Speech task based automatic classification of ALS and Parkinson’s Disease and their severity using log Mel spectrograms

Suhas BN\textsuperscript{1}, Jhansi Mallela\textsuperscript{2}, Aravind Illa\textsuperscript{2}, Yamini Belur\textsuperscript{3}, Nalini Atchayaram\textsuperscript{4}, Ravi Yadav\textsuperscript{4}, Prasanta Kumar Ghosh\textsuperscript{2}

\textsuperscript{1}Penn State University, \textsuperscript{2}SPIRE LAB, Indian Institute of Science, \textsuperscript{3,4}Dept. of SPA and Neurology, NIMHANS

Paper Presentation
July 2020
Overview

1. Introduction

2. Proposed approach

3. Data

4. Experiments and Results

5. Conclusions
Amyotrophic Lateral Sclerosis (ALS)

- A motor neuron disorder
- Neurons: communication link
- Gradual degeneration of motor neurons
- Loss of muscle control

Symptoms of ALS

- Muscle stiffness
- A hard time in holding items
- Muscle cramps
- Swallowing problems
- Speech difficulties (slurred or slowness)

---

Parkinson's Disease (PD)

- A progressive brain disorder
- Impairment of nerve cells
- Neurons: Dopamine production
- Movement issues

Introduction

Symptoms of PD

- Tremors (hands, arms, legs)
- Stiffness in limbs and trunk
- Slowness in movements
- Difficulty in swallowing and chewing
- Speech difficulties (slurred or slowness)

---

Introduction

Life expectancy with ALS or PD

- ALS affected people\(^1,2\)
  - 50% people - 3 or more years
  - 20% people - 5 or more years
  - 10% people - 10 or more years

- PD affected people\(^3\)
  - Has increased from 9.4 (1967) to 14.6 years (2016) after diagnosis

---


Diagnosis and Treatment for ALS and PD

■ Currently no specific tests that can confirm of having ALS or PD

■ Diagnosis

■ No cure for either ALS or PD although there exists some treatment for managing their symptoms

---


Motivation, Plans and Challenges

Motivation

- Objective: Automated methods
- What’s new?: Methodology + Severity classification
- Future Plan: Supplementing diagnosis

Challenges

- Identifying speech cues
- Access for people of different socio-economic backgrounds
- Data collection
Introduction

Speech waveforms and spectograms of ALS, PD, and Healthy controls
Performance of various speech tasks in automatic classification between ALS and HC using SVM and DNN\(^1\). As seen, the spectro-temporal characteristics change depending on ALS/HC (and similarly for PD)

CNNs for identifying ALS patients\(^2\) using two 1-D convolution networks - (one each for time and frequency respectively) for filter bank features (MFBE)

---


Goal of this work

- Performance of log Mel Spectrograms in 3 class ALS/PD/HC, 5 class ALS/3 class PD severity detection
- Propose a 2D CNN approach that incorporates both Time-Frequency (TF) plane which helps us to model temporal and harmonic structures of audio signals
Overview

1. Introduction
2. Proposed approach
3. Data
4. Experiments and Results
5. Conclusions
Proposed approach

2D CNN and Feature dimensions

- The work makes use of TF plane as a whole and utilizes a 2-dimensional convolutional network for log Mel spectrograms (SPEC) and Mel frequency cepstral coefficients (MFCC).
- Due to the TF plane, the feature is better understood through a 2D CNN (finer modelling of temporal-harmonic structures).

1. SPEC dimension of $96 \times 33$ with Melbins$^1 = 96$, and a audio length of 1 second represented by 33 frames.
2. MFCC dimension of $101 \times 39$ with audio length represented by 101 frames and feature dimension of 39.

---

$^1$For the same architecture, smaller number of Melbins (24 and 48) led to lower accuracies (0.6-0.75) due to a lower resolution while a higher value of 128 gave similar accuracies when compared to 96.
Proposed approach

2D CNN Architecture

- **No. conv. filters or ‘feature maps’** = 32
- **Kernel**: A $3 \times 3$ (represented by 2D Conv)
- **Convolution layer of size** $(h \times w \times d)$ **learns ‘** $d$ **’ features of size** $h \times w$.
- **Size of pooling area**: $2 \times 2$ (represented by Max Pool 2)
2D CNN architecture (contd.)

