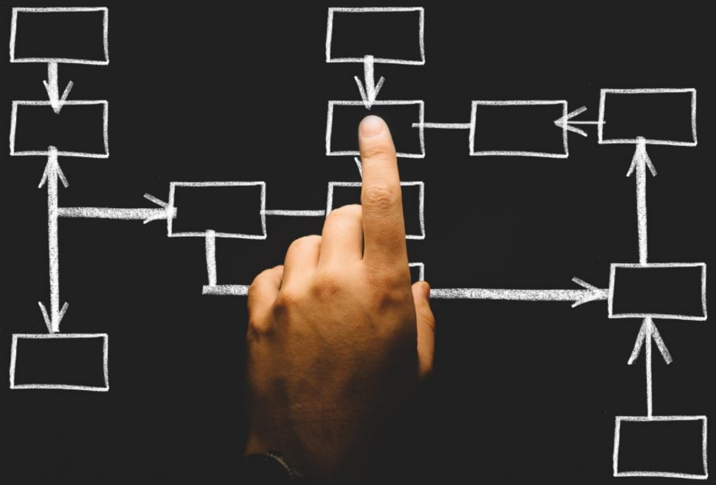
Artificial Intelligence for Dysarthria Assessment in Children With Ataxia: A Hierarchical Approach

Tartarisco et al.

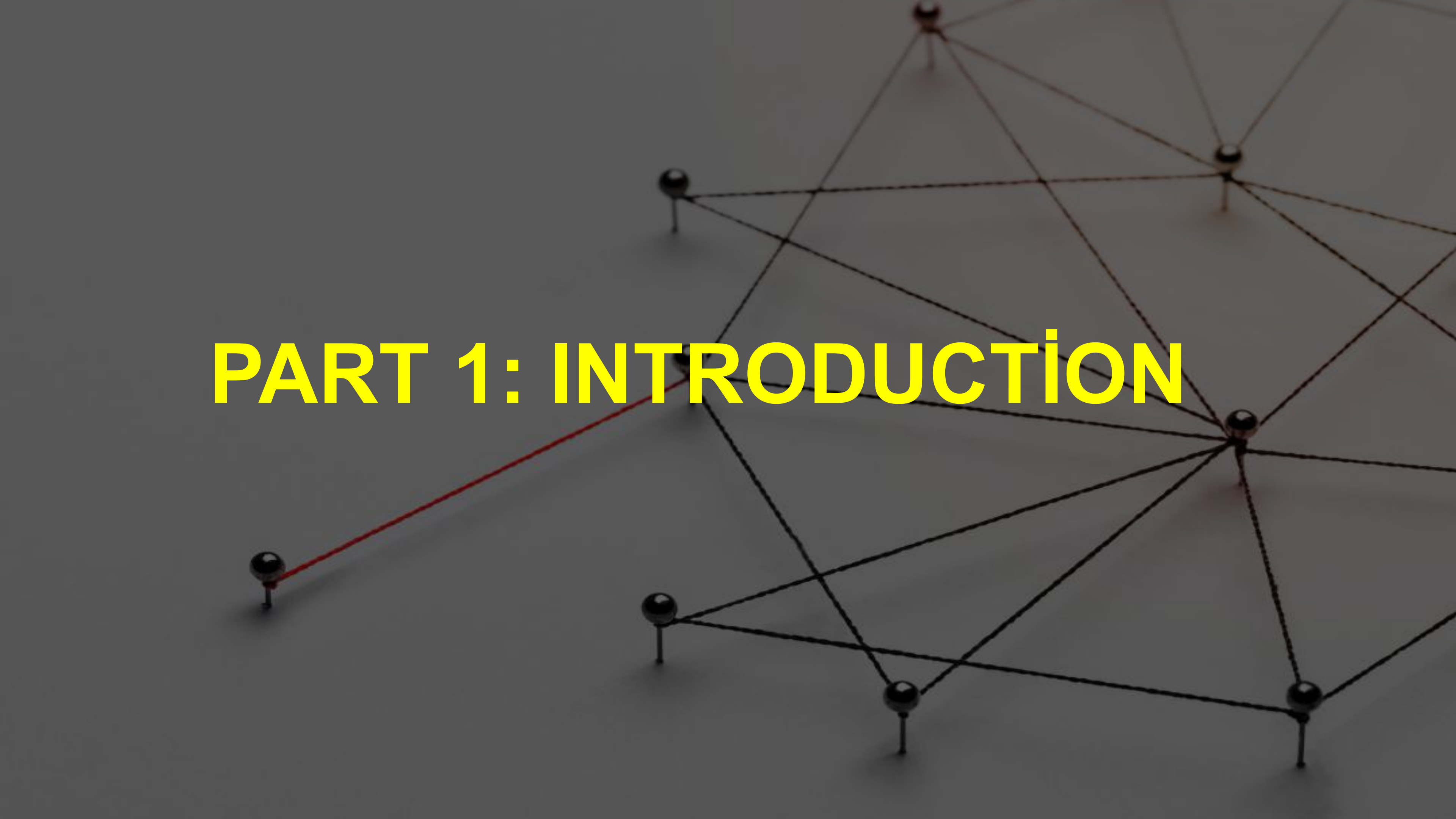
Presented by F.San Koktas & Jyri Korhonen

