 640, corresponding to analysis windows of 2.5ms and 40ms, respectively (for a sample rate of 16 kHz). For the second layer, the kernel size was changed between 20 and 320, corresponding to 2.5ms and 40ms for a sample rate of 8 kHz, this is due to the temporal max pooling layer, which reduces the sampling rate of the sequence. The results are reported in Table 3.

Table 3. Classification of PD patients vs. HC subjects at different kernel sizes. Input length is fixed at 1 second. Values are reported as mean ± standard deviation.

Kernel size layer 1	Kernel size layer 2	Accuracy(%)	Sensitivity(%)	Specificity(%)	F1-score(%)
40	20	89.1 ± 11.9	84.0 ± 22.7	92.6 ± 11.9	88.5 ± 12.7
80	40	89.1 ± 10.3	86.0 ± 21.2	91.0 ± 15.6	88.5 ± 11.1
160	**80**	**89.1 ± 9.3**	**86.0 ± 21.2**	**91.0 ± 15.6**	**88.5 ± 10.2**
320	160	80.8 ± 7.8	69.5 ± 21.4	89.0 ± 18.9	79.6 ± 8.9
640	320	80.6 ± 9.0	68.7 ± 19.1	91.3 ± 12.1	80.1 ± 9.7

The results show that a kernel size of 160 in the first convolutional layer and 80 in the second convolutional layer yield a classification accuracy of 89.1%. The corresponding AUC value is 0.87 (see Figure 3.B). It is also possible to observe that for the first 3 combinations of kernel sizes, the results are very similar. The main difference is the standard deviation which is computed along the 10 folds considering in the cross-validation stage.

Figure 3 shows the plot with the average accuracies and the corresponding standard deviation. It can be observed a kind of saturation bend in the first three values, i.e., from 40-20 to 160-80 kernel sizes. A comparison of the ROC curves for each experiment is shown on the right side of Figure 3. Notice that there are 2 groups of curves with similar trends. The group with larger areas corresponds to the solid lines with an AUC values of

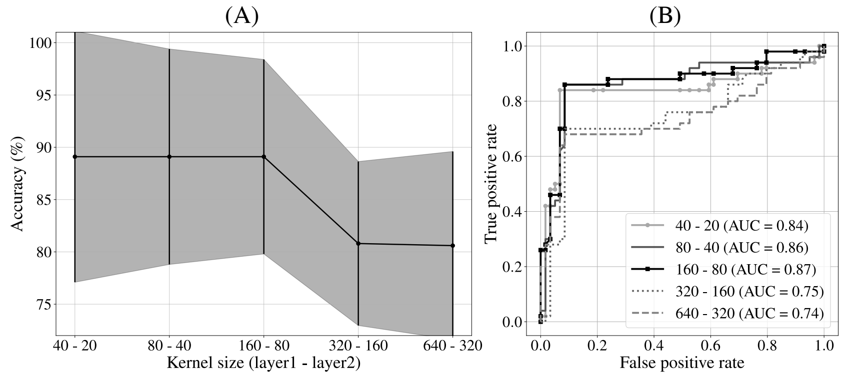

Fig. 3. (A). Mean accuracy and standard deviation (width of the gray stripe) with different kernel sizes. (B). ROC curves for different kernel sizes.

0.84, 0.86, and 0.87. And the group with smaller values correspond to the dashed lines with AUC of 0.75 and 0.74.

According to the results presented above, temporal and spectral in-deep analyses to find optimal parameters for the 1D CNN-LSTM architecture proposed in this study help in maximizing the classification performance to discriminate between PD patients and HC subjects.

4.2 Bi-class and multi-class classification

Previous experiments allowed to find the best configuration concerning input and kernel sizes. In this experiments, we want to validate how stable the network configuration is for different speech tasks. We also performed multi-class classification experiments to assess the dysarthria level of the participants according to the m-FDA scale [16].

Bi-class classification: The neural network is trained and evaluated by using different tasks performed by each participant, including: (1) the DDK task consisting in the repetition of the syllables /pa-ta-ka/, (2) read text, (3) monologue, and (4) the fusion of the 21 tasks mentioned in the Section 2. Table 4 contains the results of each experiment. Notice that the fusion of all tasks yields the best classification result with an accuracy of 89.1%, which is comparable with state of the art when the same database is used.

Figure 4.B shows the ROC curves and the corresponding AUC values obtained in each experiment. It can be observed that the fusion of tasks yields the highest AUC value. Notice also that the monologue provides a similar result. Figure 4.A illustrates the histogram and the probability density distribution obtained from the best result (the fusion of speech tasks). It can be observed that the error for the discrimination of HC subjects is small (i.e., specificity = 91%), while the discrimination of PD patients is larger (i.e., sensitivity = 86%).

Table 4. Classification of PD patients vs. HC subjects based on different speech tasks. Values are reported as mean ± standard deviation.

