Modelling of Speech Aspects in Parkinson‘s Disease by Multitask Deep Learning

Master’s thesis

Martin Strauß, 21.06.2019

Supervisor: Prof. Dr.-Ing. Elmar Nöth, M.Sc. Juan Camilo Vásquez-Correa, PD Dr.-Ing. Tino Haderlein

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Motivation
Background
Data
Methods and experiments
Results
Discussion
Conclusion and Outlook

Motivation

Parkinson‘s disease (PD) second largest neurodegenerative disorder[1]

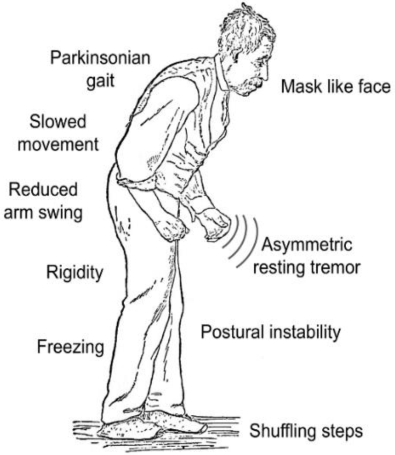
Speech impairments are one of the earliest manifestations

Speech is affected in various dimensions including articulation, phonation, prosody and intelligibility (hypokinetic dysarthria):

- **Reduced loudness**
- **Monotonic speech**
- **Breathy voice**

etc.

→ *Underrepresented in PD evaluation[2]*



http://theconversation.com

[1] Tysnes et al. (2017), Journal of Neural Transmission
[2] Ramig et al. (2008), Expert Review of Neurotherapeutics

PD assessment

Movement Disorder Society – Unified Parkinson‘s disease rating scale (MDS-UPDRS)[4]:

- Third section with 33 items evaluates disease progression (MDS-UPDRS-III)

- Rated between 0-4

→ *Only one aspect related to speech*

→ *Patient is required to be with a physician*

→ *Subjectivity*

[4] Goetz et al. (2008), Movement Disorders

PD assessment

Movement Disorder Society – Unified Parkinson‘s disease rating scale (MDS-UPDRS):

→ ***Only one aspect related to speech***

→ ***Patient is required to be with a physician***

→ ***Subjectivity***

Frenchay Dysarthria Assessment (FDA):

- Items like reflexes, respiration, lips movement etc.

→ ***Patient is required to be with a physician***

→ ***Subjectivity***

- modified-FDA (m-FDA)[5]: Assessment only relies on speech recordings

[5] Vásquez-Correa (2018), Journal of Communication Disorders

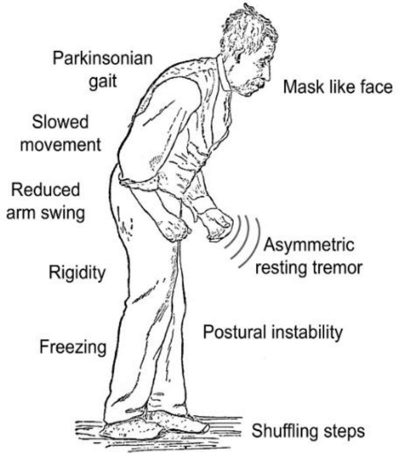
Motivation

Parkinson‘s disease (PD) second largest neurodegenerative disorder[1]

Speech impairments are one of the earliest manifestations
→ *Underrepresented in PD evaluation[2]*

Disease assessment like MDS-UPDRS-III and m-FDA are highly subjective

→ Computational based methods to increase objectivity and enable long-term monitoring



http://theconversation.com

[1] Tysnes et al. (2017), Journal of Neural Transmission
[2] Ramig et al. (2008), Expert Review of Neurotherapeutics

Background

Various feature extraction and machine learning methods already exist

Success of Deep Learning (DL) now also in PD speech assessment tasks[3]

→ ***Most studies concentrate on one aspect like PD vs. healthy controls classification***

→ ***Multitask Learning (MTL) approaches allow to evaluate multiple aspects at once***

