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

Suhas BN[1], Jhansi Mallela[2], Aravind Illa[2], Yamini BK[3], Nalini Atchayaram[4], Ravi Yadav[4],
Dipanjan Gope[5], Prasanta Kumar Ghosh[2]

[1]EE Dept., Penn State University, University Park, USA 16802
[2,5]EE & ECE Dept., Indian Institute of Science, Bangalore, India 560012
[3,4]Dept. of SPA & Neurology, NIMHANS, Bangalore, India 560029

***Abstract*—We consider the task of speech based classification of patients with amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD) and healthy controls (HC). Recent work in convolutional neural networks (CNN) to solve image classification problems raises the possibility of utilizing spectral representation of speech for detection of neurological diseases. In this paper, a spectrogram based approach is used. Feeding overlapping windows to the CNN makes sure that the temporal aspects are considered by using short signal segments or wide analysis filters. A three class (ALS, PD or HC) dysarthria classification is performed. In addition, we perform two severity classification experiments for ALS (5 class) and PD (3 class) respectively. Experiments are conducted on both baseline MFCC data [1] and log Mel spectrograms. Classification results show that for several audio lengths, models trained on log Mel spectrograms consistently outperform those of MFCC's. The ability of the network to accurately classify different classes is evaluated via the area under receiver operating characteristic curve [2], [3]. The findings from this study could aid in better detection and monitoring of ALS and PD diseases.**

***Index Terms*—spectrograms, CNN, dysarthria**

I. Introduction

Amyotrophic Lateral Sclerosis (ALS) and Parkinson's disease (PD) are progressive neurodegenerative diseases. Both the diseases currently have no cure although there exists some treatment for managing their symptoms [4]–[6]. ALS is a motor neuron disease, caused by gradual degeneration of motor neurons. Motor neurons provide communication links between the brain and voluntary muscles. In ALS, the upper and lower motor neurons degenerate and cease to send messages to the muscles. The muscles then weaken over time and begin to fasciculate and then atrophy is observed [7]. Over time, the brain loses its ability to make voluntary movements. On the other hand, Parkinson's disease (PD) is a brain disorder leading to shaking, stiffness, and movement issues. PD occurs when neurons in the area of the brain that controls movement become impaired and/or die. The neurons produce a chemical known as dopamine. When neurons become impaired, they produce less dopamine, leading to movement issues [8]. Patients affected by PD may experience mild tremors and develop a Parkinsonian gait that includes a tendency to lean forward as if hurrying. Similar to ALS, there are currently no blood or laboratory tests to diagnose non-genetic cases of PD. Patients affected by ALS sometimes show symptoms similar to that of PD, such as tremors, slow movement and body rigidness. Diagnosis of ALS and PD is primarily based on medical history and examination. Improvement after initiating medication is another important hallmark of PD [9]. Individuals with ALS may develop problems such as dyspnea, dysphagia and dysarthria. Currently, no single test can confirm ALS [7]. However, symptoms of upper or lower motor neurons (UMN/LMN) may indicate ALS. ALS symptoms in the early stages may exhibit traits of other, more common disorders. Muscle and imaging tests such as electromyography and nerve conduction study can suggest that the individual has a form of peripheral neuropathy rather than ALS [10], [11]. Magnetic resonance imaging could reveal issues that may be causing the symptoms of ALS. The above three tests are more to rule out other conditions rather than to confirm the disease itself [11].