Overview

1. Introduction
2. Methods
   a. Methods
   b. Data collection
   c. Signal Processing
   d. Feature Extraction for Machine Learning
   e. Feature Extraction for Deep Learning
   f. Classification
   g. Performance metrics
3. Results
4. Discussion
5. Limitations
6. Conclusion
   a. Main Findings
   b. Future research
7. Assignments



PART 1: INTRODUCTİON



Introduction (i)

- Ataxia → *a-taxis* → *non-order/coordination*
  - Neurological disease
  - Poor muscle control
  - Diagnosis (for speech):
    - Expert listener rates 21 parameters
      - Rater variability
      - Loss of accuracy
  - Rating:
    - The Scale for the Assessment and Rating of Ataxia (SARA)
  - No cure, no medicine to treat symptoms

- Dysarthria → dys (difficulty) + arthron (articulation) → difficulty speaking
  - Caused by weak speech muscles

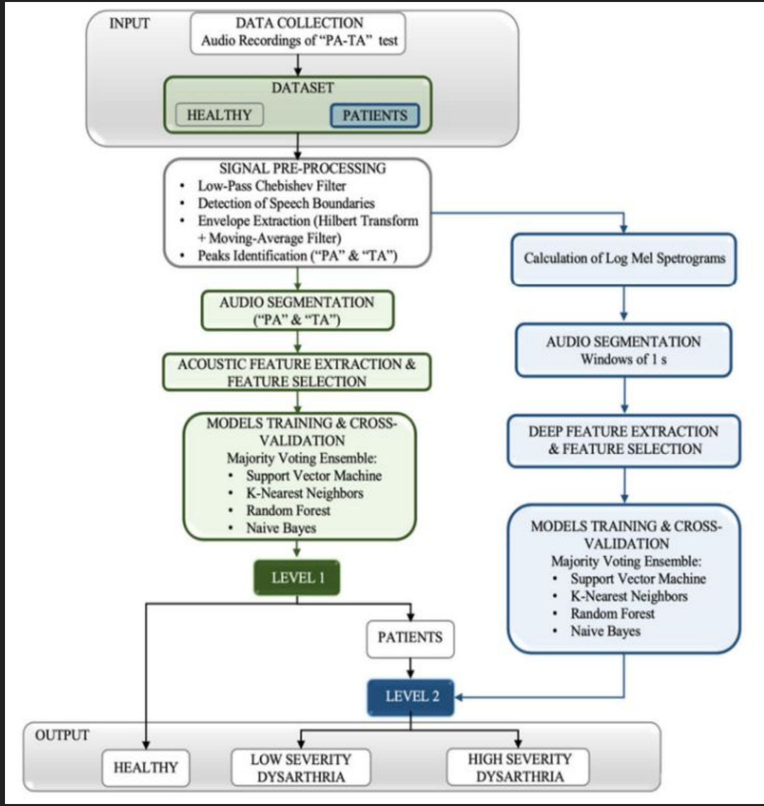
Introduction (ii)

- Why do we need a AI-powered method to diagnose ataxia?
    - For some patients, it takes so long to be diagnosed with traditional methods.
    - Expert listener might sometimes be wrong.
        - It is especially quite hard to manually rate the kids.
    - Every expert listener rates the parameters differently (rater variability).
    - For monitoring, patients need to go to clinic frequently, which is inconvenient.
        - Especially for ataxia patients in wheelchair who don’t live close to the hospital, traveling each time for monitoring must be so annoying.
        - In countries where free healthcare is not provided, frequent visits may be costly.

