Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks

Research departments in publication

1 Speech Disorders & Technology Lab, Department of Bioengineering

2 Callier Center for Communication Disorders, University of Texas at Dallas, United States

3 Department of Communication Sciences and Disorders

MGH Institute of Health Professions, United States

4 Department of Speech-Language Pathology, University of Toronto, Canada

5 MDA/ALS Center, Texas Neurology, United States

ALS Key objective of the paper

- Feasibility of automatic detection of patients with ALS at an early stage from highly intelligible speech
- The diagnosis and treatment of ALS will be significantly strengthened when objective, sensitive markers for the disease can be identified

Process

- A speech dataset was collected
- Thirteen newly diagnosed patients with ALS
- Thirteen age-gender matched healthy individuals
- Convolutional Neural Networks (CNNs), including time-domain CNN and frequency domain CNN
- Used to classify intelligible speech produced by patients with ALS and those by healthy individuals

Best sample-level sensitivity and specificity were obtained by time-CNN (71.6% and 80.9%, respectively)

The best result was obtained by frequency-CNN (76.9% sensitivity and 92.3% specificity)

ALS

What is ALS

- Rapidly progressive neurodegenerative disease
- Impairment of speech and swallowing functions
- Amyotrophic lateral sclerosis
- Fatal and progressive motor neuron disease

Current problems of diagnosis

- Provisional, primarily based on clinical observations of upper and lower motor neuron damage in the absence of other causes
- Due to the lack of clinicopathologic markers of ALS, patients are often misdiagnosed (up to 45% of the time)
- Diagnosis delayed for up to 12 months
- Consequences of delay is- time of diagnosis, a patient’s motor neurons may have been affected already

Related work

- The automatic detection of other neurological diseases from speech signals recently has shown promising results for depression [10, 11, 12], traumatic brain injury [13], and Parkinson’s disease [14, 15, 16, 17].
- Various types of acoustic features, such as formant centralization ratio, vowel space area, intonation, and prosody [18, 19] have been used for the detection of neurological diseases

Previous/ preliminary studies

Our previous preliminary study demonstrated that speech may be a sensitive measure to automatically detect ALS at early stage and monitor disease progression

Dataset

- Previous work on early detection of ALS [6], however, used a dataset that contains the non-age-matched healthy controls, which may introduce a bias in the classification performance.
- In comparison to younger speakers, elderly individuals may have slower and breathy voice characteristics.
- In addition, both speakers with ALS and senior-aged individuals show slow speech and breathy voice.
- Thus, further work with an age- and gender-matched dataset is needed to verify the previous findings.

Dataset

- In addition to predefined, hand-crafted features and artificial neural network (ANN) that were used in our previous work
- [6], we applied convolutional neural network (CNN)-based representation learning in the current work. Representation learning can learn useful features from low-level signals and it has
- Shown effectiveness in various classification applications, outperforming traditional hand-crafted features [21, 22].

Dataset

- The speech data set used in this study was collected from thirteen early-diagnosed patients with ALS (9 females and 4 males) and thirteen healthy age-matched speakers (8 females and 5 males).
- The patients with ALS were diagnosed within 6-12 months prior to data collection.
- Age interval of the patients is from 42 to 61 (mean = 53.9, SD = 6.4).
- Age range of healthy controls is from 47 to 73 (mean = 63.5, SD = 8.6).

