Comparative analysis of multiple speech tasks to recognise Parkinson's disease using pre-trained feature extractor embeddings

Attila Zoltán Jenei[1], Zalán Valálik[2], Dávid Sztahó[1]

[1] Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Telecommunication and Media Informatics
{jenei.attila.zoltan,sztaho.david}@vik.bme.hu

[2] NeuroMed
{zalanvalalik}@gmail.com

Abstract. Parkinson's disease is one of the most common neurological diseases, which is currently incurable. Speech can be a suitable biomarker for supporting the diagnosis of the disease. Therefore, many speech tasks and recording lengths are used widely in the literature. Our research compares the recognition performance measured on seven speech tasks using pre-trained out-of-domain feature extraction algorithms (x-vector, e-capa). We also examine how the voting on the speech tasks relates to the performance on the given speech tasks and which speech tasks the classifier considers essential. Our results showed that using a longer speech signal provides better recognition, and the type of the task is essential, e.g. pronouncing syllables. Furthermore, the classifier considers longer speech tasks more critical in the decision and suggests letting out sustained vowels.

Keywords: Parkinson's disease, x-vector, e-capa, voting ensemble, machine learning, feature importance

1 Introduction

Parkinson's disease (PD) is one of the most common neurological disorders nowadays, typically appearing in ageing societies. According to surveys made in 2016 Feigin et al. (2019), its incidence can be estimated at 6.1 million people worldwide. However, their number is increasing, presumably due to the ageing of society, industrialisation and environmental pollution. This fact is strengthened by that the diagnostic procedure for Parkinson's disease has not changed to such an extent recently (Bloem et al., 2021).

Early detection of the disease is essential for the patient, as the appropriate treatment and therapy can significantly slow down the progress of the disease. It is critical because the disease cannot be cured according to current knowledge

(Balestrino and Schapira, 2020). As a result, lifelong treatment and monitoring are necessary for the patient to maintain a high-quality life.

The onset of the disease presumably begins in the peripheral autonomic nervous system, then it continues to spread to the central nervous system, reaching the lower brain stem before the substantia nigra area (Katzenschlager et al., 2008). This information supports why the deterioration of the sense of smell, constipation and rapid eye movement sleep disorders occur at the beginning of the disease. The disease is described by the degeneration of the dopamine-producing neurons, accompanied by the loss of their axons extending into the striatum along the nigrostriatal pathway. Therefore, a decline in many cognitive and motor functions can be experienced (MacMahon Copas et al., 2021).

The diagnosis of the disease is mostly based on motor symptoms. However, this is usually preceded by the appearance of non-motor symptoms, even ten years before the diagnosis. This can vary from patient to patient, but the most common are decreased emotional involvement and interest, sleeping disorders, and constipation (Sveinbjornsdottir, 2016). Depression and anxiety can also appear in the pre-motor period (Pont-Sunyer et al., 2015).

The most common motor symptoms of the disease are tremors, rigidity, akinesia and/or bradykinesia (Moustafa et al., 2016). In addition to these, inadequate coordination of posture and freezing of gait appear. These symptoms also influence handwriting, which can be observed from a decrease in speed and a deterioration in quality. The symptoms become visible not only in the fine movements of handwriting but also in the formation of speech.

In the literature, many researchers report on speech-based diagnosis of the disease. Several speech tasks appear in these studies, such as sustained sounds, syllables, sentences or even spontaneous speech. However, there needs to be more agreement on how long a recording is sufficient to detect the disease. The question should be investigated all the more because feature extraction algorithms based on deep learning are becoming more and more common.

The present research examines which of several speech tasks recorded from a single person provides the best recognition and whether combining the predictions based on different speech tasks can improve recognition. The tests were performed on 7 speech tasks with pre-trained out-of-domain feature extraction algorithms and support vector machine (SVM) classifiers.

In the *Related Literature*, we present the existing results related to the topic; in the *Methodology* section, we describe the database and the experimental layout; in the *Results* section we describe the results obtained during the experiment, and finally in section *Summary and Conclusion*, we summarise the key findings.

