Aprendizaje por transferencia en redes neuronales convolucionales para el diagnóstico y monitoreo de la enfermedad de Parkinson usando señales de voz en tres idiomas diferentes

Cristian David Rios Urrego

BSc. student in Electronic Engineering

Advisor: **Prof. Juan Rafael Orozco Arroyave Ph.D.**

Co-Advisor: **MSc. Juan Camilo Vazquez Correa**

GITA research group, University of Antioquia.

cdavid.rios@udea.edu.co

June 5, 2019

Outline

Introduction
- Overview
- Hypothesis and objectives

Databases

Methodology

Data pre-processing
- Segmentation
- Short-time Fourier transform

Convolutional Neural Network
- Convolution stage
- Pooling stage

Transfer learning

Experiments and Results

Conclusions
- Conclusions and Future work

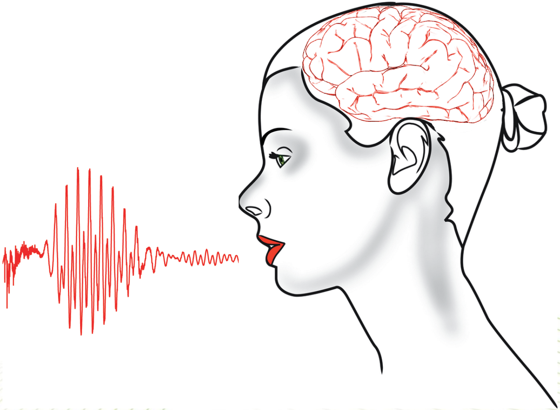
UNIVERSIDAD DE ANTIOQUIA
1803

Introduction

Context: Parkinson's disease

Parkinson's disease (PD) is a neurodegenerative disorder characterized by symptoms such as resting tremor, bradykinesia, rigidity and alterations in the gait, caused by the loss of dopaminergic neurons.

Context: Parkinson's disease



Speech symptoms

- Low voice volume
- Reduction of prosodic pitch
- Monotonous speech
- Voice tremor
- Imprecise articulation

Context: Parkinson's disease

Computational tools

- Early Detection
- Diagnostic support
- Neurological state monitoring

Hypothesis and objectives

Hypothesis

It is possible to improve the classification of patients with Parkinson's disease and healthy controls from transfer learning in monolingual data.

Objectives

General Objective

To implement and evaluate the method of transfer learning in convolutional neural networks (CNN) for three different languages in order to support diagnosis and monitor patients with PD.

Hypothesis and objectives

Objectives

Specific Objectives

1. To implement algorithms for pre-processing and segmentation of voice signals, for the extraction of onset-offset transitions.
2. To design and train CNNs with ResNet topology for different languages from time-frequency representations of the transitions.
3. To implement the transfer learning technique in the trained models of CNNs for the evaluation and monitoring of PD patients.
4. To evaluate and compare the performance of CNNs trained in different languages and the CNNs implemented with the transfer learning technique.

UNIVERSIDAD DE ANTIOQUIA
1803

Databases

Spanish

Table: Information of the speakers in PC-GITA. μ: mean, σ: standard deviation .

	PD patients		Healthy controls	
	Male	Female	Male	Female
Number of subjects	25	25	25	25
Age ($\mu \pm \sigma$)	61.3 $\pm$ 11.4	60.7 $\pm$ 7.3	60.5 $\pm$ 11.6	61.4 $\pm$ 7.0
Range of age	33 – 81	49 – 75	31 – 89	49 – 76
Disease duration ($\mu \pm \sigma$)	8.7 $\pm$ 5.8	12.6 $\pm$ 11.6		
MDS-UPDRS-III ($\mu \pm \sigma$)	37.8 $\pm$ 22.1	37.6 $\pm$ 14.1		
Range of MDS-UPDRS-III	6 – 93	19 – 71		

Tasks

- 10 sentences
- The rapid repetition of diadochokinetics (DDKs)
- Read text with 36 words
- Monologue

Databases

German

Table: Information of the speakers in the German database. $\boldsymbol{\mu}$: mean, $\boldsymbol{\sigma}$: standard deviation .