[3] Vásquez-Correa et al. (2017), In Interspeech 2017

Multitask learning

$$L(\theta) = \gamma L_1(\theta) + (1-\gamma) L_2(\theta)$$

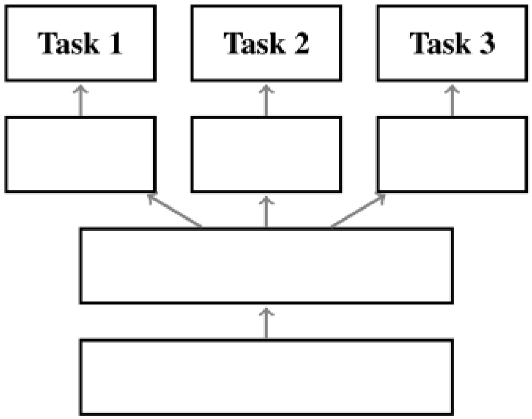
Optimize more than one loss function

→ Loss weight factor γ

$$L(\theta) = \sum_i \gamma_i L_i(\theta)$$

$$\sum_i \gamma_i = 1$$

Hard parameter sharing with shared layers and individual task layers

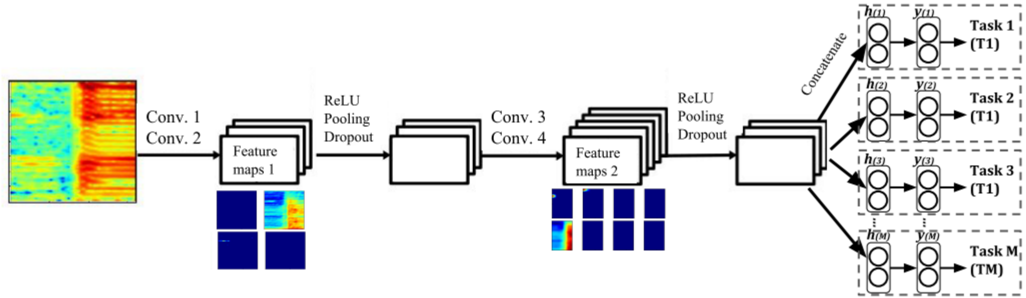


→ *Idea: Multiple tasks share representations creating more general feature maps[6]*

[6] Caruana (1997), Machine Learning

Related work

Vásquez-Correa et al.[7]:



- 11 speech aspects jointly optimized in CNN
- Voiced and unvoiced speech segments as input
- Up to 4 percent points increased accuracy by MTL approach

→ ***MTL creates more generalizable feature maps***

[7] Vásquez-Correa et al. (2018), In Interspeech 2018

Main aims

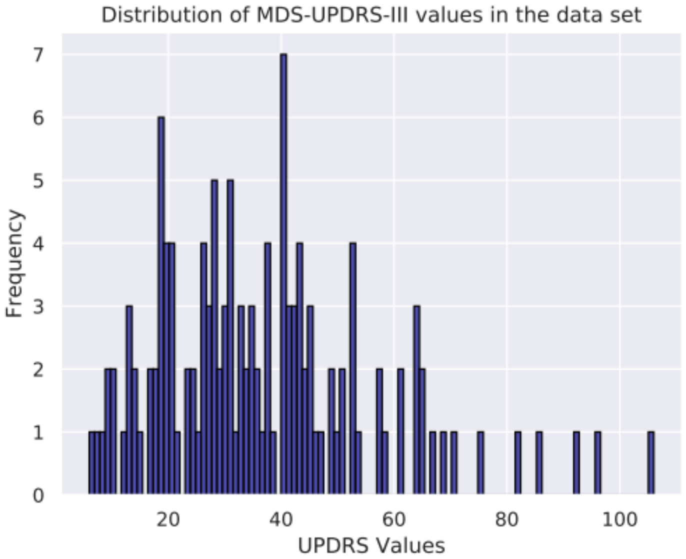
Apply a MTL neural network framework to PD speech data