Metadata about dataset

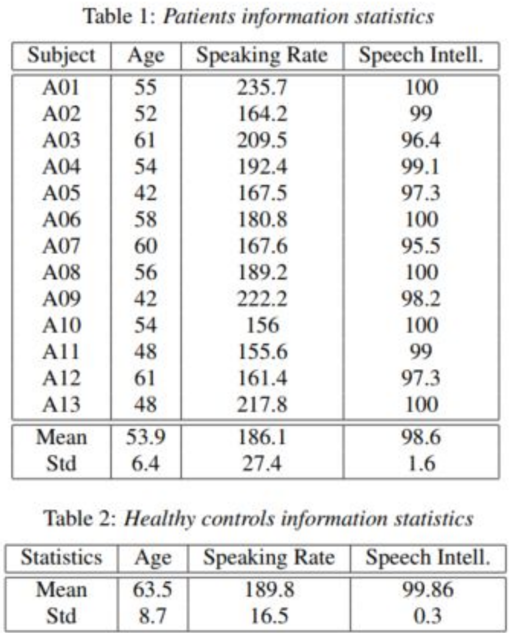


Table 1: *Patients information statistics*

Subject	Age	Speaking Rate	Speech Intell.
A01	55	235.7	100
A02	52	164.2	99
A03	61	209.5	96.4
A04	54	192.4	99.1
A05	42	167.5	97.3
A06	58	180.8	100
A07	60	167.6	95.5
A08	56	189.2	100
A09	42	222.2	98.2
A10	54	156	100
A11	48	155.6	99
A12	61	161.4	97.3
A13	48	217.8	100
Mean	53.9	186.1	98.6
Std	6.4	27.4	1.6

Table 2: *Healthy controls information statistics*

Statistics	Age	Speaking Rate	Speech Intell.
Mean	63.5	189.8	99.86
Std	8.7	16.5	0.3

Method

Performed ALS classification from speech samples using

The following three classification methods:

1) ANN with statistical (hand-crafted) features as in our previous work [6],

2) Time-domain CNN

3) Frequency-domain CNN based representation learning approaches

3.1. Baseline Approach: ANN with Statistical Features

openSMILE- used statistical features extracted by the publicly available tool

hand-crafted features openSMILE [29] was used to extract that are statistical variation of the widely used acoustic features

Mean and standard deviation of mel-frequency cepstral coefficients (MFCCs)

Quartile of the fundamental frequency contour from the speech samples

Total 7,755 acoustic features were extracted from each speech sample and fed into the ANN-based classification model

The ANN has the 2 dimensional softmax output layer: ALS and healthy

The test samples were classified by a maximum a posteriori probability obtained from ANN.

Although the network implementation was based on the ANN framework in TensorFlow [30], it had only one hidden layer.

3.2. CNN with Filterbank Energies

- Representation learning is a **feature learning method i**n which the model learns **useful feature representation from low-level signals** without hand-crafted feature extraction.
- Widely used hand-crafted features in speech recognition, de-correlation of filterbank energies in spectral domain.
- The **de-correlation may lead to loss of useful information to discriminate ALS** and healthy from speech signals

3.2.1. Time-CNN

3.2.1. Time-CNN

- Time-domain convolution applies convolution and pooling operations over time,
- Extract modulating characteristics while keeping invariance to a small shift in time [24].
- Three layers of CNNs were used with different filter sizes of 1 × 6, 1 × 5, and 1 × 3 for the corresponding layers, respectively.
- Each layer was sub-sampled by non-overlapping max pooling operation with 1 × 2, 1 × 3, and 1 × 3 with 64 feature maps, respectively

3.2.2. Frequency-CNN

- Frequency-domain convolution applies convolution and pooling operations along frequency,
- Represent useful spectral features while reducing frequency variance [24].
- Three layers of CNNs
- Filter sizes 7 × 1, 5 × 1, and 3 × 1
- Each layer was sub-sampled by non-overlapping max pooling operation
- 2 × 1, 4 × 1, and 4 × 1 with 64 feature maps, respectively

Precision measures

Accuracy, sensitivity, and specificity were the major performance indicators in this experiment.

Accuracy is the overall probability of correctly classified samples over the total number of samples.

Sensitivity is the probability of correctly predicted acoustic samples as a patient given all patient samples. Specificity is the probability of correctly classified healthy controls samples given all healthy control samples.