- Activation function: ReLU (softmax @ output)
- Optimum conv. layer dropout is $= 0.5$; dense layer dropout $= 0.6$
- Loss function: Categorical cross-entropy
- Optimizer: Adadelta
Overview

1. Introduction
2. Proposed approach
3. Data
4. Experiments and Results
5. Conclusions
Data collection

- Collected from National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India
- Recorders: Apple iPhone 7 (IPH), Motorola G5 Plus (MOT), Xiaomi Redmi 4 (XIA), Zoom H6 X/Y recorder (ZOO) and Dell XPS 15 laptop (LAP)
- Sampling frequency: 44.1 kHz
Number of subjects used in this work:

- 60 ALS (30 Male, 30 Female)
- 60 PD (34 Male, 26 Female)
- 60 healthy control (HC) (30 Male, 30 Female)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gender</th>
<th>Count</th>
<th>Age Range (Avg) in years</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS</td>
<td>M</td>
<td>30</td>
<td>33 - 76 (58.60)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>30</td>
<td>38 - 75 (56.02)</td>
</tr>
<tr>
<td>PD</td>
<td>M</td>
<td>34</td>
<td>34 - 78 (58.22)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>26</td>
<td>36 - 74 (56.99)</td>
</tr>
<tr>
<td>HC</td>
<td>M</td>
<td>30</td>
<td>26 - 68 (44.21)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>30</td>
<td>31 - 65 (46.93)</td>
</tr>
</tbody>
</table>

Table: Subject count and age range for each condition - gender pair
Dataset

- Number of subjects used in this work:
  - 60 ALS (30 Male, 30 Female)
  - 60 PD (34 Male, 26 Female)
  - 60 healthy control (HC) (30 Male, 30 Female)

- Speech tasks: considering all subjects across all devices
  - Spontaneous speech (SPON) - 21 hours
  - Diadochokinetic rate (DIDK) - 22.42 hours
  - Sustained phoneme production (PHON) - 25.84 hours
  - Image Description (IMAG) - 25.22 hours
Severity Ratings

- Provided by five speech language pathologists (SLP) from NIMHANS
- Inter-rater reliability has been calculated using the Fleiss’ kappa (\( \kappa \))
  - ALS subjects, \( \kappa = 0.9017 \) (Almost perfect agreement)
  - PD subjects, \( \kappa = 0.6995 \) (Substantial agreement)

<table>
<thead>
<tr>
<th>Finding</th>
<th>Sev</th>
<th>ALSFRS-R for Speech</th>
<th>Finding</th>
<th>Sev</th>
<th>UPDRS-III for Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>4</td>
<td></td>
<td>Normal</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Detectable speech disturbance</td>
<td>3</td>
<td></td>
<td>Slight loss of expression, diction and/or volume.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Intelligible with repeating</td>
<td>2</td>
<td></td>
<td>Monotone, slurred but understandable; moderately impaired.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Speech combined with nonvocal communications</td>
<td>1</td>
<td></td>
<td>Marked impairment, difficult to understand.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Loss of useful speech</td>
<td>0</td>
<td></td>
<td>Unintelligible.</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Table: ALSFRS-R (for ALS) and UPDRS-III (for PD) scales used for rating the subjects

Overview

1. Introduction
2. Proposed approach
3. Data
4. Experiments and Results
5. Conclusions
Experiments and Results

Experimental setup

- 5 fold cross validation (each fold consists of 12 ALS, 12 PD, 12 HC)
- Proposed approach: 2D CNN: Uses ReLU activation function (except for softmax before output)
  - Features: SPEC and MFCC (computed for window length of 20ms & shift of 10ms, analysis window of 1s).
- Baseline: SVM and DNN for ALS/PD/HC
  - Features: MFCC (suprasegmental features on 1s analysis window)
  - Kernel function in SVM: Radial basis function
  - DNN: 2-hidden layers with 128/256/512 units in each layer (for which the val. loss is minimized) and output layer with three units (ALS/PD/HC) and softmax activation
Evaluation: AUC-ROC Characteristic curves

Figure: AUC-ROC curves for 2 class classification with True Positive Rate (TPR) vs False Positive Rate (FPR) on the Y & X axis

↑ AUC ⇔ better ability of the model at distinguishing between classes.