2 Related Literature

Dysphonic speech appears in 70-90% of patients with Parkinson's disease, which can be manifested in a decrease in volume, a slower pace of speech and stuttering (Defazio et al., 2016). As the symptoms worsen, the intelligibility of speech

also decreases. Several speech tasks are commonly used in speech-based tests. These can consist of the pronunciation of sustained sounds, words, sentences or even spontaneous speech. While the sustained voice mostly provides information about the vocal cords, spontaneous speech already provides information about the process of complex speech formation (phoneme transitions, respiratory limitations, disturbances at the articulatory level) (Amato et al., 2021).

Vadovský and Paralič (Vadovský and Paralič, 2017) investigated the detection of Parkinson's disease using sustained sounds (/a/, /u/, /o/), words, numbers and sentences. 20 healthy control (HC) and 20 PD individuals were examined through manual features using several classification algorithms (C4.5, C5.0, CART and RandomForest). Their results ranged from 50.6% to 65.9% accuracy. The best result (65.9% accuracy) was achieved with the classifier scores being averaged with 5-fold cross-validation.

Sakar and her colleagues (Sakar et al., 2013) also investigated how predictive various speech tasks are regarding PD. They used the same database as *Vadovský and Paralic*. In their research, they used many characteristic features that describe a speech signal, such as frequency, pulse, amplitude, voicing, pitch and harmonicity variables. SVM and k-nearest neighbour (k-NN) were used for classification with multiple parameter settings. The best results were achieved with the SVM model for the sustained /o/ sound (72.5% accuracy) and the number four (75.0% accuracy). No significantly better results were found than when all recordings were used together.

Sztahó and his colleagues (Sztahó et al., 2019) examined the recordings of 55 PD and 33 HC. The subjects recited the sustained sound /a/, repetitive syllables (Diadochokinesis - DDK), words, sentences, a tale and a monologue. They examined the recognition performance as regards various speech tasks through several speech features. Furthermore, an attempt was made to use speech tasks together in a joint decision system. The best result was achieved with an SVM model using a radial basis function. Sustained sounds led to 77.0% accuracy, while using a monologue led to 89.3% accuracy, as two extreme points. With the joint decision, in most cases, a few percent better results were achieved.

Deep learning-based feature extraction algorithms have also become widespread recently, such as x-vector (a deep learning-based version of i-vector). *Moro-Velazquez and colleagues* studied 43 PD and 46 HC individuals (Moro-Velazquez et al., 2020). Their best results were achieved with x-vector feature extraction (compared to i-vector) on text-dependent utterances as the speech task (90% accuracy). The results of *Laetitia Jeancolas and her colleagues* showed that the x-vector technology provides a better classification performance than the traditional Mel-Frequency Cepstral Coefficients - Gaussian Mixture Model (MFCC-GMM) solution (Jeancolas et al., 2021).

Overall, it can be seen that many speech tasks are prevalent in the literature for the recognition of Parkinson's disease. However, there is less agreement about which one provides the best recognition since each speech task aims to capture different information. Furthermore, the spread of already pre-trained feature extraction algorithms can provide an opportunity for fully automatic (even

cross-language) recognition. All the more so because in the healthcare field, it is difficult to obtain enough samples to train one's own automatic deep learning feature extractor.

3 Methodology

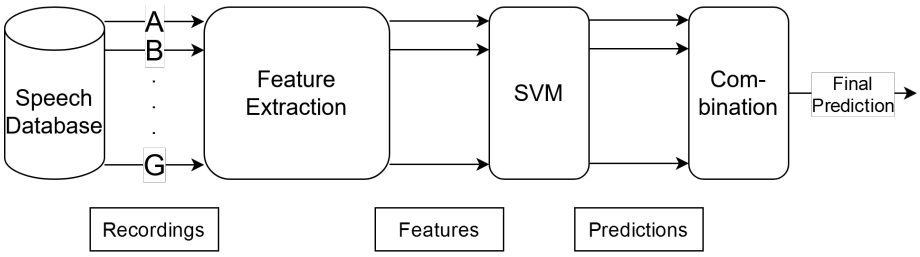

Fig. 1: Examination process: database with multiple samples, feature extraction, classification, voting.