PART 2: METHODS

Methods

- Hierarchical machine learning model (HMLM)
  - “PATA” test recordings
  - Healthy vs Low severity vs High severity
- Level 1
  - Machine Learning
  - Healthy vs patients
- Level 2
  - Deeplearning
  - Assess severity

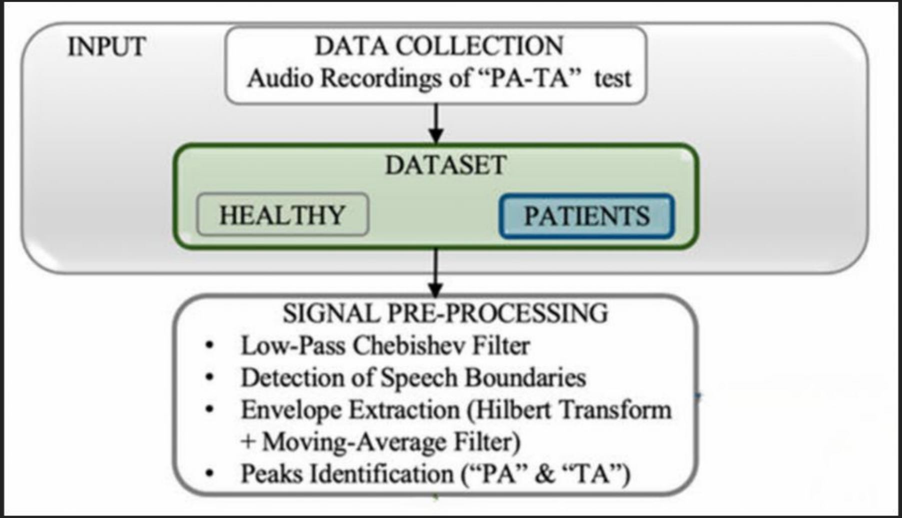


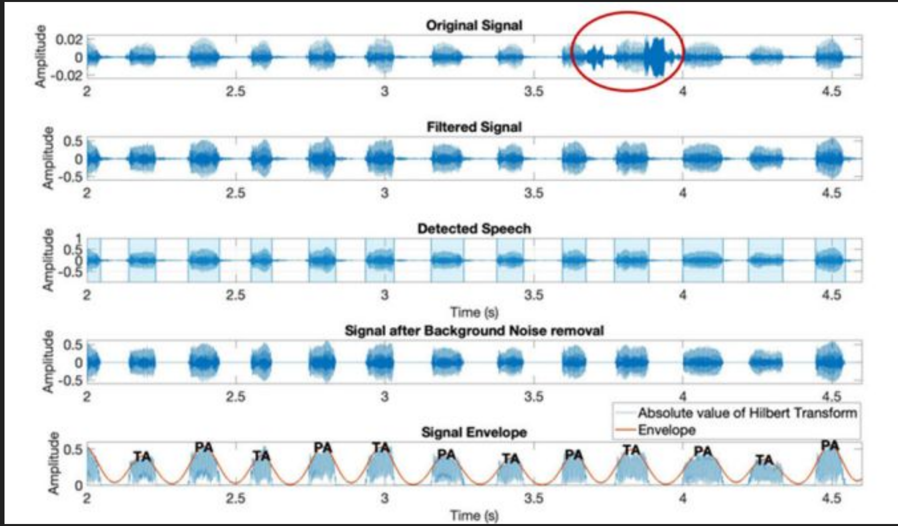
Data collection

- 55 subjects
  - 18 healthy (H)
  - 21 progressive ataxia (PA)
  - 16 congenital non progressive ataxia (CA)
- Each performed “PATA” test for 10 seconds
- Test was repeated for 21 patients with PA or CA after 12 months
- Total of 76 audio recordings
- The patients were scored using standardized clinical scale by experts
  - Scale for the Assessment and Rating of Ataxia (SARA)

