# Low Complexity Model with Single Dimensional Feature for Speech Based Classification of Amyotrophic Lateral Sclerosis Patients and Healthy Individuals

**Anjali Jayakumar**[1], Tanuka Bhattacharjee[1], Seena Vengalil[2], Yamini Belur[2], Atchayaram Nalini[2], Prasanta Kumar Ghosh[1]

[1]SPIRE LAB, EE Dept., IISc, Bangalore, India
[2] NIMHANS, Bangalore, India

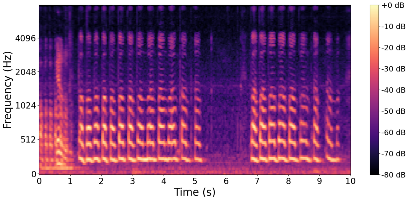
SPCOM 2024

# Overview

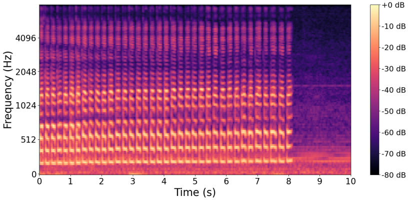
# Amyotropic Lateral Sclerosis (ALS)

- A **progressive neurodegenerative disorder** primarily affecting **motor neurons.**
- **Muscle weakness and atrophy** develop over time, affecting mobility and daily activities.
- **Dysarthria**, an early symptom of ALS, manifests as **impaired speech production.**
- As the disease advances, individuals encounter challenges in **producing sounds, modulating pitch, and maintaining proper vocal quality.**

# ALS Speech vs Healthy Speech




ALS Speech




Healthy Speech

Figure: Mel spectrogram of ALS and healthy speech: rapid repetition of monosyllabic sequence ‘pa-pa-pa’

# Motivation

- Investigate the use of **speech cues** to distinguish between ALS and healthy individuals.
- Develop **simple models** suitable for deployment on accessible platforms.

# Our Objective

- Develop **low complexity** methods for **ALS and Healthy Control (HC) classification**.
- Investigate three distinct **Deep Neural Network (DNN) models with varying complexities**, comparing them against a Convolutional Neural Network (CNN) - Bidirectional (Bi) Long Short-Term Memory (LSTM) reference model.
- Analyze 12-D **Mel Frequency Cepstral Coefficients (MFCCs), derivatives, and individual coefficients** to capture essential features.
- Explore various **temporal statistics**
  - Standard deviation (SD)
  - Autocorrelation at lag 1 (AC(1))
  - Autocorrelation at lag 2 (AC(2))

# Literature

| | | |
|---|---|---|
| **CNN-LSTM/BiLSTM** | J. Mallela et al., "Raw speech waveform based classification of patients with ALS, Parkinson's disease and healthy controls using CNN-BLSTM," in Proc. 21st Annual Conference of the International Speech Communication Association, Shanghai, China, 2020, pp. 4586–4590. | **#Param=1321832 FLOPs=2400000** |
| | J. Mallela et al., "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in ICASSP, IEEE, 2020, pp. 6784–6788. | **#Param=79,104 FLOPs=144100** |

**MFCC**

- MFCCs are known for their **low complexity and effectiveness**[*] in speech-related tasks.
- **Temporal statistics** such as mean, median and SD of MFCC are utilized for ALS-induced dysarthria diagnosis and severity assessment.
- **Individual coefficients** of 12-D MFCC has been analyzed for EMG based ALS diagnosis[**] and other speech tasks, but **not for speech based ALS vs HC classification**.

[*] A. A. Joshy and R. Rajan, "Automated dysarthria severity classification: A study on acoustic features and deep learning techniques," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1147–1157, 2022.

[**] A. B. M. S. U. Doulah and S. A. Fattah, "Neuromuscular disease classification based on mel frequency cepstrum of motor unit action potential," in International Conference on Electrical Engineering and Information Communication Technology, 2014, pp. 1–4.

# Overview

# Subject Details

- Demographic details

| Class | #Male | #Female | Mean(SD) Age (years) |
|---|---|---|---|
| **ALS** | 46 | 26 | 55.36 (10.80) |
| **HC** | 40 | 15 | 46.62 (6.85) |

- Recording conducted at **National Institute of Mental Health and Neurosciences, Bengaluru, India**.
- Subjects spoke **Bengali, Kannada, or Telugu** as native languages.

