



# Detection Of TDP-43 Using Digital Acoustic Biomarkers And Machine Learning Approaches



Pooyan Mobtahej<sub>1</sub>, Jacob Thomas<sub>1</sub>, Thiri Htet Naing<sub>1</sub>, Jamin Kim<sub>1</sub>, Maria M. Corrada<sub>1</sub>, S Ahmad Sajjadi<sub>1,2</sub>  
(1) University of California, Irvine, Irvine, CA, USA, (2) Institute for Memory Impairments and Neurological Disorders (UCI-MIND).

## Introduction

- The Need for Biomarkers:** Aggregation of TDP-43 in limbic structures causes amnesic cognitive impairment in the older population.
- There are currently no established biomarkers for detecting limbic predominant age-related TDP-43 encephalopathy (LATE).
- Acoustic Biomarkers From Speech:** We studied the utility of acoustic features from spontaneous speech for the detection of LATE neuropathological changes. We used machine learning models.

## Methods

### Participants from The 90+ Study:

- Participants were video-recorded during clinical interviews to collect spontaneous speech samples.
- Audio from baseline visit samples were extracted and processed using Audacity to reduce background noise and remove interviewer's speech.

### Acoustic Features:

- Mel-frequency cepstral coefficients (MFCC), spectral centroid (SC), zero-crossing rate (ZCR), root mean square (RMS), and chroma.
- These features were extracted and scaled from the audio files to analyze speech brightness, quality, energy, and pattern changes.

### Machine Learning Models:

- Logistic Regression, Support Vector Classifier (SVC; part of Support Vector Machine (SVM)), K-Nearest Neighbor (KNN), Extra Trees, and Gradient Boosting.
- The acoustic features were analyzed for detection of TDP-43.
- These models were selected for performance comparison and trained on 80% of the dataset. Each model's performance was evaluated using a test data consisting of unseen participants who were not included in the training set (Figure 1).

Table 1. Participant Demographics

|                                       | TDP-43                                  | No TDP-43              | Total                        |
|---------------------------------------|-----------------------------------------|------------------------|------------------------------|
| N (%)                                 | 31(27.7%)                               | 81(72.3%)              | 112                          |
| Sex (F/M)                             | 53/28<br>(72.6%/71.8%)                  | 20/11<br>(27.4%/28.2%) | 73/39<br>(65.1%/34.8%)       |
| Age (years)                           | 97.52 ± 0.86                            | 97.94 ± 3.68           | 97.64 ± 3.30                 |
| Education (college graduate vs. less) | 38/43<br>(66.7%/78.2%)                  | 19/12<br>(33.3%/21.8%) | 57/55<br>(50.9%/49.1%)       |
| Cognitive Diagnosis                   |                                         |                        |                              |
| Normal                                | Cognitive Impairment No Dementia (CIND) | Dementia               | Not given (Fatigue/Not Done) |
| 59<br>(52.7%)                         | 25<br>(22.3%)                           | 26<br>(23.2%)          | 1/1<br>(0.9%/0.9%)           |

Table 2. Model's Performances

|                     | Accuracy | F1 score | AUC    |
|---------------------|----------|----------|--------|
| Logistic Regression | 0.8260   | 0.7142   | 0.7583 |
| SVC (SVM)           | 0.6086   | 0.5833   | 0.6583 |
| KNN                 | 0.7391   | 0.6667   | 0.7000 |
| Gradient Boosting   | 0.7391   | 0.6250   | 0.7750 |
| Extra Trees         | 0.6957   | 0.5333   | 0.6792 |

## Results

### Participant Demographics (Table 1):

- 112 participants
- 81 participants had no TDP-43 at autopsy (LATE-NC stage 0-1)
- 31 had TDP-43 present (LATE-NC stage 2-3)
- 73 participants (65.1%) were female
- Mean age of 97.64 years at death

### Performance Results (Table 2):

- Logistic Regression showed the best detection results:
- F1 score = **71.4%**
- Accuracy score = **82.6%**
- AUC = **75.8%**
- Followed by KNN and Gradient Boosting

## Conclusion

- Analysis of acoustic features of speech holds promise as a novel and noninvasive approach for the detection of LATE.
- Availability of a reliable biomarker for LATE could substantially improve the clinical management by offering clinicians an opportunity to diagnose the disease in its initial stages as opposed to post-mortem detection.
- Integrating both acoustic and linguistic features of speech and using a larger sample size might further enhance the accuracy of detection models.
- Speech biomarkers could also be useful for detection of Alzheimer's Disease Neuropathologic Change (ADNC).

## Acknowledgment

This work was supported in part by the National Institutes of Health (R01 AG062706), R01 AG021055, the UCI Institute for Memory Impairments and Neurological Disorders (IRPA-01-2024), the Institute for Precision Health at UCI, UCI ADRC (P30 AG066519), and NIH TITAN T32 Award at UCI MIND.

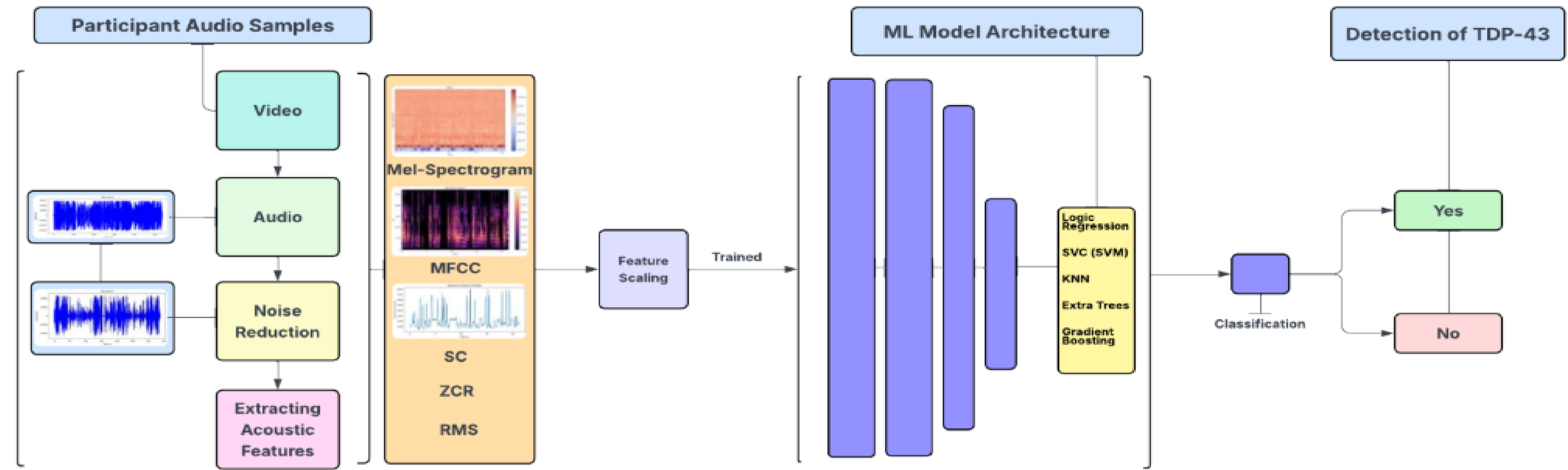


Figure 1: Machine Learning Model Overview for Detection