Signal processing

- Reduce background noise
  - Patients voice is mostly under 1 kHz
  - Low-Pass Chebishev filter with 1 kHz cut-off
  - Hanning window
- Detection of Speech Boundaries
  - threshold short-term energy
  - spectral spread
- Envelope Extraction
  - Hilbert Transform
  - zero-phase moving-average filter
- “PA” & “TA” peaks Identification
  - Detected from the envelope

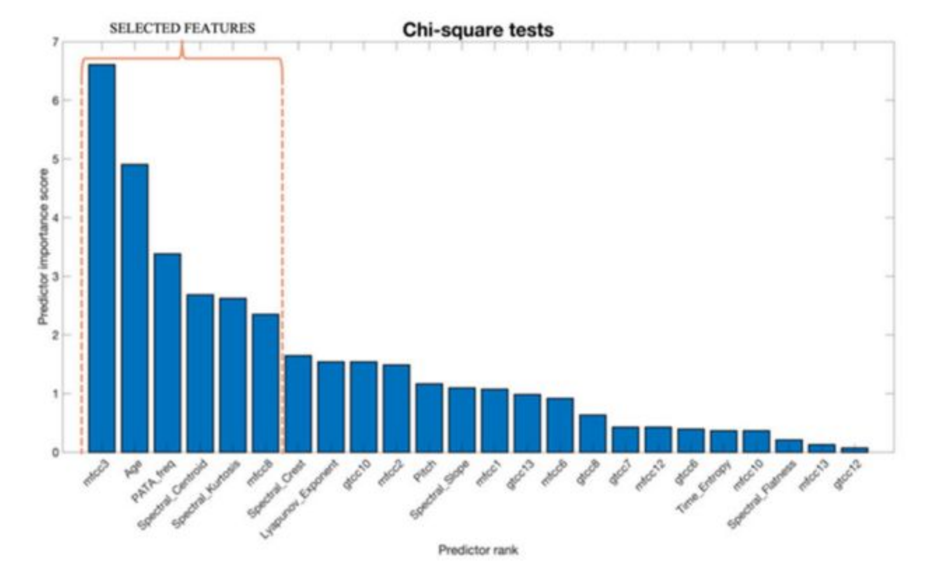




Feature extraction for Machine Learning

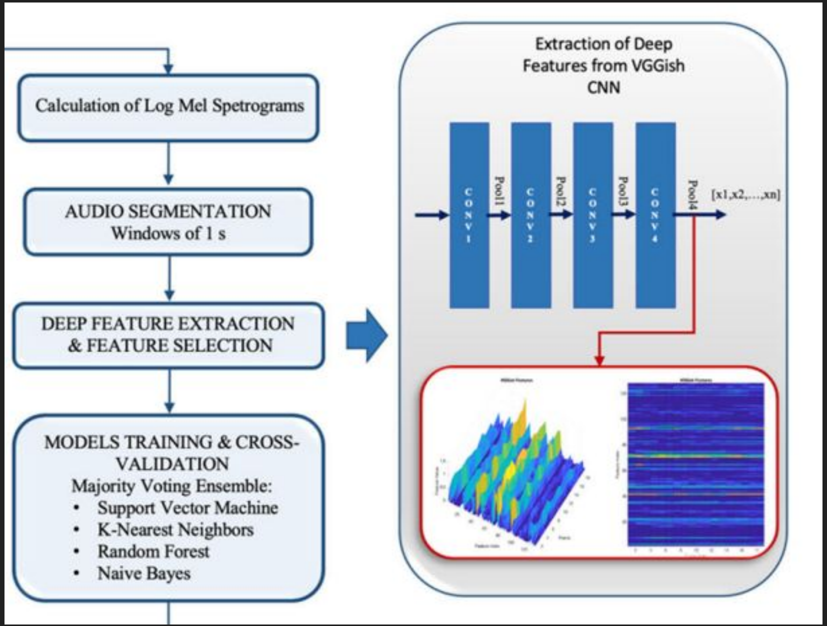
Features were ranked according to the predictor importance score and optimal subset was chosen