The mean diagnosis time for the detection and confirmation of ALS adds up to 14 months [7], [12]. Early diagnosis and assessment becomes crucial. Patients have an average survival of 1.75 to 4 years with a worldwide annual incidence of 1.9 per 100,000 [13], [14]. In India, ALS has a prevalence rate of 4/100,000. The male to female ratio of incidence is about 5:7 [14]. Revised El Escorial criteria is used for the diagnosis of ALS [12], while to monitor the progress of ALS, the ALS Functional Rating Scale-Revised (ALSFRS-R) is used [7]. Since ALS is difficult to diagnose, biomarkers could potentially help clinicians diagnose ALS earlier and faster. Additionally, biomarkers are needed to help predict and accurately measure disease progression and enhance clinical studies aimed at developing more effective treatments. There have been attempts to use Electromyography to assess neuromuscular disorder [15], and perform automatic classification [16]. There have been studies on rate of utterance of ALS and PD patients [17], [18] and are found to be lower than those of healthy controls (HCs). Gomez et al. [19] has used running speech segments to infer articulation kinematics to detect early symptoms and monitor the evolution of the condition. About 90% of all PD and 85% of all bulbar ALS patients [20] experience speech issues as a primary symptom. Symptoms such as strained, hoarse, breathy, slurry or a monotonic voice may be observed. Thus, audio features could embed such characteristic symp-

978-1-7281-8895-9/20/$31.00

Condition	Gender	Count	Age Range (Avg) in years
ALS	M	30	33 - 76 (58.60)
	F	30	38 - 75 (56.02)
PD	M	34	34 - 78 (58.22)
	F	26	36 - 74 (56.99)
HC	M	30	26 - 68 (44.21)
	F	30	31 - 65 (46.93)

TABLE I
SUBJECT COUNT AND AGE RANGE FOR EACH CONDITION - GENDER PAIR

toms and help in better classification. Although CNNs have been used in identifying ALS patients, KwangHoon et al. [21] used two 1-D convolution networks - (time and frequency respectively) for filter bank features (MFBE). As seen in [1], the spectro-temporal characteristics change depending on ALS/HC (and similarly for PD). The work presented here makes use of time-frequency 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 features are better captured through a 2D CNN (finer modelling of temporal-harmonic structures). Also, a comparison among the two features namely, SPEC, MFCC has been performed to find the best feature for classification purposes.

II. DATASET

For the experiments in this work, speech data is considered from 60 ALS, 60 PD and 60 HC subjects. Patients are recruited from the National Institute of Mental Health and Neurosciences (NIMHANS) in Bangalore, India. Data collection has been approved by the ethics committee of NIMHANS and consent forms have been signed by the subjects prior to the data collection. The age range (avg) vs gender for each condition has been summarized in Table I. All patients included in this study are confirmed to have ALS or PD by Neurologists at NIMHANS as per the El Escorial and UPDRS-III criteria [7], [22] respectively. It is to be noted that while an ALSFRS-R 0 refers to loss of speech for an ALS subject, UPDRS-III 0 refers to normal speech for a PD subject as summarized in Table II. The severity ratings of speech disturbance have been provided by five speech language pathologists (SLP) from NIMHANS. The inter-rater reliability has been calculated using the Fleiss' kappa (κ). For ALS subjects, κ = 0.9017 (Almost perfect agreement) while for PD subjects, κ = 0.6995 (Substantial agreement). The data was initially recorded at 44.1 kHz which was downsampled to 16 kHz for classification experiments. The recording setup consists of five devices (of varying performances [1]), namely, Apple iPhone 7 (IPH), Motorola G5 Plus (MOT), Xiaomi Redmi 4 (XIA), Zoom H6 X/Y recorder (ZOO) and Dell XPS 15 laptop (LAP). The four speech tasks (and their duration across all devices) namely SPON (Spontaneous speech/ monologue: 21 hrs), IMAG (describing images: 25.22 hrs), PHON (Sustained Phonation: 25.84 hrs) & DIDK (Diadochokinetic rate: 22.42 hrs) and their significance in understanding speech disorders have been discussed in detail in the baseline paper [1].