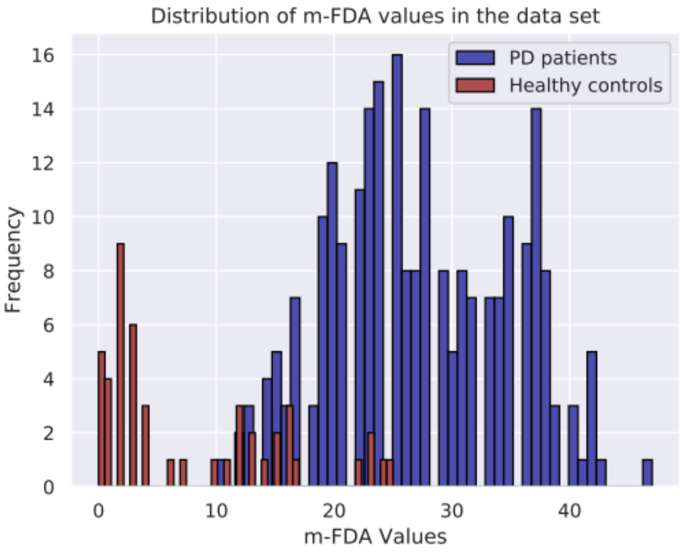
Evaluate the proposed approach compared to single task networks and a baseline algorithm

Motivation
Background
Data
Methods and experiments
Results
Discussion
Conclusion and Outlook

Data set

94 PD and 87 HC Spanish native speakers from Colombia





Recordings obtained in various sessions performing different exercises

→ *In total 5145 utterances included into the data set*

Task description

Exercises: DDK, Sentences, Monologue, Reading text

Acoustic condition: Soundproof booth, Portable soundproof booth, Headset, At-home

Task number	Task name	Number of classes
1.	PD vs. HC	2
2.	MDS-UPDRS-III	4
3.	m-FDA	4
4.	Acoustic condition	4
5.	Exercise	4
6.	Gender	2
7.	Age	4

Task description

Tasks: PD vs. HC MDS-UPDRS-III m-FDA

Knowledge features

Exercises: DDK, Sentences, Monologue, Reading text

Acoustic condition: Soundproof booth, Portable soundproof booth, Headset, At-home

Age

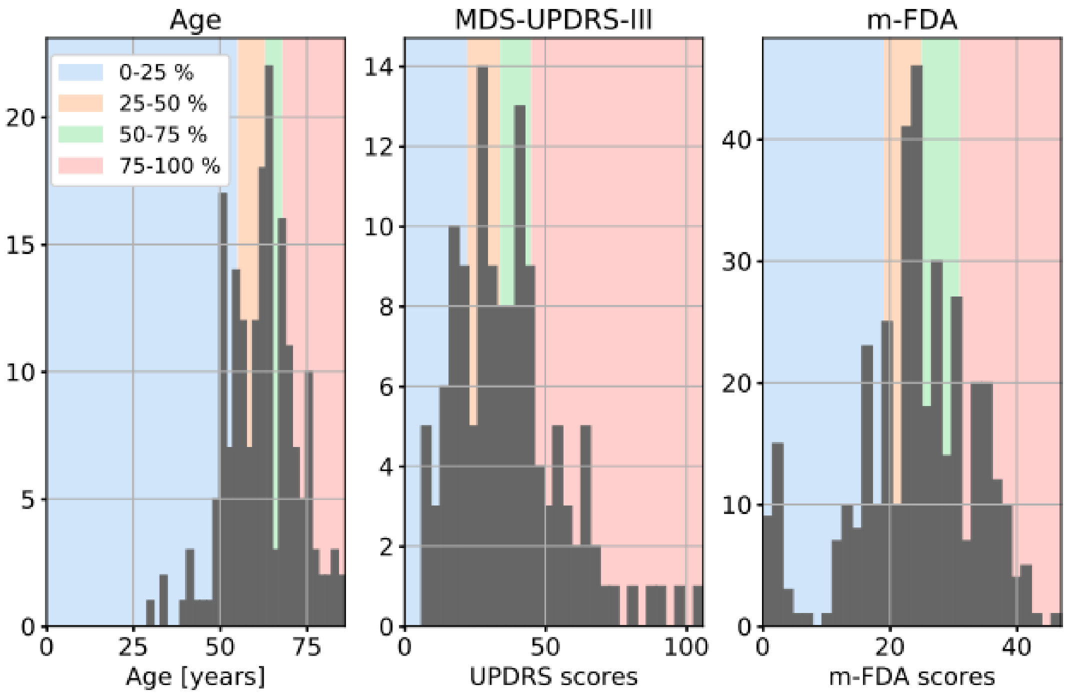
Gender

Labelling procedure

Classes defined using percentiles

HC are assumed to have lower UPDRS values then PD patients
→ *Part of the first class*