The examination process is illustrated in Figure 1. The speech database includes several recordings (marked A - G) from one individual. From these, features were determined using deep learning-based feature extraction algorithms (x-vector and e-capa). Then, classification was performed on each speech task separately. Finally, multiple voting approaches were applied to the predictions obtained from each speech task.

3.1 Speech Database

As regards the speech database, 39 PD and 39 HC people were selected for examination (where all speech tasks [mentioned below] were available). 18 males and 21 females were in both classes, with a median age of 68 years. The PD patients were recorded in two institutes: Semmelweis and Virányos Clinic. The HC people were recorded at the Budapest University of Technology and Economics. The severity of the PD was measured on the Modified Hoehn and Yahr (H&Y) scale, where 1 means minimal or no functional disability (unilateral) and 5 means confinement to bed or wheelchair (Goetz et al., 2004). The severity distribution of the patients can be seen in Figure 2. The average H&Y score is 2.6, with a 1.2 standard deviation.

Seven speech tasks were examined: **A** - sustained /a/, **B** - sustained /a/ with pitch increase, **C** - syllables (/pa/-/ta/-/ka/), **D** - word (/agárral/), **E** - sentence (/Karcsi eltörte a lábát, amikor kerékpározott/), **F** - The North Wind and the Sun tale (bound text), **G** - free monologue. The recordings were made using a

clip-on microphone and an external audio device. The samples were stored at a 44.1 kHz sampling frequency and 16-bit quantisation.

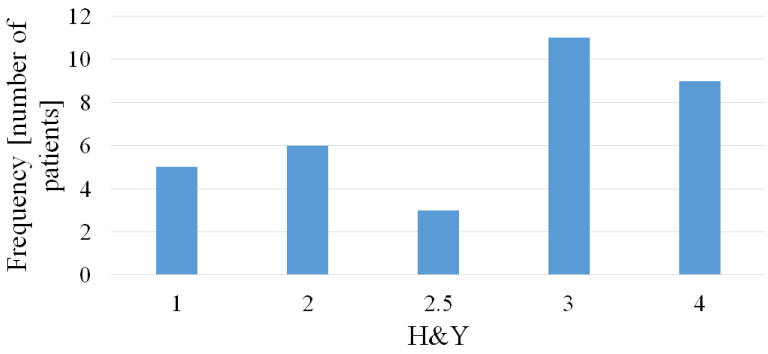

Fig. 2: Severity distribution of the patients according to their H&Y.

3.2 Preprocessing and Feature Extraction

The speech recordings were resampled at 16 kHz sampling frequency and normalised to -20 decibels relative to full scale (dB FS) according to loudness.

The *x-vector* technology is a feature extraction method related to speaker recognition tasks. This system consists of a feed-forward deep neural network, which maps speech segments of variable length into embeddings called x-vectors. The system divides the speech segment into frames and then examines the temporal environment of the frames in the first five layers. The subsequent (statistical pooling) layer aggregates the information over the entire segment. The output of this layer is the mean and standard deviation of its input. Finally, the dimensionality of these features is reduced in the segment-level layers. The network is trained for a speaker recognition task (using speaker IDs as targets), and the activation of one of the last layers is used as embedding vectors ('x-vector'). Its output dimension is set to 512 (number of features) (Snyder et al., 2018).

The *ECAPA-TDNN (e-capa)* system is an improved version of the previously mentioned x-vector system. One of its modifications performs an extended temporal attention to channel levels. Furthermore, the authors extended the frame-level features with 1-dimensional Squeeze-Excitation (SE) blocks. This makes the global context available in the attention module. Finally, the final feature vector is built from the concatenation of all SE-Res2Blocks instead of using only the last layer. The resulting feature vector has an output dimension of 192 (Desplanques et al., 2020).