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

TABLE II
ALSFRS-R AND UPDRS-III SCALES USED FOR RATING THE SUBJECTS

III. ALS VS PD VS HEALTHY CLASSIFICATION

Three kinds of classification experiments are carried out: 1) ALS vs PD vs HC, 2) 5 class ALS severity classification, 3) 3 class PD severity classification. Audio files are split into windows of length N_w = {0.5, 0.8, 1, 1.2, 1.5, 2, 3} seconds with a shift of N_{sh} = 0.1 seconds. The files are then converted to log Mel spectrograms [23] (saved into numpy arrays [24] during implementation). Choi et al. [25] in their paper on audio tagging achieved similar classification accuracies using the Short Term Fourier Transforms and Mel spectrograms. It is observed that log-scaling of Mel spectrograms improves the accuracy over regular spectrograms. For this reason, log Mel spectrograms are used for all following experiments. The Melgram dimension is determined by number of Melbins, audio length and feature map. Overlapping windows are taken to make sure that the temporal dynamics are considered. The convolutional neural network (CNN) architecture used in this work is shown in Fig 1. CNNs can learn nested features, thus making it robust to translation and distortions.

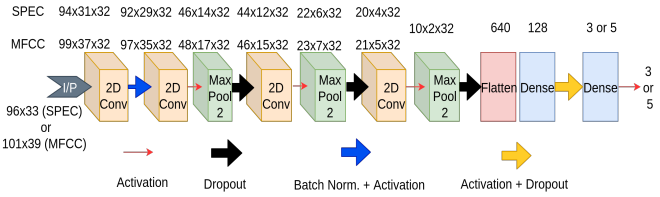

Fig. 1. Illustration of 2D CNN architecture proposed for classification.

The input feature SPEC has dimension of 96×33 with Melbins[1] = 96, and a audio length of 1 second represented by 33 frames. On the other hand, Mel frequency cepstral coefficients (MFCCs), used as features for comparison, has a dimension of 101×39 with audio length represented by 101 frames and feature dimension of 39 [1]. The number of convolutional filters or 'feature maps' is 32. The dimensions at the top of each block represent the output feature dimension after that layer. The activation function used is ReLU (except for softmax before output). A 3×3 convolutional kernel (represented by 2D Conv) is used. Every convolution layer of size $h \times w \times d$ learns ' d ' features of size $h \times w$, where h and w refer to the height and width of the kernels learnt. The

[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

Speech Task	Window lengths (sec)				
	0.5	**0.8**	**1**	**1.2**	**1.5**
SPON	0.814 (0.003)	0.832 (0.002)	**0.848 (0.009)**	0.844 (0.013)	0.815 (0.004)
DIDK	0.877 (0.005)	0.896 (0.003)	**0.918 (0.013)**	0.911 (0.014)	0.878 (0.005)
PHON	0.812 (0.015)	0.83 (0.01)	**0.846 (0.008)**	0.84 (0.013)	0.813 (0.01)
IMAG	0.808 (0.006)	0.826 (0.004)	**0.842 (0.01)**	0.84 (0.005)	0.809 (0.007)

TABLE III
ALS VS PD VS HC CLASSIFICATION ACCURACY (SD IN BRACKETS) ACROSS DEVICES USING CNN OVER EACH TASK & WINDOW LENGTH

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	**0.86 (0.01)**	0.85 (0.01)	0.85 (0.01)	0.84 (0.01)	0.84 (0.02)
	MFCC	0.67 (0.04)	0.68 (0.04)	0.67 (0.08)	**0.68 (0.03)**	0.64 (0.01)
DIDK	SPEC	**0.93 (0.01)**	0.90 (0.01)	0.92 (0.02)	0.90 (0.01)	0.89 (0.01)
	MFCC	**0.74 (0.04)**	0.72 (0.05)	0.73 (0.06)	0.73 (0.01)	0.75 (0.03)
PHON	SPEC	**0.89 (0.0)**	0.80 (0.01)	0.86 (0.01)	0.80 (0.01)	0.83 (0.01)
	MFCC	**0.72 (0.04)**	0.70 (0.05)	0.70 (0.05)	0.68 (0.02)	0.67 (0.07)
IMAG	SPEC	0.84 (0.01)	0.82 (0.01)	**0.86 (0.00)**	0.80 (0.01)	0.81 (0.01)
	MFCC	0.72 (0.01)	**0.77 (0.04)**	0.75 (0.03)	0.66 (0.03)	0.72 (0.01)