Missing m-FDA labels for HC were estimated using SVR[8]



[8] Vásquez-Correa et al. (2018), Journal of Communication Disorders

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

openSMILE[9] features

https://www.audeering.com/opensmile/

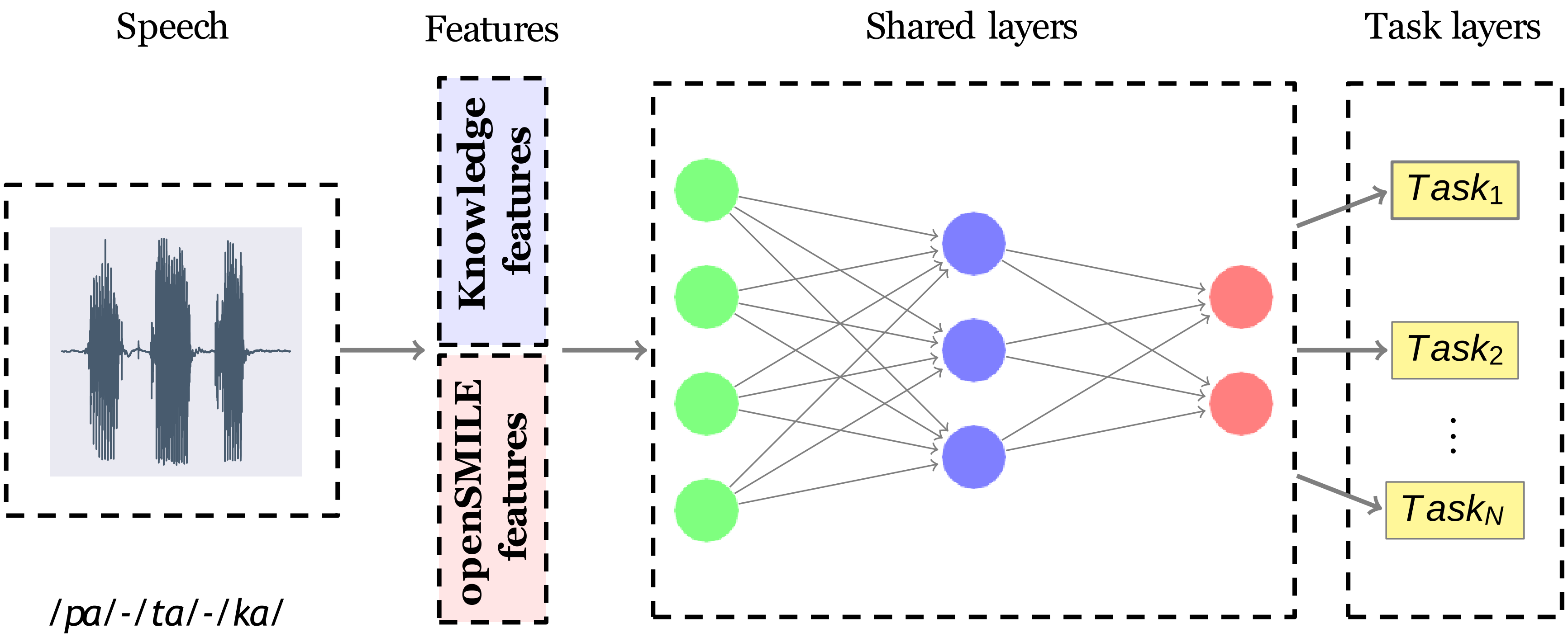
Precomputed features using the openSMILE software

Open source feature extractor also used e.g. in the Interspeech ComPar challenge[10]

1428 features per utterance as input data

[9] Eyben et al. (2013), In Proceddings of the 21st International conference on Multimedia
[10] Schuller et al. (2010), In Interspeech 2010