Person-level ALS detection performance

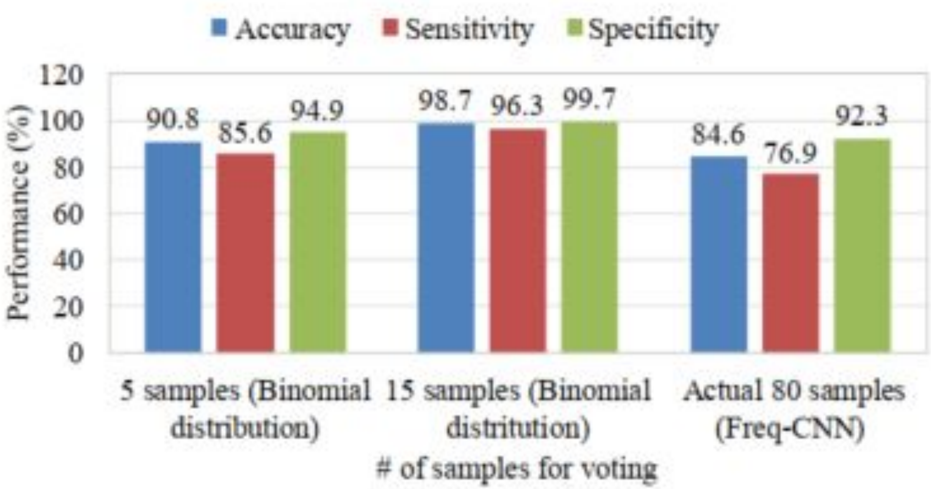
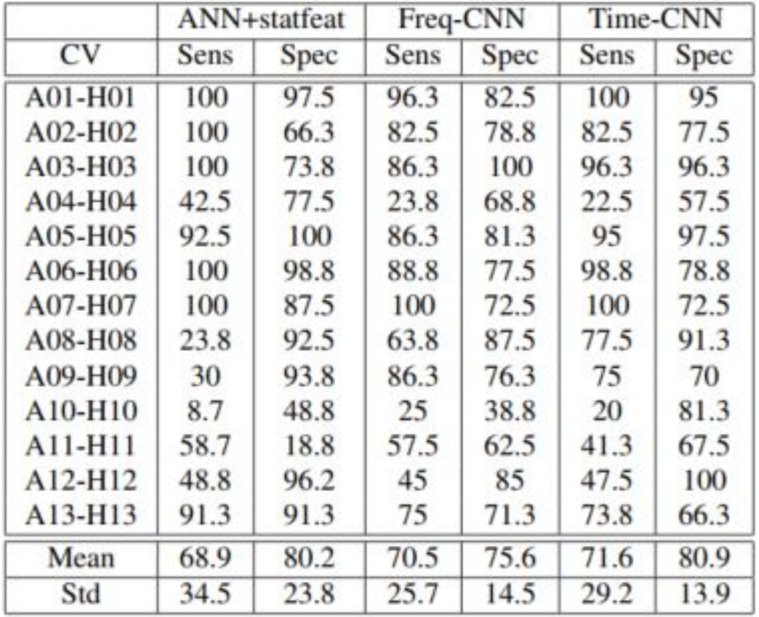


Figure 3: *Person-level ALS detection performance.*

Individual subject’s performance in cross validations (CV)



	ANN+statfeat		Freq-CNN		Time-CNN	
CV	Sens	Spec	Sens	Spec	Sens	Spec
A01-H01	100	97.5	96.3	82.5	100	95
A02-H02	100	66.3	82.5	78.8	82.5	77.5
A03-H03	100	73.8	86.3	100	96.3	96.3
A04-H04	42.5	77.5	23.8	68.8	22.5	57.5
A05-H05	92.5	100	86.3	81.3	95	97.5
A06-H06	100	98.8	88.8	77.5	98.8	78.8
A07-H07	100	87.5	100	72.5	100	72.5
A08-H08	23.8	92.5	63.8	87.5	77.5	91.3
A09-H09	30	93.8	86.3	76.3	75	70
A10-H10	8.7	48.8	25	38.8	20	81.3
A11-H11	58.7	18.8	57.5	62.5	41.3	67.5
A12-H12	48.8	96.2	45	85	47.5	100
A13-H13	91.3	91.3	75	71.3	73.8	66.3
Mean	68.9	80.2	70.5	75.6	71.6	80.9
Std	34.5	23.8	25.7	14.5	29.2	13.9

Questions

What is the main objective of the research paper?

What are the current problems in today’s diagnosis methods of ALS?

Which architectures are giving the best result while detecting ALS?

Explain the dataset- 2. why are we considering gender and age while selecting the samples from healthy control?