TABLE IV
COMPARISON OF ACCURACY (WITH SD IN BRACKETS) FOR T=1SEC ACROSS 5 FOLDS FOR DIFFERENT TASKS AND DEVICES BETWEEN MFCC BASELINE AND SPECTROGRAMS

size of pooling area for max pooling (represented by Max Pool 2) is 2×2. The optimum convolutional layer dropout is experimentally set to 0.5 while the dense layer dropout is set to 0.6. The output at the final dense layer is either 3 (for ALS vs PD vs HC and 3 class PD severity classification) or 5 (for 5 class ALS severity classification). The training is done using Categorical cross-entropy as the loss function with Adadelta optimizer working the best among the chosen ones. Keras library was used for the implementation [26].

IV. EXPERIMENTAL SETUP

For classification using MFCC and SPEC, a five-fold cross-validation setup is used. Five groups, each with thirty-six subjects with twelve from each of ALS, PD and HC are formed. Subjects in each group are chosen such that they are balanced in all aspects as mentioned earlier. For computing the MFCC baseline features, the methodology in [1] has been followed. For all the feature based classification experiments, in each fold, four groups are used for training and the remaining group is used as the test set in a round robin fashion with 0.25 split off from training for validation. Batch sizes of {64, 128, 256, 512} are used for hyper-parameter tuning. Early stopping criterion using validation loss has been imposed to prevent overfitting.

V. RESULTS

ALS vs PD vs HC classification with window lengths of N_w = {0.5, 0.8, 1, 1.2, 1.5, 2, 3}s with a shift of N_{sh} = 0.1s is performed for all devices and speech tasks. The average accuracy across devices for each speech task along with their standard deviations (SD) is tabulated in Table III. From this table, it is observed that N_w = 1s consistently performs the best among different window lengths with N_w = 1.2s providing the next best performance. However, considering the SD between N_w = 1 & 1.2s, it is seen that the former is more consistent across different tasks. N_w = 0.8s is consistent across all devices with the least SD for each speech task. It is found that batches of 256 and 512 perform equally and, a batch size of 256 is preferred for all experiments. Further, in the 3 class ALS vs PD vs HC classification, for N_w = 1s, the comparison of averaged 5 fold accuracies for each device and speech task pair has been tabulated for both SPEC and MFCC features in Table IV. It is observed that in the case of SPON, DIDK and PHON, the combination of MOT and SPEC perform the best. In the case of IMAG, the combination of IPH and SPEC performs the best. From the above, for the case of 3 class ALS vs PD vs HC classification, it can be concluded from Table IV that regardless of the speech task and device pair, SPEC consistently performs better than MFCC. Along with the classification accuracies for different experiments, the

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	0.761 (0.08)	0.726 (0.09)	**0.807 (0.09)**	0.741 (0.04)	0.740 (0.06)
	MFCC	0.701 (0.09)	0.616 (0.01)	0.717 (0.01)	0.613 (0.05)	0.656 (0.07)
DIDK	SPEC	0.741 (0.01)	0.771 (0.07)	**0.791 (0.013)**	0.784 (0.01)	0.763 (0.01)
	MFCC	0.641 (0.01)	0.671 (0.08)	0.700 (0.01)	0.652 (0.013)	0.654 (0.02)
PHON	SPEC	0.703 (0.07)	0.684 (0.01)	**0.725 (0.01)**	0.713 (0.08)	0.697 (0.08)
	MFCC	0.643 (0.08)	0.546 (0.01)	0.645 (0.013)	0.595 (0.09)	0.557 (0.09)
IMAG	SPEC	**0.771 (0.05)**	0.752 (0.06)	0.74 (0.07)	0.735 (0.08)	0.735 (0.01)
	MFCC	0.671 (0.06)	0.662 (0.08)	0.717 (0.08)	0.695 (0.09)	0.675 (0.02)

TABLE V
COMPARISON OF ACCURACY (WITH SD IN BRACKETS) FOR T=1SEC ACROSS 5 FOLDS FOR DIFFERENT TASKS AND DEVICES BETWEEN SPECTROGRAMS AND MFCC FOR 5 CLASS ALS SEVERITY CLASSIFICATION