Architecture



Experimental setup

10 fold cross validation with parameter optimization (lr, dropout prob., hidden layers)

Architectures:

- Adaboost baseline
- Single task networks
- MTL different experiments
- MTL all seven tasks

Experiment	Loss function weight factor γ PD vs. HC	MDS-UPDRS-III	m-FDA
1	0.8	0.1	0.1
2	0.1	0.8	0.1
3	0.1	0.1	0.8
4	learned	learned	learned

Metrics:

- Accuracy
- Unweighted average recall
- Session based evaluation

Motivation
Background
Data
Methods
Results
Discussion
Conclusion and Outlook

Experimental results

Adaboost results

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	80.70	77.21	85.71
MDS-UPDRS-III	53.29	32.14	63.77
m-FDA	36.56	33.78	40.28

Single task results

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	71.82	71.12	78.22
MDS-UPDRS-III	29.46	29.13	36.23
m-FDA	36.83	34.50	40.52

MTL with focus on PD vs. HC

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	73.16	72.49	81.73
MDS-UPDRS-III	46.79	34.45	52.45
m-FDA	37.30	35.94	43.56

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Adaboost

The algorithm is able to differentiate PD vs. HC

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	80.70	77.21	85.71
MDS-UPDRS-III	53.29	32.14	63.77
m-FDA	36.56	33.78	40.28

Related to the imbalance of the data

MDS-UDPRS-III

Reference \ Prediction	1	2	3	4
1	1719	165	112	116
2	311	76	76	79
3	234	109	98	108
4	264	41	110	75

(a) Confusion matrix

Reference \ Prediction	1	2	3	4
1	0.81	0.08	0.05	0.05
2	0.57	0.14	0.14	0.15
3	0.43	0.20	0.18	0.20
4	0.54	0.08	0.22	0.15

(b) Normalized confusion matrix

→ ***Putting a high weight on the first class still reaches decent results***

Single task networks

Showing the worst results

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	71.82	71.12	78.22
MDS-UPDRS-III	29.46	29.13	36.23
m-FDA	36.83	34.50	40.52

Parameter optimization shows more complex models (more layers) are chosen

→ *More regularization necessary*

<u>MDS-UDPRS-III</u>

Reference \ Prediction	1	2	3	4
1	611	73	1384	44
2	68	15	430	29
3	64	15	431	39
4	46	11	402	31

(a) Confusion matrix

Reference \ Prediction	1	2	3	4
1	0.29	0.03	0.66	0.02
2	0.13	0.03	0.79	0.05
3	0.12	0.03	0.79	0.07
4	0.09	0.02	0.82	0.06

(b) Normalized confusion matrix

Problems with overfitting or local minimums

Multitask networks

Best results for focusing on PD vs. HC

Task	ACC (%)	UAR (%)	Session (%)
PD vs. HC	73.16	72.49	81.73
MDS-UPDRS-III	46.79	34.45	52.45
m-FDA	37.30	35.94	43.56

Comparable results to the baseline, but better than individual networks

<u>MDS-UDPRS-III</u>

Reference \ Prediction	1	2	3	4
1	1351	129	300	332
2	208	57	90	187
3	216	65	90	178
4	211	28	21	230