Task	Accuracy(%)	Sensitivity(%)	Specificity(%)	F1-score(%)
/pa-ta-ka/	65.8 ± 18.8	51.7 ± 25.5	79.0 ± 20.3	64.3 ± 19.6
Read text	77.9 ± 12.2	78.3 ± 17.6	76.5 ± 26.9	76.9 ± 13.3
Monologue	78.9 ± 14.4	86.3 ± 13.4	70.5 ± 26.1	78.1 ± 15.3
Fusion	**89.1 ± 9.4**	**86.0 ± 21.2**	**91.0 ± 15.6**	**88.5 ± 10.2**

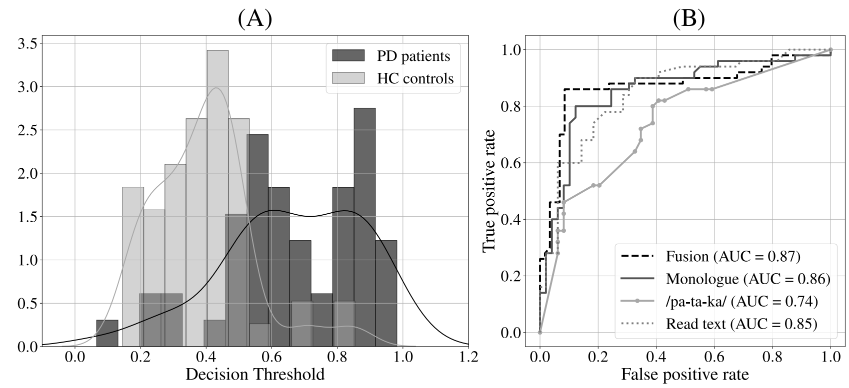

Fig. 4. (A). Histogram and the corresponding probability density distribution of the scores obtained from the classification of PD patients and HC subjects for the fusion of speech tasks. (B). ROC curves for several speech tasks performed by the participants.

Multi-class classification: Besides the bi-class classification, we evaluated the dysarthria level of the speakers according to the m-FDA scale, which is a modified version of the Frenchay dysarthria assessment scale [4]. The m-FDA evaluates several aspects of speech, including: respiration, lips movement, palate/velum movement, larynx, tongue, monotonicity, and intelligibility. The participants are divided into four groups according to their m-FDA score. The distribution of the groups is illustrated in Figure 5.A. Note that the white bars correspond to HC subjects while the others are for PD patients.

The result for the multi-class classification yielded an accuracy of 59.9 ± 6.7 and an F1-score of 55.7 ± 13.1. Figure 5.B shows the confusion matrix resulting from this experiment. Note that most speakers were correctly classified in the intermediate levels. Most of the errors occurred in groups 0 and 3, where the scores of the m-FDA are the smallest and the largest. To the best of our knowledge, this is one of the first works that includes an end-to-end architecture for the dysarthria level classification of PD patients.

4.3 Classification using the independent test set

This experiment is performed with the aim to evaluate the proposed approach in a more realistic scenario. This will provide a better impression regarding the generalization capability of the proposed models. We considered 20 independent PD patients and 20

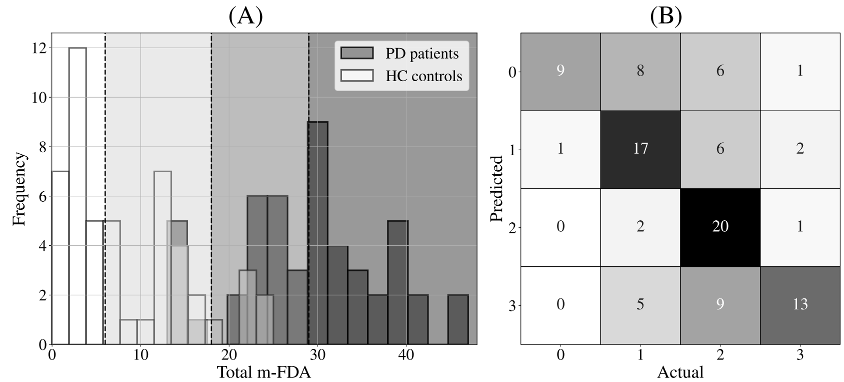

Fig. 5. (A). Distribution of the m-FDA scores for the participants of this study. The scores for the PD patients are grouped into three classes: low, intermediate, and severe according to the severity of the disease. The scores for the m-FDA scale also include HC subjects, represented with the white bars. (B). Confusion matrix for the multi-class classification.

HC subjects that were not included in the training and validation processes introduced before. Notice that this additional test set represents a group of subjects that arrived at the clinic and performed speech recordings to decide whether to continue with further tests to define their neurological condition.

The results are shown in Table 5. Note that the fusion of speech tasks yielded the best result with an accuracy of 77.5% and an F1-score of 75.3%. Furthermore, it can be observed that the same trend of the trained network is preserved with higher specificity (85%) than sensitivity (70%). These results confirm that it is possible to design deep neural network models to support the PD diagnosis and screening. We hypothesize that by adding more data in the training process, we could improve the stability and robustness of this classifier in order to obtain results closer to those reported with the cross-validation strategy (see Table 4).