	PD patients		Healthy controls	
	Male	Female	Male	Female
Number of subjects	47	41	44	44
Age ($\mu \pm \sigma$)	66.7 ± 8.4	66.2 ± 9.7	63.8 ± 12.7	62.6 ± 15.2
Range of age	44 – 82	42 – 84	26 – 83	28 – 85
Disease duration ($\mu \pm \sigma$)	7.0 ± 5.5	7.1 ± 6.2		
MDS-UPDRS-III ($\mu \pm \sigma$)	22.1 ± 9.9	23.3 ± 12.0		
Range of MDS-UPDRS-III	5 – 43	6 – 5		

Tasks

- 5 sentences
- The rapid repetition of the syllables /pa-ta-ka/
- Read text with 81 words
- Monologue

Czech

Table: Information of the speakers in the Czech database. $\boldsymbol{\mu}$: mean, $\boldsymbol{\sigma}$: standard deviation .

	PD patients		Healthy controls	
	Male	Female	Male	Female
Number of subjects	30	20	30	19
Age ($\mu \pm \sigma$)	65.3 $\pm$ 9.6	60.1 $\pm$ 8.7	60.3 $\pm$ 11.5	63.5 $\pm$ 11.1
Range of age	43 – 82	41 – 72	41 – 77	40 – 79
Disease duration ($\mu \pm \sigma$)	6.7 $\pm$ 4.5	6.8 $\pm$ 5.2		
MDS-UPDRS-III ($\mu \pm \sigma$)	21.4 $\pm$ 11.5	18.1 $\pm$ 9.7		
Range of MDS-UPDRS-III	4 – 54	6 – 38		

Tasks

- The rapid repetition of the syllables /pa-ta-ka/
- Read text with 80 words
- Monologue

Methodology

Methodology

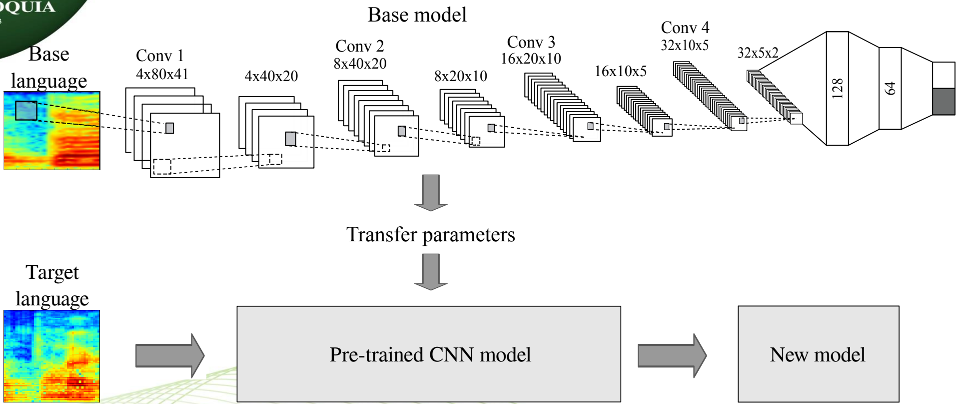


Figure: Transfer learning strategy proposed in this study to classify PD from speech with utterances from different languages.

Data pre-processing

Segmentation

Voiced and unvoiced segments are identified by the presence of the fundamental frequency of the voice (pitch) in frames of short duration.

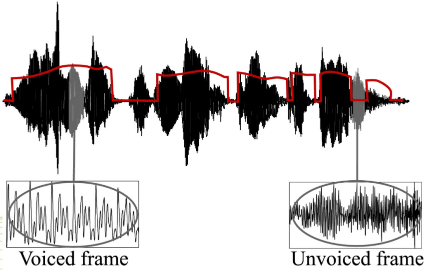


Figure: Voiced/unvoiced segments. Figure taken from Arias-Vergara et al. 2018.

Segmentation

Onset and offset transitions are considered to model difficulties of the PD patients to start and to stop a movement like the vocal fold vibration.