# Recording Details

SPIRE LAB

- Speech tasks:
  - **Spontaneous (SPON)**: Subjects describe a festival and a recent place they visited in their native language (approx. 1 minute each).
  - **Diadochokinetic rate (DIDK)**: Subjects rapidly repeat mono-syllabic or tri-syllabic sequences (e.g., 'pa-pa-pa', 'ta-ta-ta', 'ka-ka-ka', 'pataka', 'badaga').
- Recording duration

| Task | Class | Mean (SD) Duration (sec) | Total Duration (min) |
|---|---|---|---|
| SPON | ALS | 60.75 (18.15) | 138.7 |
| | HC | 59.06 (20.78) | 107.28 |
| DIDK | ALS | 15.92 (9.44) | 93.65 |
| | HC | 18.80 (8.33) | 85.85 |

J. Mallela et al., "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 6784–6788.

# Overview

# Classification Pipeline

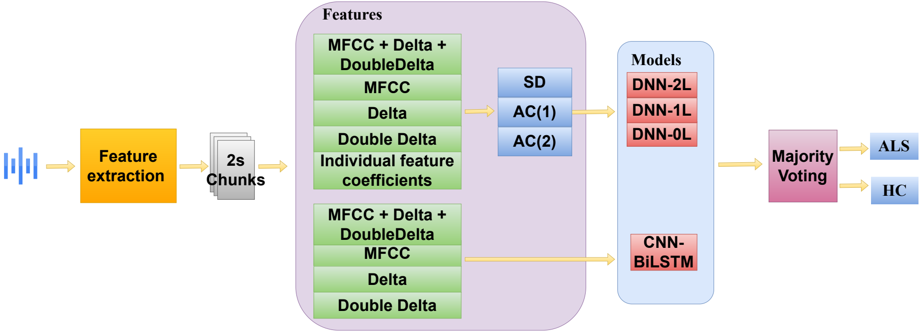



Figure: ALS vs HC classification methodology

# Methods for Complexity Reduction

## 1. Model Complexity Reduction

- Transitioned from CNN-BiLSTM to **simpler DNN models**.
- Decreased model complexity: Lower **number of parameters (#params) and Floating Point Operations (FLOPs)**.
- Simplifying with DNNs:
  - **DNN-2L:** 2 dense layers
  - **DNN-1L:** 1 dense layer
  - **DNN-0L:** 0 dense layer

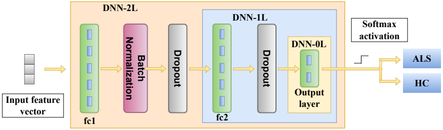



Figure: DNN models of varying complexity; here, fc1 and fc2 are fully connected layers with 128 units each, and output layer is a dense layer with 2 units

# Methods for Complexity Reduction

## 2. Feature Dimensionality Reduction

| Model | Feature | Dimension |
|---|---|---|
| CNN-BiLSTM | MFCC + delta + double delta (36D Matrix) | 36 |
| | MFCC Matrix | 12 |
| | Delta Matrix | 12 |
| | Double delta Matrix | 12 |
| DNN | SD, AC(1), and AC(2) of: | |
| | 36D Matrix | 36 |
| | MFCC Matrix | 12 |
| | Delta Matrix | 12 |
| | Double delta Matrix | 12 |
| | Individual coefficients of 36D Matrix | 1 |

# Model Complexity

Table: Model complexity for different models and feature dimensions

| Model | Feature dimension | #params | FLOPs |
|---|---|---|---|
| CNN-BiLSTM* | 36 | 1321832 | 2400000 |
| | 12 | 1307462 | 2380000 |
| DNN-2L | 36 | 22018 | 21500 |
| | 12 | 18946 | 18430 |
| | 1 | 17538 | 17020 |
| DNN-1L | 36 | 4994 | 9730 |
| | 12 | 1992 | 3580 |
| | 1 | 514 | 768 |
| DNN-0L | 36 | 74 | 9540 |
| | 12 | 26 | 3350 |
| | 1 | 4 | 514 |