(a) Confusion matrix

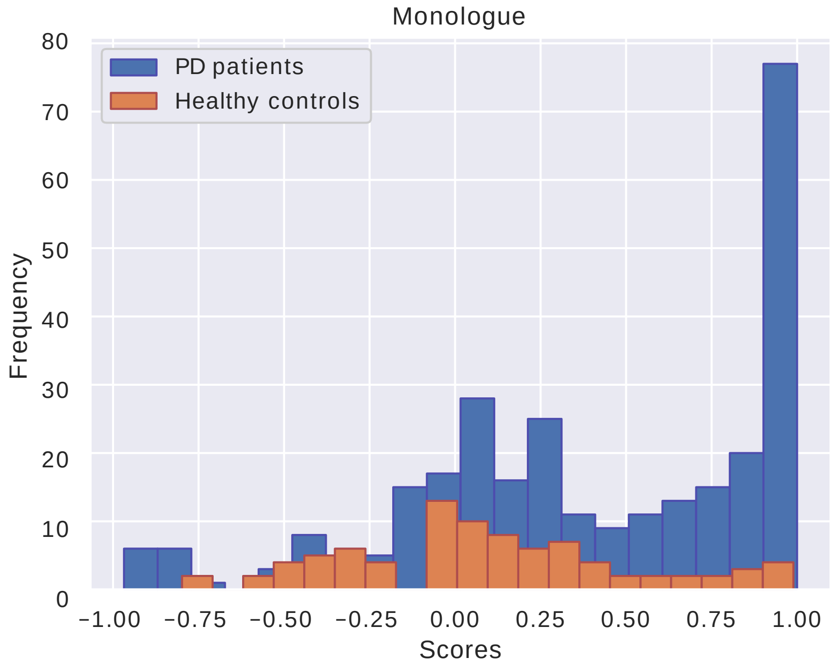
Reference \ Prediction	1	2	3	4
1	0.64	0.06	0.14	0.16
2	0.38	0.11	0.17	0.35
3	0.39	0.12	0.16	0.32
4	0.43	0.06	0.04	0.47

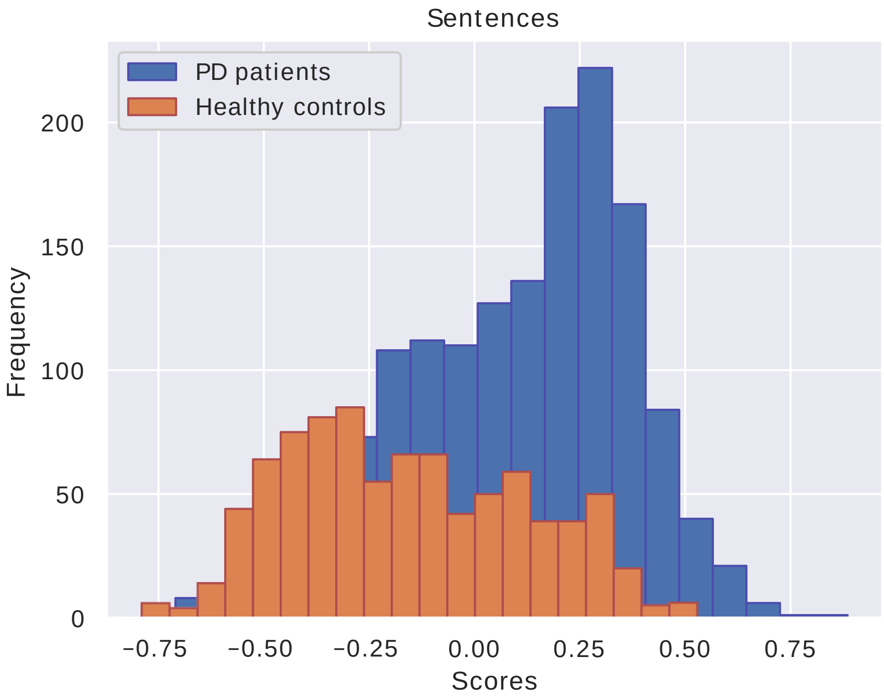
(b) Normalized confusion matrix

Adding more tasks does not help

Confidence score

Based on the softmax output and the true PD vs. HC labels





Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Motivation
Background
Data
Methods and experiments
Results
Discussion
Conclusion and Outlook

Conclusion

- **Simple Adaboost baseline delivers solid results**
- **MTL approach superior to single task networks**
- **No clear advantadge over the Adaboost baseline**
- **Adding more tasks does not increase the performance**

Outlook

- **Add a fifth class to the MDS-UPDRS-III task for the HC samples**
- **Find the best trade-off model working for all folds**
- **Convert the MDS-UPDRS-III and m-FDA tasks into a regression problem**
- **Combine proposed approach with other approaches (e.g. CNN)**
- **Investigate larger feature sets**
- **Obtaining the missing labels**

References

[1] O. - B. Tysnes and A. Storstein. “Epidemiology of Parkinson‘s disease”. *Journal of Neural Transmission*, 124(8):901-905, 2017

[2] L. O. Ramig, C. Fox and S. Sapir. “Speech treatment for Parkinson‘s disease”. *Expert Review of Neurotherapeutics*, 8(2):297-309, 2008

[3] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth. “Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson’s Disease”. In Proceedings of Interspeech 2017, pages 314–318, 2017.