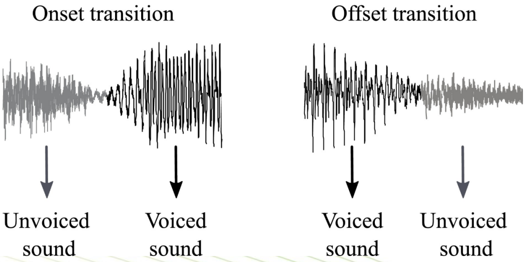


Figure: Onset/offset transitions. Figure taken from Arias-Vergara et al. 2018.

Time-frequency representations

The short-time Fourier transform (STFT), is a Fourier-related transform used to determine the frequency content (Ω) of local sections of a signal as it changes over time.

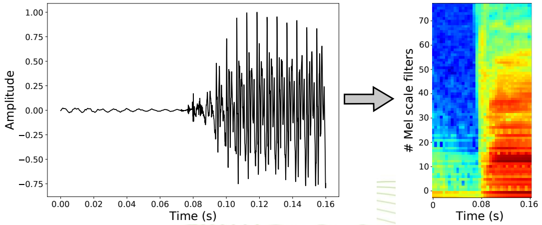
$$X_m(\Omega) = \sum_{n=-\infty}^{\infty} x(n) f(n-m) e^{-j\Omega n} \tag{1}$$

Where $x(n)$ is the signal to be transformed, and $f(n)$ is the window function, commonly a Blackman, Hamming or Hanning window.

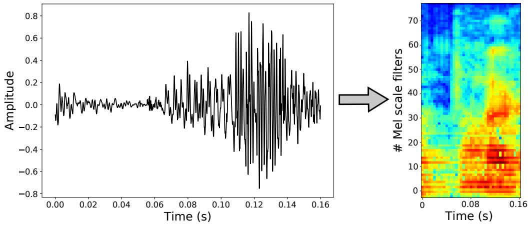
Time-frequency representations

UNIVERSIDAD DE ANTIOQUIA

Male healthy control
Age: *54*



Male PD patient.
Age: *48;* MDS-UPDRS: *9*



UNIVERSIDAD DE ANTIOQUIA 1803

Convolutional Neural Network

A CNN typically consists of 3 stages: a convolution stage in parallel to produce a set of linear activations, a pooling stage to modify the output of the layer and a classification stage.

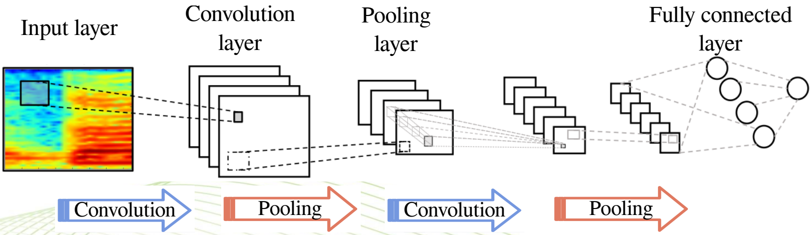


Figure: Typical structure of a CNN.

Convolution Stage

The convolution operation has the effect of filtering the input image with a trainable kernel.

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \qquad (2)$$

Where (x) is known as the input, and the second argument (w) is the kernel, and the output is the feature map.

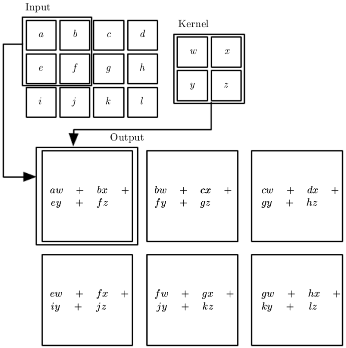


Figure: Example of a 2-D convolution.Figure taken from Goodfellow, Bengio, and Courville 2016.

Pooling Stage

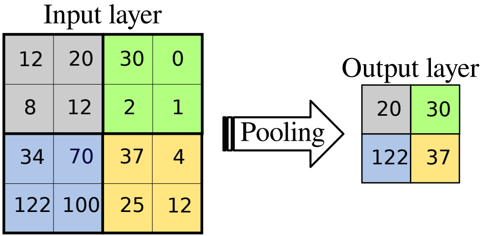


Figure: Pooling layer using the max pooling method.