The SpeechBrain toolkit (v.0.5.13) was used to implement the feature extractors, where the modules were previously trained on the Voxceleb database (Voxceleb 1+ Voxceleb2 training data) (Ravanelli et al., 2021). The present feature extraction algorithms are not specific for Parkinson's disease, but their use

is motivated by the fact that they are capable of high-dimensional representation and discrimination.

3.3 Classification and Classifier Fusion

The *Support Vector Machine (SVM)* is a supervised non-probabilistic classification algorithm that maps the input features into one of the classes. During training, the algorithm adjusts the parameters of the separating line (the boundary between the classes) in such a way that the distances from the class elements and the line (or hyperplane) are maximal (thus maximising the margin between the two categories). Test examples are then mapped into the same feature space and predicted into one of the categories. In the experiments, a linear SVM was used with default settings ($C = 1.0$) provided by the scikit-learn python package (*sklearn.svm.SVC class*) (Buitinck et al., 2013). We have chosen this kernel because feature importance can also be extracted from it.

Voting aggregation (classifier fusion) can be used when the predictions of multiple inputs or classifiers are available. The central aspect of the voting ensemble is to provide a more reliable outcome for solving a given problem (Brownlee, 2021). There are two major approaches to aggregate predictions by voting: 1) *soft voting* - predict final probabilities by averaging the initial probabilities, 2) *hard voting* - predicting the class with the most common label from models.

During experiments, soft, hard and linear SVM voting were probed. First, classification was done separately for each speech task, and the prediction (from 0 to 1) was stored. This prediction measured the probability of the sample belonging to the positive (PD) class. From all speech tasks, the predictions were gathered together per speaker. By combination via a linear SVM model, we mean that a further linear SVM was trained where the input features were the predictions from the above-mentioned speech tasks.

3.4 Model evaluation

The experiments were performed using *Leave-One-Out Cross-Validation (LOO-CV)* procedure. In this case, one sample is drawn as a test element, and the other samples in the database are used for training. This splitting, training and evaluation took place until every database element had been a testing element. Finally, we aggregated the results for the whole database at the end of the LOOCV.

The *sensitivity, specificity, balanced accuracy and area under the ROC Curve (AUC)* metric were used to evaluate the models. In addition, Receiver Operating Characteristic curves (ROC) were plotted for both feature extraction cases. The ROC curves show the development of the false positive rate (1-specificity) and the true positive rate (sensitivity) along different decision limits. The AUC value gives the size of the area under the ROC curves.

Finally, we used *t-distributed stochastic neighbour embedding (t-SNE)* to visualise the two classes (PD, HC) based on the predictions of the speech tasks

(Van der Maaten and Hinton, 2008). This makes it possible to observe the samples in the two-dimensional space created from the features. Our study examined the distribution of the original labels and the gender distribution via t-SNE mapping.

The experimental process and evaluation were implemented in a Python (v.3.7.16) environment. For feature extraction, the pytorch (v.1.12.1) framework was used (with the SpeechBrain toolkit).

4 Results

The results are summarized in Table 1. The table can be divided into two major parts vertically: results with e-capa feature extractor and results with x-vector feature extractor. The second column contains the experimental cases where *A-G* are speech tasks, while *soft*, *hard* and *SVM* refer to the voting approaches. The last four columns contain the sensitivity (sens), specificity (spec), balanced accuracy (b. acc) and AUC (auc) values. The best performances are marked with **bold** per feature extractor.