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	0.75 (0.01)	0.76 (0.01)	0.76 (0.01)	0.76 (0.01)	**0.77 (0.06)**
	MFCC	0.65 (0.04)	0.68 (0.01)	0.69 (0.03)	0.63 (0.01)	0.61 (0.01)
DIDK	SPEC	**0.87 (0.01)**	0.87 (0.01)	0.85 (0.01)	0.85 (0.00)	0.86 (0.01)
	MFCC	0.80 (0.01)	0.82 (0.01)	0.84 (0.02)	0.80 (0.01)	0.79 (0.01)
PHON	SPEC	**0.85 (0.01)**	0.81 (0.00)	0.82 (0.01)	0.79 (0.01)	0.77 (0.01)
	MFCC	0.68 (0.04)	0.75 (0.05)	0.62 (0.08)	0.66 (0.09)	0.70 (0.01)
IMAG	SPEC	**0.81 (0.00)**	0.79 (0.01)	0.81 (0.01)	0.76 (0.01)	0.77 (0.01)
	MFCC	0.70 (0.05)	0.74 (0.04)	0.75 (0.02)	0.70 (0.01)	0.67 (0.03)

TABLE VI
COMPARISON OF ACCURACY (WITH SD IN BRACKETS) FOR T=1SEC ACROSS 5 FOLDS FOR DIFFERENT TASKS AND DEVICES BETWEEN SPECTROGRAMS AND MFCC FOR 3 CLASS PD SEVERITY CLASSIFICATION

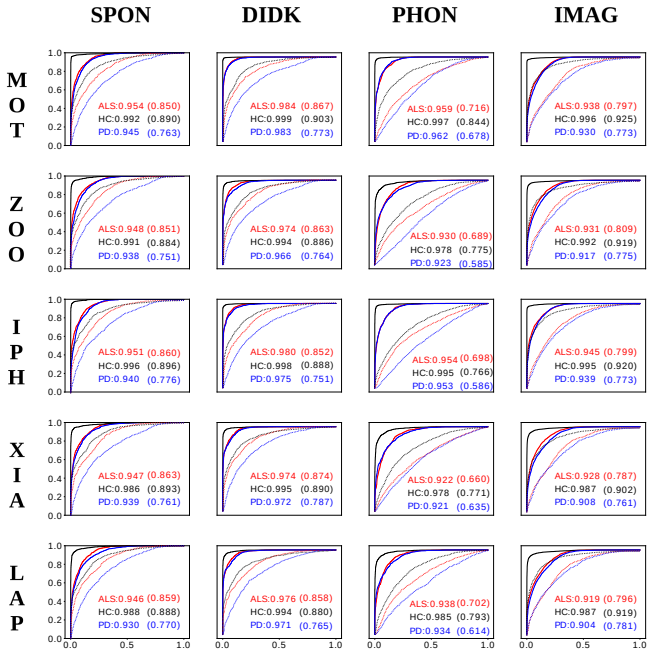

Fig. 2. AUC-ROC curves for 3 class ALS vs PD vs HC classification from different tasks (columns) and devices (rows) with True Positive Rate (TPR) vs False Positive Rate (FPR) on the Y & X axis respectively. The bold lines correspond to the SPEC feature while the dashed lines correspond to the MFCC. The AUC scores for MFCC have been presented in brackets next to the SPEC values for reference