- Mel-Frequency Cepstral Coefficients (MFCCs)
- Subject age
- PATA frequency
- Spectral Centroid
  - Indicates where most of signal energy is contained
- Spectral Kurtosis
  - Measure of the flatness of the spectrum around its mean value



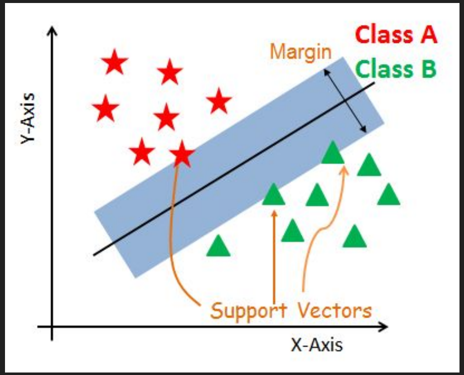
Feature extraction for Deep Learning

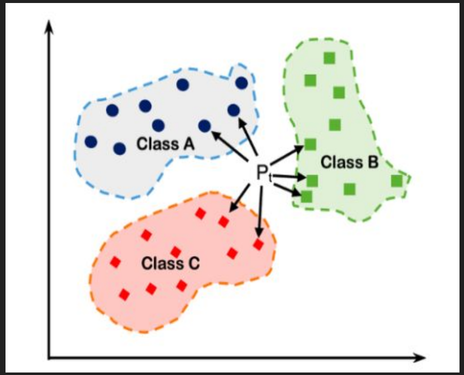
- Training Deep Learning Network requires a large amount of data
- Transfer Learning with Feature Extraction
- Pre-trained VGGish Convolutional Neural Network
  - Developed by Google for audio classification tasks
- Out of 12288 extracted features, 1444 were selected with addition of age and PATA frequency

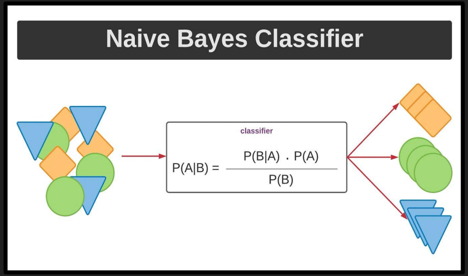


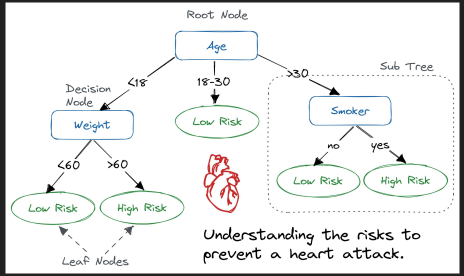
Classification

- Two binary classifiers
  - Level 1 healthy vs patient
  - Level 2 low vs high severity
- Cross-validation
- Majority voting ensemble
  - Support Vector Machine (SVM)
  - k-Nearest Neighbours (k-NN)
  - Naïve Bayes (NB)
  - Decision Tree









Performance metrics

- Accuracy
- Precision
- Recall
- F1-Score

Measure	Binary Classification	Multi-class Classification	Hierarchical Classification
Accuracy	$\frac{tp + tn}{tp + tn + fp + fn}$	$\frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l}$	$\frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l}$
Precision	$\frac{tp}{tp + fp}$	$P_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fp_i)}$ $P_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}$	$P_\downarrow = \frac{\lvert C_\downarrow^c \cap C_\downarrow^d \rvert}{\lvert C_\downarrow^c \rvert}$
Recall	$\frac{tp}{tp + fn}$	$R_\mu = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fn_i)}$ $R_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}$	$R_\downarrow = \frac{\lvert C_\downarrow^c \cap C_\downarrow^d \rvert}{\lvert C_\downarrow^d \rvert}$
F1-Score	$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$	$F1S_\mu = 2 \cdot \frac{P_\mu \cdot R_\mu}{P_\mu + R_\mu}$ $F1S_M = 2 \cdot \frac{P_M \cdot R_M}{P_M + R_M}$	$F1S_\downarrow = 2 \cdot \frac{P_\downarrow \cdot R_\downarrow}{P_\downarrow + R_\downarrow}$