The main function is to reduce the spatial dimensions of the input layer from a statistical summary of the nearest outputs in the layer.

Transfer learning

Transfer learning

The initial idea of transfer learning is to reuse the experience gained to improve the learning of new models.

Transfer learning can take advantage of the knowledge (features and weights) of previously created models to train new models and even address model problems with small amounts of data.

Transfer learning

Unlike traditional learning that is isolated and based exclusively on specific tasks, data sets and training on separate models, learning by transfer takes advantage of knowledge from previously created models.

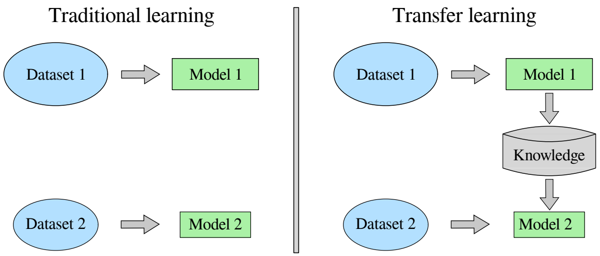


Figure: Comparison between traditional learning and transfer learning.

Experiments and Results

Results: Validation and Regularization

- 10-fold Cross-Validation strategy, speaker independent.
- Regularization:
  a) L^2 Regularization.
  b) Dropout.
  c) Early stopping.

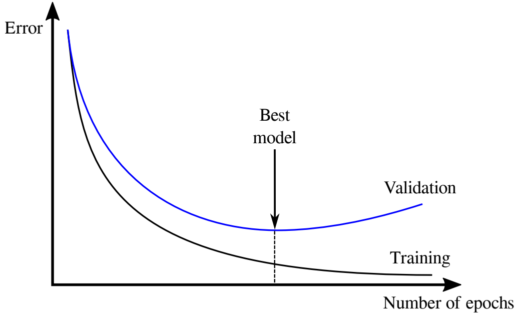


Figure: Early stopping.

Results: Architectures implemented

Table: ResNet20 Architecture.

Stage	Layer type	Output size
Input	Conv (1x16x3,1)	16x80x41
Block 1	Conv (16x16x3,1)	16x80x41
	Conv (16x16x3,1)	
	Conv (16x16x3,1)	
	Conv (16x16x3,1)	
	Conv (16x16x3,1)	
	Conv (16x16x3,1)	
Block 2	Conv (16x32x3,2)	32x40x21
	Conv (32x32x3,2)	
	Conv (32x32x3,2)	
	Conv (32x32x3,2)	
	Conv (32x32x3,2)	
	Conv (32x32x3,2)	
Block 3	Conv (32x64x3,2)	64x20x11
	Conv (64x64x3,2)	
	Conv (64x64x3,2)	
	Conv (64x64x3,2)	
	Conv (64x64x3,2)	
	Conv (64x64x3,2)	
Pooling	Avg Pool (11)	1x1x64
Output	Lineal (64,2)	1x1x2

Table: LeNet Architecture.

Layer type	Output size
Conv (1x4x3,1) + dropout	4x80x41
Max Pool (2,2)	4x40x20
Conv (4x8x3,1) + dropout	8x40x20
Max Pool (2,2)	8x20x10
Conv (8x16x3,1) + dropout	16x20x10
Max Pool (2,2)	16x10x5
Conv (16x32x3,1) + dropout	32x10x5
Max Pool (2,2)	32x5x2
Lineal (320,128) + dropout	1x1x128
Lineal (128,64) + dropout	1x1x64
Lineal (64,2)	1x1x2

Results: CNN monolingual

Table: Classification results for the architectures implemented with CNN models trained in three different languages. Acc: Accuracy. Sen: Sensitivity. Spe: Specificity.