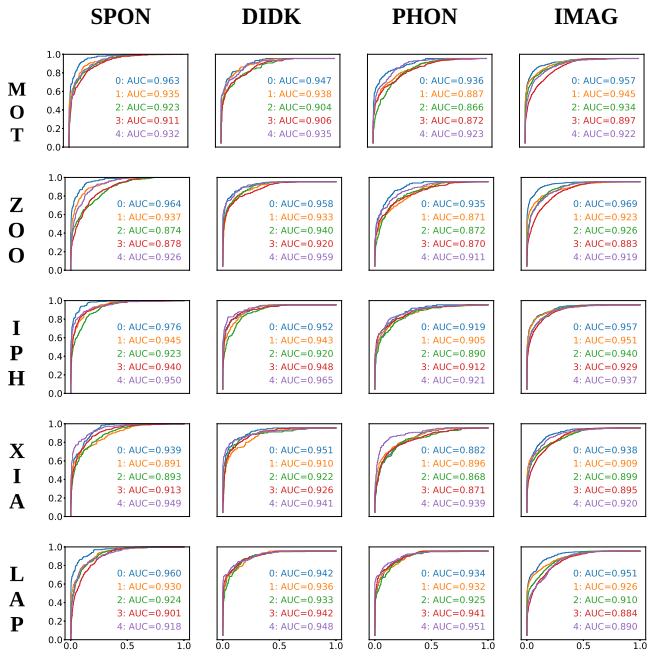

Fig. 3. AUC-ROC curves for 5 class ALS severity classification from different tasks (columns) and devices (rows) with TPR vs FPR on the Y & X axis respectively using SPEC features

area under the receiver operating characteristic curve (AUC-ROC) is plotted. AUC-ROC is a performance measure for the classification problem at various threshold settings. ROC is a probability curve while the AUC represents the measure of separability [3]. In the 3 class ALS vs PD vs HC classification,

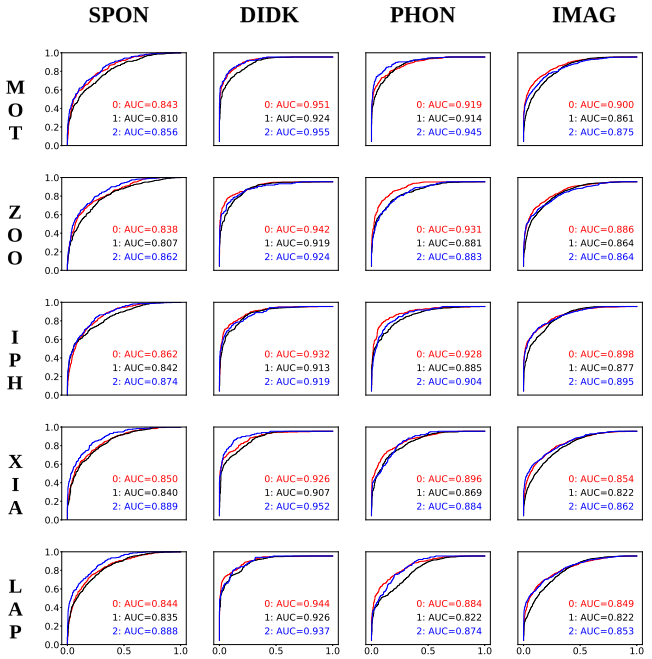

Fig. 4. AUC-ROC curves for 3 class PD severity classification from different tasks (columns) and devices (rows) with TPR vs FPR on the Y & X axis respectively using SPEC features

3 AUC-ROC curves can be plotted using the one vs rest methodology. A model with an AUC close to 1(0) reflects a good(poor) measure of separability between the classes. Thus, higher the AUC, better is the model at distinguishing between ALS, PD and HC. The AUC-ROC curves for SPEC (bold lines) and MFCC (dotted lines with AUC scores in brackets) are plotted for each device-task pair in Fig. 2. It is seen that the AUC scores for HC are above 0.98 which represents good performance of the model with unseen test case in distinguishing HC and ALS or PD. Apart from HC, the worst to best performance of different speech tasks varies from 0.90 for IMAG to 0.99 for DIDK across different devices. The lowest AUC for the three classes is with the combination of LAP and IMAG (the best AUC for MOT and DIDK combination). The best average AUC score across devices for different speech tasks is seen in MOT while the best average AUC score across speech tasks is seen in DIDK. This observation matches with the tabular results seen in Table IV. SPEC performs better than MFCC as observed from Table IV and Fig. 2. For this reason, we proceed to plot the AUC curves for SPEC for all following experiments. To the best of our knowledge, speech based severity classification of a neurological disease such as ALS or PD has not been attempted earlier in literature. Considering the case of 5 class ALS severity experiment, the averaged 5 fold accuracies for each device and speech task pair have been tabulated for the two features in Table V. The combination of IPH and SPEC consistently perform the best for 3 speech tasks, namely SPON, DIDK and PHON while MOT and SPEC combination performs the best for IMAG speech task. The AUC-ROC curves for 5 class ALS severity classification are