Table 1. Experimental results with multiple speech tasks using e-capa and x-vector feature extractors.

	speech and voting tasks	sens	spec	b. acc	auc
e-capa	sustained vowel (A)	43.6%	46.2%	44.9%	0.467
	(A) with p. increase (B)	56.4%	59.0%	57.7%	0.618
	syllable (C)	74.4%	71.8%	73.1%	0.738
	word (D)	56.4%	51.3%	53.8%	0.526
	sentence (E)	71.8%	82.1%	76.9%	0.805
	bound text (F)	**79.5%**	**84.6%**	**82.1%**	**0.902**
	monologue (G)	74.4%	76.9%	75.6%	0.859
	soft	74.4%	84.6%	79.5%	0.895
	hard	74.4%	76.9%	75.6%	0.756
	SVM	79.5%	79.5%	79.5%	0.874
x-vector	sustained vowel (A)	56.4%	66.7%	61.5%	0.669
	(A) with p. increase (B)	69.2%	69.2%	69.2%	0.636
	syllable (C)	84.6%	76.9%	80.8%	0.911
	word (D)	56.4%	59.0%	57.7%	0.636
	sentence (E)	79.5%	71.8%	75.6%	0.815
	bound text (F)	89.7%	94.9%	92.3%	0.980
	monologue (G)	89.7%	89.7%	89.7%	0.934
	soft	87.2%	94.9%	91.0%	0.960
	hard	89.7%	89.7%	89.7%	0.897
	SVM	**92.3%**	**97.4%**	**94.9%**	**0.985**

The results show that shorter speech tasks containing fewer phonemes have a lower recognition performance. Using sustained vowels, a random decision (around 50% b. acc) can be seen with the e-capa, while the x-vector had a

61.5% b. acc. In the case of recordings containing longer and more varied text, a higher performance can be seen. The monologue, for example, has 75.6% b. acc value for e-capa, and 89.7% b. acc for the x-vector approach.

It can also be observed that the specifically formed syllable (also known as the DDK test) has better results (73.1% [e-capa], 80.8% [x-vector] b. acc) than a more general word test (53.8% [e-capa], 57.7% [x-vector] b. acc).

It can also be seen that reading the bound text (F task) with both feature extraction algorithms provided almost the best results. In the case of e-capa, 82.1% b. acc was achieved while using x-vector features, 92.3 % b. acc value was experienced. In comparison, the monologue had 75.6 % and 89.7% b. acc respectively for e-capa and x-vector. The difference in the results of the two speech tasks is the correct classification of 5 and 2 persons, respectively.

Finally, the voting results on the predictions with e-capa features did not obtain better results than using only one speech task (Task F). From the three voting approaches, the soft and SVM votings approached the performance of the F task. In the case of x-vector, however, SVM-based voting achieved the top result (94.9% b. acc). The difference compared to the F task is 5.2% in b. acc (two correctly recognised people).

Figure 3 presents the ROC curves of the experiments. The results obtained with x-vector features are shown on the left, and the results obtained with e-capa are shown on the right side. The colour coding of the curves shows the speech tasks and voting cases. The two figures are consistent: tasks A, B and D performed poorly (small area under the curve), while for task F, soft, and SVM votings show a large area under the curve. The auc value, according to Table 1, might be used for a compact representation of the ROC curves.

The importance of the features (speech tasks) extracted from the voting SVM model can be seen in Figure 4. The x-vector feature extractor was used on the left side, the e-capa was used on the right side. The bound text (task F) and syllable (task C) tasks were the most important in the model's prediction in both cases. In both models, the monologue (task G) also appears as a valuable task.

When using x-vector features, however, the two forms of the sustained sound (tasks A and B) and the spoken word (task D) tasks proved confusing. In the case of e-capa, the model also used word pronunciation as a beneficial task, but the sustained sound task appeared with a negative importance.