	ResNet20 Architecture			LeNet Architecture		
Language	Acc ($\mu \pm \sigma$)	Sen ($\mu \pm \sigma$)	Spe ($\mu \pm \sigma$)	Acc ($\mu \pm \sigma$)	Sen ($\mu \pm \sigma$)	Spe ($\mu \pm \sigma$)
Spanish	71.0 $\pm$ 11.0	58.0 $\pm$ 17.5	84.0 $\pm$ 15.8	71.0 $\pm$ 15.9	74.0 $\pm$ 25.0	68.0 $\pm$ 28.6
German	70.9 $\pm$ 9.90	74.8 $\pm$ 22.1	66.9 $\pm$ 15.9	63.1 $\pm$ 11.7	43.1 $\pm$ 38.0	83.1 $\pm$ 17.7
Czech	61.9 $\pm$ 12.0	90.0 $\pm$ 14.1	33.5 $\pm$ 29.1	68.5 $\pm$ 14.1	94.0 $\pm$ 13.5	42.0 $\pm$ 33.2

Results: Transfer language to Spanish

Table: Classification results for the transfer learning to Spanish.

Language	η	Drop	L^2	Acc ($\mu \pm \sigma$)	Sen ($\mu \pm \sigma$)	Spe ($\mu \pm \sigma$)
Spanish	0.005	0.3	0.0005	71.0 ± 15.9	74.0 ± 25.0	68.0 ± 28.6
Czech–Spanish	0.005	0.3	0.0005	**72.0** ± **13.1**	67.0 ± 11.6	78.0 ± 23.9
German–Spanish	0.005	0.3	0.0005	70.0 ± 12.5	62.0 ± 19.9	78.0 ± 29.0

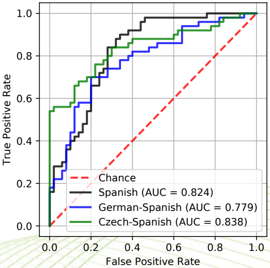


Figure: ROC curve for the transfer learning to Spanish.

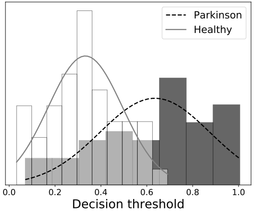


Figure: Histogram and the corresponding probability density distribution for Czech-Spanish model.

Results: Transfer language to German

Table: Classification results for the transfer learning to German.

Language	η	Drop	L^2	Acc ($\mu \pm \sigma$)	Sen ($\mu \pm \sigma$)	Spe ($\mu \pm \sigma$)
German	0.006	0.4	0.0005	63.1 ± 11.7	43.1 ± 38.0	83.1 ± 17.7
Czech–German	0.006	0.4	0.0005	76.7 ± 7.9	87.5 ± 11.0	66.0 ± 15.6
Spanish–German	0.006	0.4	0.0005	**77.3 ± 11.3**	86.2 ± 13.8	68.3 ± 14.3

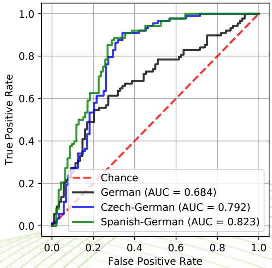


Figure: ROC curve for the transfer learning to German.

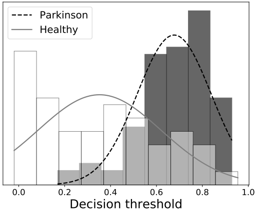


Figure: Histogram and the corresponding probability density distribution for Spanish-German model.

Results: Transfer language to Czech

Table: Classification results for the transfer learning to Czech.

Language	η	Drop	L^2	Acc ($\mu \pm \sigma$)	Sen ($\mu \pm \sigma$)	Spe ($\mu \pm \sigma$)
Czech	0.005	0.1	0.001	68.5 $\pm$ 14.1	94.0 $\pm$ 13.5	42.0 $\pm$ 33.2
German–Czech	0.005	0.1	0.001	70.7 $\pm$ 14.5	80.0 $\pm$ 16.3	62.5 $\pm$ 26.3
Spanish–Czech	0.005	0.1	0.001	**72.6** $\pm$ **13.9**	82.0 $\pm$ 14.8	62.0 $\pm$ 28.9

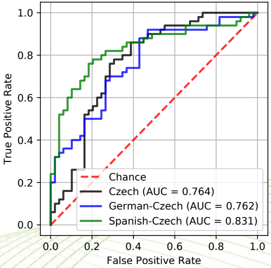


Figure: ROC curve for the transfer learning to Czech.