plotted for each device-task pair for SPEC in Fig. 3. Using one vs rest strategy, the classifier is able to accurately predict the severity of the condition with a max(min) AUC score of 0.976(0.882) for Severity 0, 0.951(0.871) for Severity 1, 0.940(0.866) for Severity 2, 0.948(0.870) for Severity 3 and 0.965(0.890) for Severity 4 among all device-task combinations. It is observed that across all severity-speech task-device combinations, the max(min) AUC score is 0.976(0.866) which shows good separability between severity classes. The averaged 5 fold accuracies for 3 class PD severity classification using each device - speech task pair have been tabulated for both SPEC and MFCC in Table VI. Also, The best feature and device combination turns out to be SPEC and MOT (except for SPON - SPEC and LAP). This is consistent with the previous 3 class ALS vs PD vs HC and 5 class ALS severity classification results. The corresponding AUC-ROC curves is plotted for each device-task pair for SPEC in Fig. 4. We observe max(min) AUC score of 0.951(0.838) for Severity 0, 0.926(0.807) for Severity 1 and 0.955(0.853) for Severity 2. The results are slightly lesser or comparable to ALS severity AUC scores and show good separability between the classes.

The SPEC features of dimension (96×33) consistently out performs MFCC features of dimension (101×39) across all tasks and devices. One explaination for this could be that the fully connected networks (consisting of conv. layers and subsampling) improve performance of CNNs as it reduces total parameters by sharing weights [25]. The features learnt are invariant to the location on the TF plane of SPEC. This allows the network to model temporal and harmonic structures of audio signals and could have led to improved performance.

VI. Conclusions

In this work, we have considered three classification problems (3 class ALS vs PD vs HC, 5 class ALS Severity and 3 class PD Severity) using four different speech tasks (SPON, DIDK, PHON and IMAG) with a 2-dimensional Convolutional Neural Network. In the case of ALS vs PD vs HC classification, the experiments showed that the classification with log Mel spectrograms performs better than the baseline scheme of MFCCs in all the speech tasks. It is observed that, regardless of the recording device used, similar accuracies were obtained. Further, severity classification (of any kind) of a neurological disease such as ALS or PD having not been attempted earlier shows good promise in identifying the condition and severity at an early stage. The code for this work can be obtained here

References

[1] B. N. Suhas., D. Patel, N. Rao, Y. Belur, P. Reddy, N. Atchayaram, R. Yadav, D. Gope, and P. K. Ghosh, "Comparison of Speech Tasks and Recording Devices for Voice Based Automatic Classification of Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis," in *Proc. Interspeech 2019*, 2019, pp. 4564–4568.

[2] S. N. Awan, N. Roy, D. Zhang, and S. M. Cohen, "Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores," *Journal of Voice*, vol. 30, no. 2, pp. 130–144, 2016.

[3] S. Narkhede, "Understanding AUC - ROC Curve," May 2019. [Online]. Available: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[4] V. Neuro, "ALS vs Parkinson's - How Do These Conditions Differ?" Aug 2019. [Online]. Available: https://alstreatment.com/als-vs-parkinsons/

[5] A. N. Lieberman, "Update on Parkinson disease." *New York State Journal of Medicine*, 1987.

[6] J.-P. Julien, "ALS: Astrocytes move in as deadly neighbors," *Nature Neuroscience*, vol. 10, no. 5, pp. 535–537, 2007.