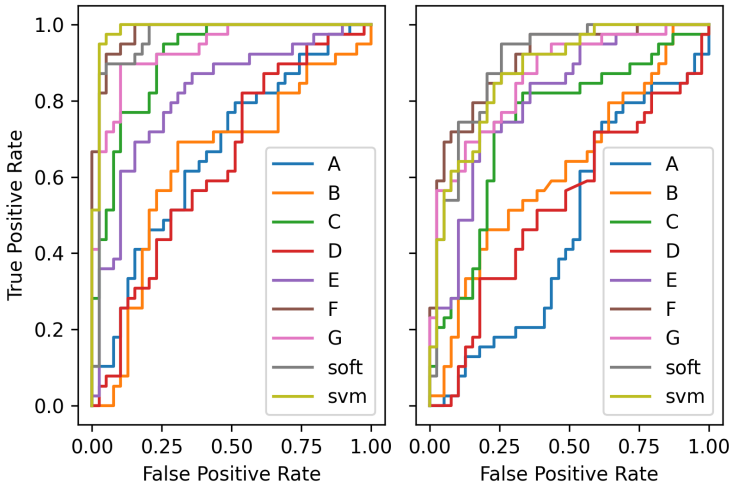

Fig. 3: ROC curves of the experiments. The left figure illustrates the x-vector, the right figure the e-capa results.

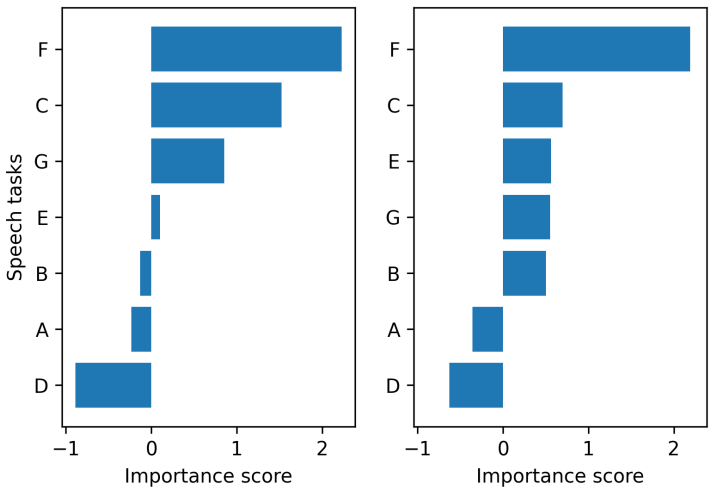

Fig. 4: The importance of each speech task based on the voting SVM model. On the left side, x-vector was used for feature extraction, on the right side, the e-capa.

Figure 5 illustrates the t-SNE maps applied on the database using the voting SVM model using speech task-based prediction of PD with x-vector feature extraction. This approach was selected for t-SNE because it reached superior performance. On the left, the data points have been coloured based on the original label (1 - PD, 0 - HC). It can be seen that the subjects can be well categorised using their predictions based on speech tasks (shrunk to 2 dimensions) since the colours are separated. Furthermore, the distribution of genders is shown on the right side of the figure. It can be seen that the distribution of men and women is homogeneous in both classes. Thus, the gender distribution did not cause bias in the classification, and the system probably used the property of disease exposure. The result is as expected since we strived to eliminate bias during the database assembly.

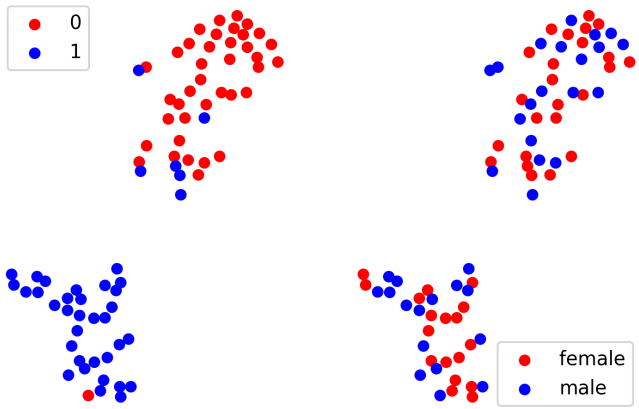

Fig. 5: t-SNE maps of the subjects (speech tasks predictions) using x-vector. The left figure is coloured with the original label, while the right figure is coloured according to gender.