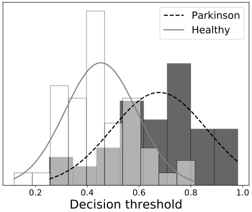
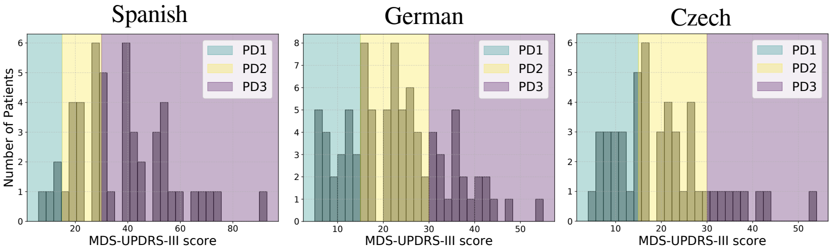


Figure: Histogram and the corresponding probability density distribution for Spanish-Czech model.

Results: Multiclass classification



Healthy controls and PD Patients were chosen in 4 groups:

- Healthy Controls (HC).
- PD1: Patients with MDS-UPDRS-III scores between 0 and 15.
- PD2: Patients with MDS-UPDRS-III scores between 16 and 30.
- PD3: Patients with MDS-UPDRS-III scores above 31

Results: Multiclass classification

Table: Confusion matrices with results of classifying HC subjects and PD patients in different stages of the disease, Acc: Accuracy, κ: Cohen kappa coefficient. The results are expressed in (%).

	Spanish				German				Czech			
	Acc = 60.0		κ = 0.38		Acc = 50.6		κ = 0.30		Acc = 41.4		κ = 0.13	
	HC	PD1	PD2	PD3	HC	PD1	PD2	PD3	HC	PD1	PD2	PD3
HC	**72.0**	4.0	0.0	24.0	**51.1**	5.7	35.2	8.0	**57.1**	18.4	8.2	16.3
PD1	20.0	**20.0**	0.0	60.0	7.4	**14.8**	66.7	11.1	63.1	**21.1**	0.0	15.8
PD2	22.2	11.1	**5.6**	61.1	10.0	5.0	**80.0**	5.0	34.8	8.7	**21.7**	34.8
PD3	14.8	3.7	0.0	**81.5**	14.3	9.5	38.1	**38.1**	37.5	12.5	0.0	**50.0**

UNIVERSIDAD DE ANTIOQUIA 1803

Conclusions

Conclusions

- Transfer learning can improve the performance of monolingual models, with increases of up to 14% in accuracy. It is also possible to evaluate the severity of patients from the models created, obtaining results of up to 60% accuracy.

Conclusions

- Transfer learning can improve the performance of monolingual models, with increases of up to 14% in accuracy. It is also possible to evaluate the severity of patients from the models created, obtaining results of up to 60% accuracy.
- The method of knowledge transfer in other languages gets good results as long as the basic model is robust enough, i.e. it performs well with its training and test data.

Conclusions

- Transfer learning can improve the performance of monolingual models, with increases of up to 14% in accuracy. It is also possible to evaluate the severity of patients from the models created, obtaining results of up to 60% accuracy.
- The method of knowledge transfer in other languages gets good results as long as the basic model is robust enough, i.e. it performs well with its training and test data.
- Deep learning outperforms traditional learning strategies, as long as the input data is good enough, the appropriate architecture is used, and regularization measures are implemented avoiding overfitting.

Future Work

- Creation of more robust base models, increasing the number of training data by combining 2 of the 3 databases and transfering knowledge to the remaining language.
- Implement a Bayesian optimization algorithm in order to obtain the optimal parameters for each network.
- Implement a learning by transference between different pathologies.