Experiments

Three sets of classification experiments are carried out:

1. 3 class ALS vs PD vs HC
2. 5 class ALS severity classification
3. 3 class PD severity classification
3 class ALS vs PD vs HC

<table>
<thead>
<tr>
<th>Speech Task/Device</th>
<th>MOT</th>
<th>ZOO</th>
<th>IPH</th>
<th>XIA</th>
<th>LAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEC</td>
<td>0.86 (0.01)</td>
<td>0.85 (0.01)</td>
<td>0.85 (0.01)</td>
<td>0.84 (0.01)</td>
<td>0.84 (0.02)</td>
</tr>
<tr>
<td>MFCC</td>
<td>0.67 (0.04)</td>
<td>0.68 (0.04)</td>
<td>0.67 (0.08)</td>
<td>0.68 (0.03)</td>
<td>0.64 (0.01)</td>
</tr>
<tr>
<td>DIDK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEC</td>
<td>0.93 (0.01)</td>
<td>0.90 (0.01)</td>
<td>0.92 (0.02)</td>
<td>0.90 (0.01)</td>
<td>0.89 (0.01)</td>
</tr>
<tr>
<td>MFCC</td>
<td>0.74 (0.04)</td>
<td>0.72 (0.05)</td>
<td>0.73 (0.06)</td>
<td>0.73 (0.01)</td>
<td>0.75 (0.03)</td>
</tr>
<tr>
<td>PHON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEC</td>
<td>0.89 (0.00)</td>
<td>0.80 (0.01)</td>
<td>0.86 (0.01)</td>
<td>0.80 (0.01)</td>
<td>0.83 (0.01)</td>
</tr>
<tr>
<td>MFCC</td>
<td>0.72 (0.04)</td>
<td>0.70 (0.05)</td>
<td>0.70 (0.05)</td>
<td>0.68 (0.02)</td>
<td>0.67 (0.07)</td>
</tr>
<tr>
<td>IMAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEC</td>
<td>0.84 (0.01)</td>
<td>0.82 (0.01)</td>
<td>0.86 (0.00)</td>
<td>0.80 (0.01)</td>
<td>0.81 (0.01)</td>
</tr>
<tr>
<td>MFCC</td>
<td>0.72 (0.01)</td>
<td>0.77 (0.04)</td>
<td>0.75 (0.03)</td>
<td>0.66 (0.03)</td>
<td>0.72 (0.01)</td>
</tr>
</tbody>
</table>

Table: Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between MFCC baseline and SPEC
Experiments and Results

AUC-ROC for 3 class ALS/PD/HC

- Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis
- Bold lines: SPEC
  Dashed lines: MFCC.
- AUC scores for MFCC are in brackets next to the SPEC values for reference
Three sets of classification experiments are carried out:

1. 3 class ALS vs PD vs HC
2. 5 class ALS severity classification
3. 3 class PD severity classification
5 class ALS severity classification

<table>
<thead>
<tr>
<th>Speech Task/Device</th>
<th>MOT</th>
<th>ZOO</th>
<th>IPH</th>
<th>XIA</th>
<th>LAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPON SPEC</td>
<td>0.76 (0.08)</td>
<td>0.73 (0.09)</td>
<td><strong>0.80 (0.09)</strong></td>
<td>0.74 (0.04)</td>
<td>0.74 (0.06)</td>
</tr>
<tr>
<td>SPON MFCC</td>
<td>0.70 (0.09)</td>
<td>0.62 (0.01)</td>
<td>0.72 (0.01)</td>
<td>0.61 (0.05)</td>
<td>0.66 (0.07)</td>
</tr>
<tr>
<td>DIDK SPEC</td>
<td>0.74 (0.01)</td>
<td>0.77 (0.07)</td>
<td><strong>0.79 (0.01)</strong></td>
<td>0.78 (0.01)</td>
<td>0.76 (0.01)</td>
</tr>
<tr>
<td>DIDK MFCC</td>
<td>0.64 (0.01)</td>
<td>0.67 (0.08)</td>
<td>0.70 (0.01)</td>
<td>0.65 (0.01)</td>
<td>0.65 (0.02)</td>
</tr>
<tr>
<td>PHON SPEC</td>
<td>0.70 (0.07)</td>
<td>0.68 (0.01)</td>
<td><strong>0.72 (0.01)</strong></td>
<td>0.71 (0.08)</td>
<td>0.70 (0.08)</td>
</tr>
<tr>
<td>PHON MFCC</td>
<td>0.64 (0.08)</td>
<td>0.55 (0.01)</td>
<td>0.65 (0.01)</td>
<td>0.60 (0.09)</td>
<td>0.58 (0.09)</td>
</tr>
<tr>
<td>IMAG SPEC</td>
<td><strong>0.77 (0.05)</strong></td>
<td>0.75 (0.06)</td>
<td>0.74 (0.07)</td>
<td>0.74 (0.08)</td>
<td>0.74 (0.01)</td>
</tr>
<tr>
<td>IMAG MFCC</td>
<td>0.67 (0.06)</td>
<td>0.66 (0.08)</td>
<td>0.72 (0.08)</td>
<td>0.70 (0.09)</td>
<td>0.68 (0.02)</td>
</tr>
</tbody>
</table>