5 Discussion and Conclusion

PD is one of the most common neurological diseases in the elderly population, and it is incurable according to current clinical knowledge. It is essential to recognise the disease in the early stage to reduce its progression and maintain appropriate therapy. Speech-based research is an intensive field, providing non-invasive support for diagnosing the disease. In addition, other modalities, such as the examination of drawing or movement data, are also widespread.

Speech is a valuable biomarker, as a speech segment of a few seconds may be sufficient to recognise PD. That is why many speech tasks and recordings of varying lengths are applied in the literature studies. The different speech tasks have their test objectives. However, the overall question is how well these can be used for the recognition of the disease.

In this study, we examined seven speech tasks from the same people, looking for the one that best supports the recognition of PD among the seemingly healthy population. In addition, we attempted to combine the predictors trained on various speech tasks using several approaches (*soft, hard, SVM voting*).

The speech recordings were preprocessed, and then the features were determined using pre-trained out-of-domain feature extraction algorithms (*x-vector, e-capa*). With the help of these features, we performed classification with linear-kernel SVM models.

The results showed that x-vector feature extraction could perform better when using the same classifier than e-capa. In speaker recognition, e-capa is considered an improved version of the x-vector; still, the latter performed better in our case. Presumably, the disease had a significant influence on the extracted features. All the more so since no fine-tuning was applied to the embeddings. From the feature importance plot (Figure 4), it also can be seen that the e-capa tries to use more speech tasks, but its performance is still lagging behind that of the x-vector representation.

In addition, there was an agreement between the two feature extractors that the bound text recordings (task F) produced the best performance among the speech tasks. Presumably, using the same text in the two classes helped the classifier to identify the target difference (disease exposition). Furthermore, these feature extraction algorithms were trained on more extended recordings, so it can be expected that they can extract less essential information from shorter texts (like sustained vowels).

The results also showed that the syllabic task performed better than the word or sentence tasks. This is also proved in the literature, as the syllabic task requires fine articulation precision and the ability to quickly change articulators between two consecutive segments (Godino-Llorente et al., 2017). Our results are consistent with this literature's findings. However, it probably depends also on the complexity of the chosen word (D task).

Among the voting approaches, linear SVM using x-vector features proved outstanding. In the case of e-capa, the votings provided a performance decrease by some percentage compared to the result of task F. Examining the importance of the speech tasks, it was seen that the bounded text (task F), the syllable (task C), the sentence (task E) and the monologue (task G) tasks proved to be helpful with both feature sets. Opposite, the words (task B) and sustained sounds (task A and B) tasks did not contribute positively to the classification. This is consistent with the literature since the syllable task was designed directly to capture the symptoms of the disease, and the bound text also contains a variety of word and phoneme relationships. The latter also needs to be read aloud at an appropriate pace.

Finally, we also performed t-SNE mapping to check whether the model used the presence of the disease and whether it did not include, for example, a gender bias. By examining Figure 5, it became visible that the groups can be separated based on the original label and that the gender distribution is homogeneous along the two categories. This was as expected since we aimed to filter out disturbing factors when assembling the database. This resulted in a set of 39 PD and 39 HC subjects, which may be a limitation of the results.

Acknowledgment

This work was partly funded by project no. K143075, which has been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the K_22 funding scheme.

References

Amato, F., Borzì, L., Olmo, G., Orozco-Arroyave, J.R.: An algorithm for parkinson's disease speech classification based on isolated words analysis. Health Information Science and Systems 9, 1–15 (2021)

Balestrino, R., Schapira, A.: Parkinson disease. European journal of neurology 27(1), 27–42 (2020)

Bloem, B.R., Okun, M.S., Klein, C.: Parkinson's disease. The Lancet 397(10291), 2284–2303 (2021)

Brownlee, J.: How to develop voting ensembles with python (Apr 2021), https://machinelearningmastery.com/voting-ensembles-with-python/

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122 (2013)

Defazio, G., Guerrieri, M., Liuzzi, D., Gigante, A.F., Di Nicola, V.: Assessment of voice and speech symptoms in early parkinson's disease by the robertson dysarthria profile. Neurological Sciences 37, 443–449 (2016)