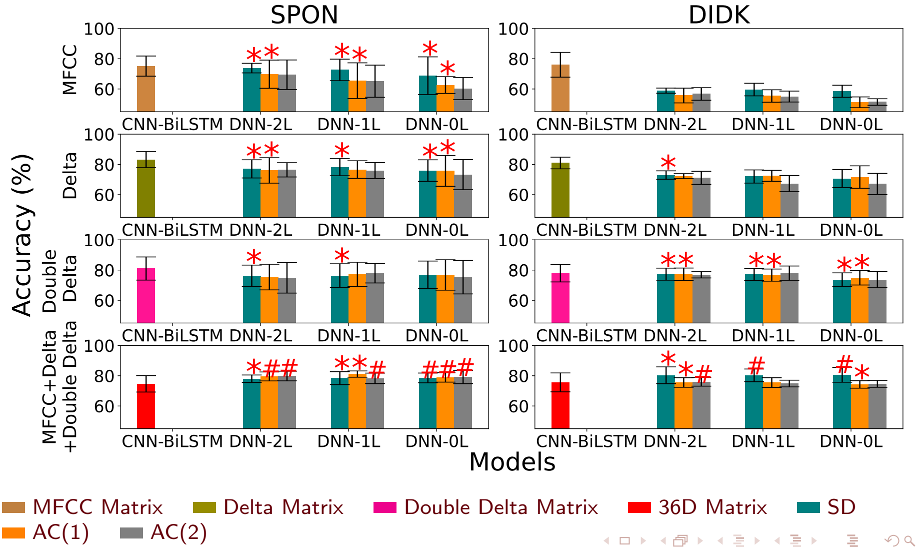
* J. Mallela et al., "Raw speech waveform based classification of patients with ALS, Parkinson's disease and healthy controls using CNN-BLSTM," in Proc. 21st Annual Conference of the International Speech Communication Association, Shanghai, China, 2020, pp. 4586–4590.

# Training and Evaluation

- Trained using **5-fold cross validation** method
  - 3 folds for training, 1 for validation and 1 for testing
- **Mean and SD of balanced accuracy scores** computed over 5 test sets reported.
- Wilcoxon signed-rank test (1% significance level) used to compare classification accuracies across different feature and model configurations.

# Overview

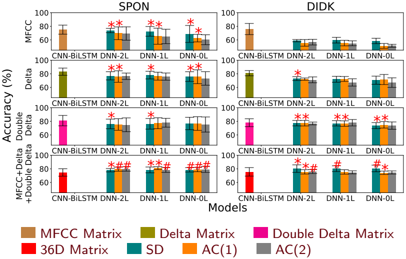
# Model Complexity Reduction



# Model Complexity Reduction

Table: Comparison with baseline model

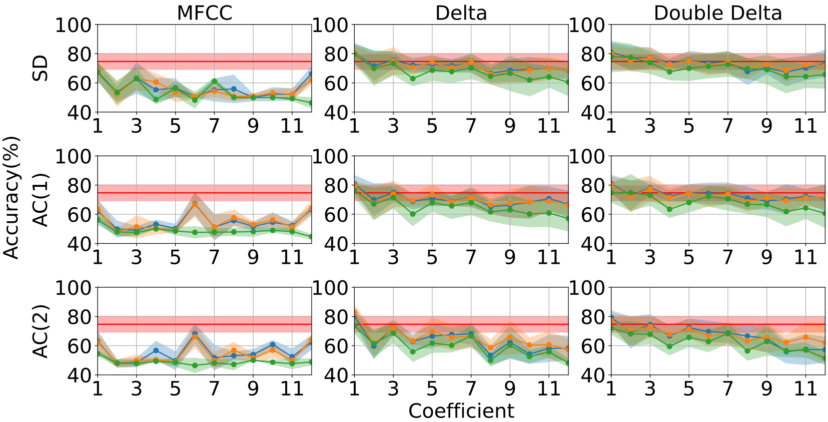
| | SPON Task | DIDK Task |
|---|---|---|
| **Best Configuration** | DNN-0L using AC(2) of 36D vector | DNN-0L using SD of 36D vector |
| **Reduction in #Param** | 99.99% | 99.99% |
| **Reduction in FLOPs** | 99.60% | 99.60% |
| **Mean Balanced Accuracy** | Increase: 5.67% | Increase: 6.69% |
| **#Configurations Statistically Similar to Baseline** | 16/36 | 10/36 |
| **#Configurations Outperforming Baseline** | 6/36 | 3/36 |




On average, **SPON task outperforms DIDK task** in mean accuracy score by:

- 4.84% for DNN model.
- 0.75% for CNN-BiLSTM model.