Table: Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between SPEC and MFCC for 5 class ALS Severity classification
AUC-ROC for 5 class ALS Severity

Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis

Severity: 0: Severe (loss of useful speech) and 4: Normal speech (but has ALS)
Three sets of classification experiments are carried out:

1. 3 class ALS vs PD vs HC
2. 5 class ALS severity classification
3. 3 class PD severity classification
### Experiments and Results

#### 3 class PD severity classification

<table>
<thead>
<tr>
<th>Speech Task/Device</th>
<th>MOT</th>
<th>ZOO</th>
<th>IPH</th>
<th>XIA</th>
<th>LAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPON</td>
<td>SPEC</td>
<td>0.75(0.01)</td>
<td>0.76(0.01)</td>
<td>0.76(0.01)</td>
<td>0.77(0.06)</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>0.65(0.04)</td>
<td>0.68(0.01)</td>
<td>0.69(0.03)</td>
<td>0.63(0.01)</td>
</tr>
<tr>
<td>DIDK</td>
<td>SPEC</td>
<td>0.87(0.01)</td>
<td>0.87(0.01)</td>
<td>0.85(0.01)</td>
<td>0.85(0.00)</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>0.80(0.01)</td>
<td>0.82(0.01)</td>
<td>0.84(0.02)</td>
<td>0.80(0.01)</td>
</tr>
<tr>
<td>PHON</td>
<td>SPEC</td>
<td>0.85(0.01)</td>
<td>0.81(0.00)</td>
<td>0.82(0.01)</td>
<td>0.79(0.01)</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>0.68(0.04)</td>
<td>0.75(0.05)</td>
<td>0.62(0.08)</td>
<td>0.66(0.09)</td>
</tr>
<tr>
<td>IMAG</td>
<td>SPEC</td>
<td>0.81(0.00)</td>
<td>0.79(0.01)</td>
<td>0.81(0.01)</td>
<td>0.76(0.01)</td>
</tr>
<tr>
<td></td>
<td>MFCC</td>
<td>0.70(0.05)</td>
<td>0.74(0.04)</td>
<td>0.75(0.02)</td>
<td>0.70(0.01)</td>
</tr>
</tbody>
</table>

**Table:** Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between SPEC and MFCC for 3 class PD Severity classification.
AUC-ROC for 3 class PD Severity

Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis

Severity:
0: Normal speech (but has PD)
2: Monotone, slurred but understandable; moderately impaired
Overview

1. Introduction
2. Proposed approach
3. Data
4. Experiments and Results
5. Conclusions
Conclusions

Summary

- 2D CNN+SPEC outperforms 2D CNN+MFCC across all speech task-device pairs
- 2D CNN+MFCC in turn performs better than SVM+MFCC or DNN+MFCC for ALS/PD/HC classification
- 5 class ALS Severity: Across all severity-speech task-device combinations, the max(min) AUC score is 0.976(0.866) - good separability between severity classes
- 3 class PD Severity: Across all severity-speech task-device combinations, the max(min) AUC score is 0.951(0.807) - (which may further improve with an increase in Fleiss’ kappa ($\kappa$))
Conclusions

Key Takeaways

1. SPEC features performed better than MFCC and regardless of the recording device used, similar accuracies were obtained.

2. Severity classification (to the best of our knowledge) of a neurological disease such as ALS or PD having not been attempted earlier shows good promise in identifying the condition and also its severity at an earlier stage.
References


Acknowledgement

This work is supported by the Pratiksha Trust and the Department of Science and Technology (DST), Government of India.
THANK YOU

Have Questions/Suggestions?
Write to us at spirelab.ee@iisc.ac.in