Table 5. Classification of PD patients vs. HC subjects based on several tasks performed by the participants in the independent test set.

Task	**Accuracy(%)**	**Sensitivity(%)**	**Specificity(%)**	**F1-score(%)**
/pa-ta-ka/	62.5	52.0	73.0	61.6
Read text	73.0	78.0	68.0	72.9
Monologue	71.0	78.5	63.5	69.1
Fusion	**77.5**	**70.0**	**85.0**	**75.3**

5 Conclusions

This paper proposes an end-to-end architecture of an 1D CNN-LSTM network for the classification of patients suffering from PD vs. HC subjects. We evaluated and determined the best configuration of the proposed architecture for different input and kernel sizes in the convolutional layer. The results showed that the best configuration with respect to the input is obtained from windows of 1 sec. We noticed that a correct window analysis for the convolutional layers corresponds to a kernel size of 160 and 80 for the first and second layers, respectively. (i.e., windows of 10 ms for two layers). Then, we evaluated how the network classifies patients and controls based on several tasks performed by the participants. The fusion of speech tasks yielded the best results in the classification experiments with an accuracy of 89.1%. Motivated by this result, we included an experiment for the evaluation of the dysarthria level according to the m-FDA scale. We obtained an accuracy of up to 60%, and we could observe that the error in classification was mainly in the subjects with the smallest and largest m-FDA scores. Finally, we evaluated the robustness and generalization capability of the network with an independent test set of 20 PD patients and 20 HC subjects. For this last experiment, we observed that the fusion of speech tasks yielded also the best result, with an accuracy of 77.5%, which demonstrates that despite the small amount of data, it is possible to generate deep neural network models for the automatic diagnosis of PD.

Future experiments will include the training of architectures with larger amounts of data, as well as other pathologies in different languages. In addition, we will also address experiments for the interpretation of the hidden states of RNNs, preliminary results showed that in some cases, the hidden states of the network carry the same behavior as the fundamental frequency of the participant.

Acknowledgements This work was financed by CODI from University of Antioquia grant # 2017–15530. This project received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement # 766287.

References

1. Arias-Vergara, T., et al.: Automatic detection of voice onset time in voiceless plosives using gated recurrent units. Digital Signal Processing **104**, 102779 (2020)
2. Caliskan, A., et al.: Diagnosis of the parkinson disease by using deep neural network classifier. IU-Journal of Electrical & Electronics Engineering **17**(2), 3311–3318 (2017)
3. El Maachi, I., Bilodeau, G.A., Bouachir, W.: Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait. Expert Systems with Applications **143**, 113075 (2020)
4. Enderby, P.: Frenchay dysarthria assessment. British Journal of Disorders of Communication **15**(3), 165–173 (1980)
5. Goetz, C.G., et al.: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Movement disorders **23**(15), 2129–2170 (2008)

6. Jankovic, J.: Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry **79**(4), 368–376 (2008)
7. Kim, H., et al.: Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. Journal of Clinical Medicine **9**(11), 3415 (2020)
8. Mallela, J., et al.: Voice based classification of patients with amyotrophic lateral sclerosis, parkinson's disease and healthy controls with cnn-lstm using transfer learning. In: Proceedings of ICASSP. pp. 6784–6788. IEEE (2020)
9. McKinlay, A., et al.: A profile of neuropsychiatric problems and their relationship to quality of life for parkinson's disease patients without dementia. Parkinsonism & related disorders **14**(1), 37–42 (2008)
10. Orozco-Arroyave, J.R., et al.: New spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of LREC. pp. 342–347 (2014)
11. Orozco-Arroyave, J.R., et al.: Apkinson: the smartphone application for telemonitoring parkinson's patients through speech, gait and hands movement. Neurodegenerative Disease Management **10**(3), 137–157 (2020)
12. Pinto, S., et al.: Treatments for dysarthria in parkinson's disease. The Lancet Neurology **3**(9), 547–556 (2004)
13. Rizvi, D.R., et al.: An lstm based deep learning model for voice-based detection of parkinson's disease. Int. J. Adv. Sci. Technol **29**(8) (2020)
14. Spencer, K.A., Rogers, M.A.: Speech motor programming in hypokinetic and ataxic dysarthria. Brain and Language **94**(3), 347–366 (2005)
15. Trinh, N.H., O'Brien, D.: Pathological speech classification using a convolutional neural network (2019)
16. Vásquez-Correa, J.C., et al.: Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease. Journal of communication disorders **76**, 21–36 (2018)
17. Vavrek, L., et al.: Deep convolutional neural network for detection of pathological speech. In: Proceedings of SAMI. pp. 000245–000250. IEEE (2021)
18. Wodzinski, M., et al.: Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In: Proceedings of EMBC. pp. 717–720. IEEE (2019)