[7] J. M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, A. Nakanishi, B. A. S. Group, A. complete listing of the BDNF Study Group *et al.*, "The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function," *Journal of the neurological sciences*, vol. 169, no. 1-2, pp. 13–21, 1999.

[8] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases." *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 55, no. 3, pp. 181–184, 1992.

[9] N. I. of Neurological Disorders and Stroke, "Parkinson's disease," 2017. [Online]. Available: https://www.nia.nih.gov/health/parkinsons-disease

[10] N. C. Joyce and G. T. Carter, "Electrodiagnosis in persons with amyotrophic lateral sclerosis," *PM&R*, vol. 5, no. 5, pp. S89–S95, 2013.

[11] N. I. of Neurological Disorders, O. o. C. Stroke, and P. Liaison, "Amyotrophic lateral sclerosis (ALS) fact sheet," 2018. [Online]. Available: https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Amyotrophic-Lateral-Sclerosis-ALS-Fact-Sheet

[12] B. R. Brooks, R. G. Miller, M. Swash, and T. L. Munsat, "El escorial revisited: revised criteria for the diagnosis of Amyotrophic Lateral Sclerosis," *Amyotrophic Lateral Sclerosis and other motor neuron disorders*, vol. 1, no. 5, pp. 293–299, 2000.

[13] K. C. Arthur, A. Calvo, T. R. Price, J. T. Geiger, A. Chio, and B. J. Traynor, "Projected increase in Amyotrophic Lateral Sclerosis from 2015 to 2040," *Nature communications*, vol. 7, p. 12408, 2016.

[14] A. Nalini, K. Thennarasu, M. Gourie-Devi, S. Shenoy, and K. Dinakar, "Clinical characteristics and survival pattern of 1153 patients with Amyotrophic Lateral Sclerosis: experience over 30 years from India," *Journal of the Neurological Sciences*, vol. 272, no. 1-2, pp. 60–70, 2008.

[15] F. Buchthal, C. Guld, and P. Rosenfalck, "Action potential parameters in normal human muscle and their dependence on physical variables." *Acta physiologica scandinavica*, vol. 32, no. 2-3, pp. 200–218, 1954.

[16] C. S. Pattichis and A. G. Elia, "Autoregressive and cepstral analyses of motor unit action potentials," *Medical Engineering & Physics*, vol. 21, no. 6-7, pp. 405–419, 1999.

[17] H. Hirose, S. Kiritani, and M. Sawashima, "Patterns of dysarthric movement in patients with Amyotrophic Lateral Sclerosis and pseudobulbar palsy," *Folia Phoniatrica et Logopaedica*, vol. 34, pp. 106–112, 1982.

[18] J. R. Green, C. A. Moore, and K. J. Reilly, "The sequential development of jaw and lip control for speech," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 1, pp. 66–79, 2002.

[19] P. Gomez, D. Palacios, A. Gomez, V. Rodellar, and A. Londral, "Articulation acoustic kinematics in ALS speech," in *International Conference and Workshop on Bioinspired Intelligence*. IEEE, 2017, pp. 1–6.

[20] S. Shellikeri, J. R. Green, M. Kulkarni, P. Rong, R. Martino, L. Zinman, and Y. Yunusova, "Speech movement measures as markers of bulbar disease in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 5, pp. 887–899, 2016.

[21] K. An, M. J. Kim, K. Teplansky, J. R. Green, T. Campbell, Y. Yunusova, D. Heitzman, and J. Wang, "Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks." in *Interspeech*, 2018, pp. 1913–1917.

[22] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[23] B. McFee, V. Lostanlen, and M. M. and, "librosa/librosa: 0.7.0," Jul. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3270922

[24] S. H. Hawley, "Panotti: A Convolutional Neural Network classifier for multichannel audio waveforms," Mar. 2017. [Online]. Available: https://doi.org/10.5281/zenodo.1275605

[25] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "A comparison of audio signal preprocessing methods for deep neural networks on music tagging," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1870–1874.

[26] F. Chollet *et al.*, "Keras," https://keras.io, 2015.