Desplanques, B., Thienpondt, J., Demuynck, K.: Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification. arXiv preprint arXiv:2005.07143 (2020)

Feigin, V.L., Nichols, E., Alam, T., Bannick, M.S., Beghi, E., Blake, N., Culpepper, W.J., Dorsey, E.R., Elbaz, A., Ellenbogen, R.G., et al.: Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. The Lancet Neurology 18(5), 459–480 (2019)

Godino-Llorente, J., Shattuck-Hufnagel, S., Choi, J., Moro-Velázquez, L., Gómez-García, J.: Towards the identification of idiopathic parkinson's disease from the speech. new articulatory kinetic biomarkers. PloS one 12(12), e0189583 (2017)

Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R.G., Moore, C.G., Wenning, G.K., et al.: Movement disorder society task force report on the hoehn and yahr staging scale: status and recommendations the movement disorder society task force on rating scales for parkinson's disease. Movement disorders 19(9), 1020–1028 (2004)

Jeancolas, L., Petrovska-Delacrétaz, D., Mangone, G., Benkelfat, B.E., Corvol, J.C., Vidailhet, M., Lehéricy, S., Benali, H.: X-vectors: New quantitative biomarkers for early parkinson's disease detection from speech. Frontiers in Neuroinformatics 15, 578369 (2021)

Katzenschlager, R., Head, J., Schrag, A., Ben-Shlomo, Y., Evans, A., Lees, A., et al.: Fourteen-year final report of the randomized pdrg-uk trial comparing three initial treatments in pd. Neurology 71(7), 474–480 (2008)

Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research 9(11) (2008)

MacMahon Copas, A.N., McComish, S.F., Fletcher, J.M., Caldwell, M.A.: The pathogenesis of parkinson's disease: a complex interplay between astrocytes, microglia, and t lymphocytes? Frontiers in Neurology 12, 666737 (2021)

Moro-Velazquez, L., Villalba, J., Dehak, N.: Using x-vectors to automatically detect parkinson's disease from speech. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1155–1159 (2020)

Moustafa, A.A., Chakravarthy, S., Phillips, J.R., Gupta, A., Keri, S., Polner, B., Frank, M.J., Jahanshahi, M.: Motor symptoms in parkinson's disease: A unified framework. Neuroscience & Biobehavioral Reviews 68, 727–740 (2016)

Pont-Sunyer, C., Hotter, A., Gaig, C., Seppi, K., Compta, Y., Katzenschlager, R., Mas, N., Hofeneder, D., Brücke, T., Bayés, A., et al.: The onset of nonmotor symptoms in parkinson's disease (the onset pd study). Movement Disorders 30(2), 229–237 (2015)

Ravanelli, M., Parcollet, T., Plantinga, P., Rouhe, A., Cornell, S., Lugosch, L., Subakan, C., Dawalatabad, N., Heba, A., Zhong, J., Chou, J.C., Yeh, S.L., Fu, S.W., Liao, C.F., Rastorgueva, E., Grondin, F., Aris, W., Na, H., Gao, Y., Mori, R.D., Bengio, Y.: SpeechBrain: A general-purpose speech toolkit (2021), arXiv:2106.04624

Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O.: Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics 17(4), 828–834 (2013)

Snyder, D., Garcia-Romero, D., McCree, A., Sell, G., Povey, D., Khudanpur, S.: Spoken language recognition using x-vectors. In: Odyssey. vol. 2018, pp. 105–111 (2018)

Sveinbjornsdottir, S.: The clinical symptoms of parkinson's disease. Journal of neurochemistry 139, 318–324 (2016)

Sztahó, D., Valálik, I., Vicsi, K.: Parkinson's disease severity estimation on hungarian speech using various speech tasks. In: 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). pp. 1–6 (2019)

Vadovský, M., Paralič, J.: Parkinson's disease patients classification based on the speech signals. In: 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI). pp. 000321–000326 (2017)