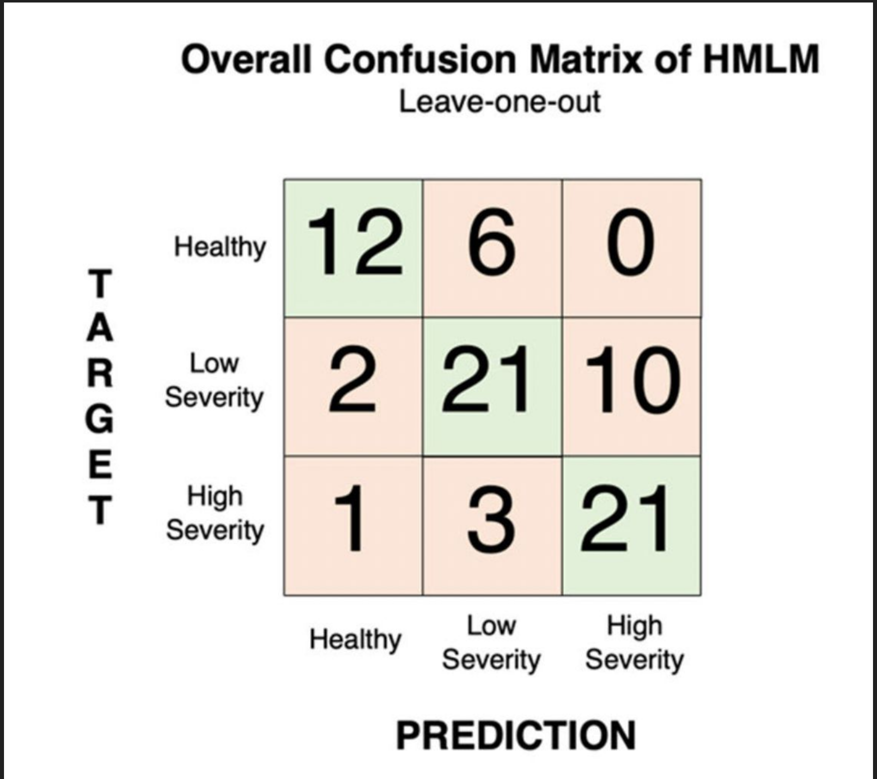
PART 3: RESULTS

Results (i)

TABLE 5. Hierarchical approach performance metrics [Level 1: Healthy vs Patients – LEVEL 2: Low Severity vs High Severity].

Measure	Machine Learning			Transfer Learning			Combination: Machine Learning (Level 1) + Transfer Learning (Level 2)		
	5-FOLD	10-FOLD	LEAVE-ONE-OUT	5-FOLD	10-FOLD	LEAVE-ONE-OUT	5-FOLD	10-FOLD	LEAVE-ONE-OUT
Accuracy	1: 88.08% 2: 61.02% Total: 57.89%	1: 86.84% 2: 51.72% Total: 51.32%	1: 88.16% 2: 60.66% Total: 56.58%	1: 78.95% 2: 75% Total: 60.53%	1: 84.11% 2: 77.42% Total: 65.79%	1: 84.21% 2: 78.33% Total: 67.11%	**1: 93.42% 2: 78.69% Total: 76.32%**	**1: 88.16% 2: 75.44% Total: 69.74%**	**1: 88.16% 2: 78.69% Total: 71.05%**
Precision	1: 91.53% 2: 58.52% Total: 73%	1: 91.38% 2: 48.95% Total: 68.48%	1: 90.16% 2: 56.37% Total: 72.37%	1: 87.50% 2: 79.18% Total: 69.96%	1: 87.10% 2: 77.81% Total: 74.73%	1: 88.33% 2: 78.70% Total: 75.66%	**1: 93.44% 2: 78.55% Total: 84.67%**	**1: 92.98% 2: 75.62% Total: 78.93%**	**1. 90.16% 2. 78.87% Total: 78.61%**
Recall	1: 93.10% 2: 55.36% Total: 73%	1: 91.38% 2: 49.02% Total: 68.48%	1: 94.83% 2: 53.66% Total: 72.37%	1: 84.48% 2: 75.52% Total: 69.96%	1: 93.10% 2: 77.81% Total: 79.28%	1: 91.33% 2: 79.86% Total: 75.66%	**1: 98.28% 2: 79.50% Total: 84.67%**	**1: 91.38% 2: 76.60% Total: 78.93%**	**1. 94.83 2. 80.24% Total: 78.61%**
F1-Score	1: 92.31% 2: 56.90% Total: 73%	1: 91.38% 2: 48.98% Total: 48.95%	1: 92.44% 2: 54.98% Total: 72.37%	1: 85.96% 2: 75.26% Total: 69.96%	1: 90% 2: 78.54% Total: 74.73%	1: 89.33% 2: 79.28% Total: 75.66%	**1: 95.80% 2: 79.02% Total: 84.67%**	**1: 92.98% 2: 75.62% Total: 78.93%**	**1. 92.44 2. 79.55% Total: 78.61%**