[4] C. G. Goetz, 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: official journal of the Movement Disorder Society, 23(15):2129–2170, 2008.

References

[5] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth. “Towards an automatic evaluation of the dysarthria level of patients with Parkinson’s disease”. Journal of Communication Disorders, 76:21–36, 2018.

[6] R. Caruana. “Multitask Learning“. *Machine Learning*, 28(1): 41-75, 1997

[7] J. C. Vàsquez-Correa, T. Arias, J. R. Orozco-Arroyave and E. Nöth. ``A Multitask Learing Approach to Assess the Dysarthria Severity in Patients with Parkinson‘s Disease”. In *Interspeech 2018*, pages 456-460, ISCA, 2018

[8] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth. “Towards an automatic evaluation of the dysarthria level of patients with Parkinson’s disease”. Journal of Communication Disorders, 76:21–36, 2018

References

[9] F. Eyben, F. Weninger, F. Gross and B. Schuller. “Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor”.In *Proceedings of the ACM international conference on Multimedia – MM ’13*, pages 835-838. ACM Press, 2013

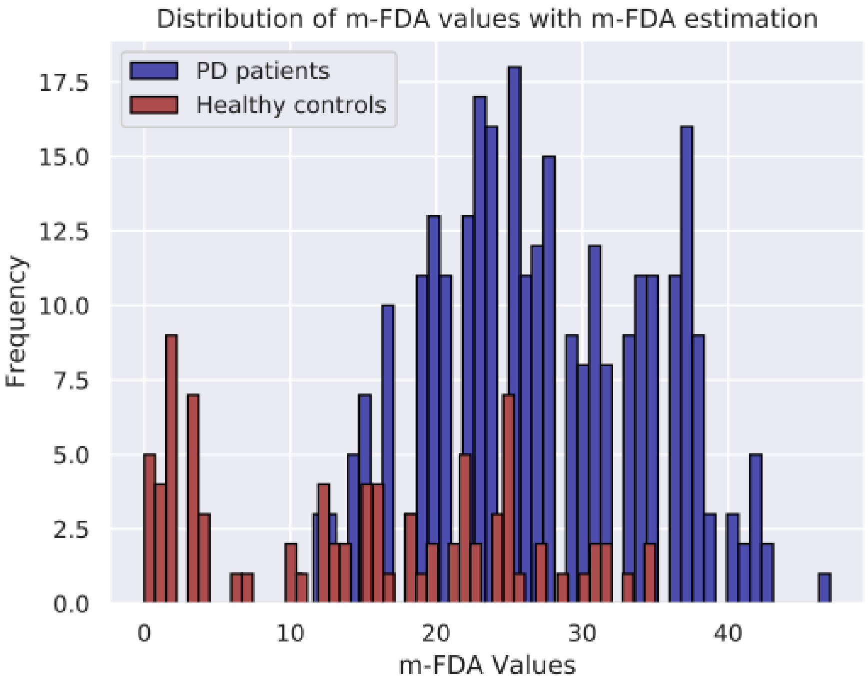
[10] B. Schuller, S. Steidl, A. Batlinger, S. Hantke, F. Hönig, J.R. Orozco-Arroyave, E. Nöth, Y. Zhang and F. Weninger. “The INTERSPEECH 2010 Paralinguistic Challenge”. In *Proceedings of Interspeech 2010*, ISCA, 2010

More information?