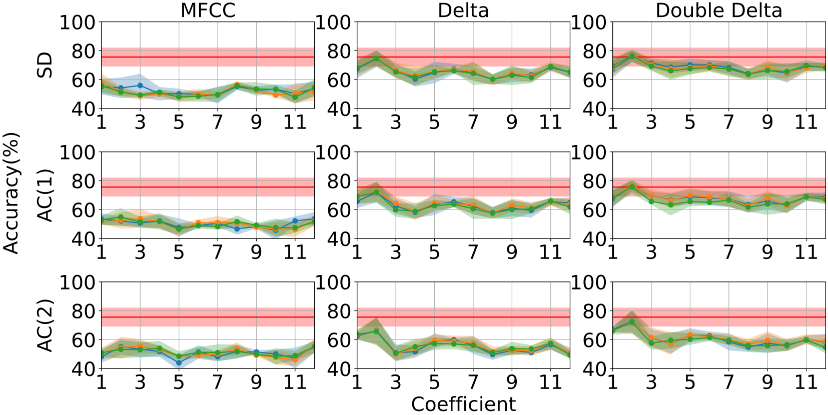
# Feature Dimensionality Reduction




SPON

CNN-BiLSTM (36D Matrix) (#params=1321832) DNN-2L (#params=17538) DNN-1L (#params=514) DNN-0L (#params=4)

# Feature Dimensionality Reduction




DIDK

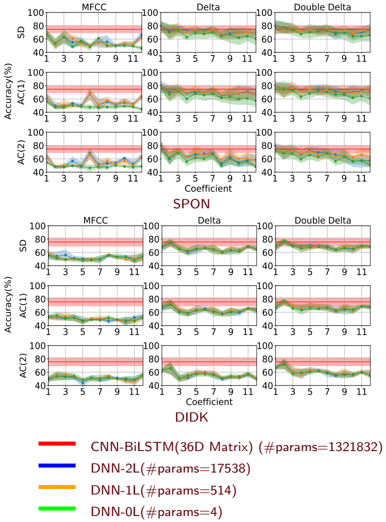
CNN-BiLSTM (36D Matrix) (#params=1321832) DNN-2L (#params=17538) DNN-1L (#params=514) DNN-0L (#params=4)

# Feature Dimensionality Reduction

Table: Comparison with the best configuration of model complexity reduction

| | SPON Task | DIDK Task |
|---|---|---|
| **Best Configuration** | DNN-0L using SD of first delta coefficient | DNN-0L using SD of second delta coefficient |
| **Reduction in #Param** | 94.59% | 94.59% |
| **Reduction in FLOPs** | 94.61% | 94.61% |
| **Mean Balanced Accuracy** | Decrease: 1.76% | Decrease: 5.17% |

- On average, **SPON task outperforms DIDK task** in mean accuracy score by 4.24% for DNN model using individual coefficients.
- First delta and double delta coefficients excel in SPON task; second in DIDK task.



# Overview

# Key-Takeaways

- DNN models perform comparably to CNN-BiLSTM with fewer resources.
- Simplifying DNN complexity maintains high performance while reducing resource usage.
- DNN-0L with a mere softmax matches CNN-BiLSTM for lower delta and double delta coefficients.
- Standard deviation and auto-correlations offer similar performance, allowing flexible feature selection.
- DNN models show promise for ALS vs HC classification under resource constraints.
- Lower delta and double delta coefficients are crucial for accurate and efficient classification.

# Future Work

- Extending the study to various datasets.
- Exploration of additional methodology to further enhance the accuracy of the classification task.

# Acknowledgement


- We sincerely thank all the subjects who contributed to the speech dataset.
- We express our sincere gratitude to the Department of Science and Technology (DST), Government of India for supporting this work.

**THANK YOU**

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