Results (ii)


Overall Confusion Matrix of HMLM
Leave-one-out
TARGET
Healthy
Low Severity
High Severity
12
6
0
2
21
10
1
3
21
Healthy
Low Severity
High Severity
PREDICTION

Results (iii)

	Hierarchical App. (ML+DL)	Flat-multi class App.
Level 1 Accuracy	about 90%	-
Level 2 Accuracy	about 80%	-
Overall accuracy	about 76%	about 65%
Overall precision, recall, and f1 score	about 85%	about 55%

PART 4: DISCUSSION

Discussion

- Conventional features work better at the 1st Level (healthy vs patient)
- Transfer learning based method was more suitable to assess the severity of dysarthria
- HMLM outperforms flat multi-class approach
  - overall accuracy 76.32% vs 65.58% for 5-fold cross-validation


PART 5: LIMITATIONS

Limitations

- Lack of variability of speech disturbance score
  - None of the patients had a higher severity than 3 while the highest number is 6.

- Small number of available subjects
  - Ataxic symptoms are very rare (26/100,000).
  - Still, the collected dataset was the first and biggest.

- Sometimes same clinical scores were used for different classes (e.g. healthy & low severity)
  - Reason of most of the errors.
  - It’s really hard for doctors to tell the difference between subtle changes in speech problems.

PART 6: CONCLUSIONS

Main Findings

- Voice and speech can help doctors to track ataxia at every stage of the disease (early or advanced).

- Patients can record their voices at home using phones. Doctors can track patients condition even though they don’t come to clinic.
    - This can reduce cost and make patients’ and their families’ lives easier.

- Since AI objectively evaluates the condition, doctors can have more reliable results.
    - Manual evaluations conducted by humans can vary due to individual differences

- The model described in the study can be a useful tool to screen for ataxia and monitor patients by just recording their voices.

Future Research

- Larger dysarthric speech databases are needed.
  - TORGO and NEMOURS have no more than 15 subjects.
- For AI-powered approach to fully replace manual approach, increasing the accuracy even more is needed.
- A better scale than SARA might be beneficial to combat the issues caused by different classes having the same rating from time to time.
- Model needs to be validated on greater number of subjects.

PART 7: ASSIGNMENT

Assignment

1. Please listen to the speech patterns of this individual with ataxia, then contrast them with your own. Identify at least three distinctions you observe in how they speak compared to yourself.

2. Watch this video about ataxia. By combining the experience of Tallullah and Dan with your own thoughts on ataxia, explain why AI-powered ataxia diagnosis might be useful for patients.

3. In one sentence, explain why did the researchers of this paper use low-pass Chebishev filter for pre-processing?

Thanks for listening

We would be happy to answer your questions about the paper.

Citations

1. Tartarisco, Gennaro & Bruschetta, Roberta & Summa, Susanna & Ruta, Liliana & Favetta, Martina & Busa, Mario & Romano, Alberto & Castelli-gattinara, Guido & Flavia, Marino & Cerasa, Antonio & Schirinzi, Tommaso & Petrarca, Maurizio & Bertini, Enrico & Vasco, Gessica & Pioggia, Giovanni. (2021). Artificial Intelligence for Dysarthria Assessment in Children With Ataxia: A Hierarchical Approach. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3135078.

Diagrams in slides 6-11, charts in slides 13-14, as well as all the information on this paper are taken from this paper.