Portable soundproof booth

m-FDA label estimation

Based on a Support vector regression approach



PD vs. HC results

Adaboost

		Prediction PD	Prediction HC
Reference	PD	3083	479
	HC	496	1069

(a) Confusion matrix

		Prediction PD	Prediction HC
Reference	PD	0.86	0.14
	HC	0.32	0.68

(b) Normalized confusion matrix

Single task

		Prediction PD	Prediction HC
Reference	PD	2610	970
	HC	480	1085

(a) Confusion matrix

		Prediction PD	Prediction HC
Reference	PD	0.73	0.27
	HC	0.31	0.69

(b) Normalized confusion matrix

Multitask

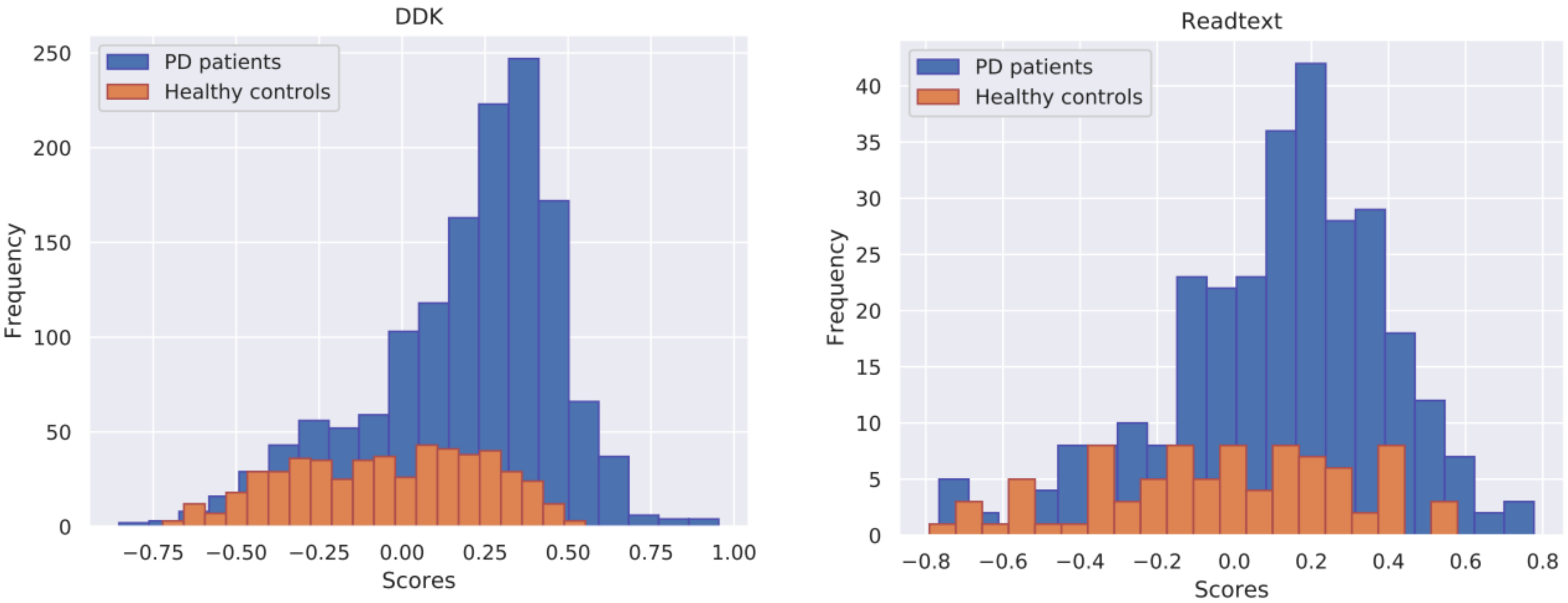
		Prediction PD	Prediction HC
Reference	PD	2656	924
	HC	457	1108

(a) Confusion matrix

		Prediction PD	Prediction HC
Reference	PD	0.74	0.26
	HC	0.29	0.71

(b) Normalized confusion matrix

Additional confidence scores



More tables

Samples per Exercise

Exercise	PD	HC	
DDK	1411	552	1933
Sentences	1588	870	2458
Read text	292	87	379
Monologue	289	86	375
Total	3580	1565	5145

Percentile ranges

Task	0–25	25–50	50–75	75–100
Age	< 55	55–63	63–68	$68 <$
MDS-UPDRS-III	< 22	22–34	34–45	$45 <$
m-FDA	< 19	19–25	25–31	$31 <$

Number of samples per percentile

Task	0–25	25–50	50–75	75–100
Age	1246	1507	1253	1079
MDS-UPDRS-III	2112	542	549	490
m-FDA	1586	1165	1412	982