Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's Disease and Healthy Controls with CNN-LSTM using transfer learning

Jhansi Mallela[⋆], *Aravind Illa*[⋆], *Suhas B N*[⋆], *Sathvik Udupa*[⋆], *Yamini Belur*[†],
Nalini Atchayaram[††], Ravi Yadav[††], Pradeep Reddy[††], Dipanjan Gope[⋆⋆], Prasanta Kumar Ghosh[⋆]

[⋆]EE Department, [⋆⋆]ECE Department, Indian Institute of Science, Bengaluru 560012, India
[†]Department of Speech Pathology and Audiology , [††]Department of Neurology,
National Institute of Mental Health and Neurosciences, Bengaluru 560029, India

ABSTRACT

In this paper, we consider 2-class and 3-class classification problems for classifying patients with Amyotrophic Lateral Sclerosis (ALS), Parkinson's Disease (PD), and Healthy Controls (HC) using a CNN-LSTM network. Classification performance is examined for three different tasks, namely, Spontaneous speech (SPON), Diadochokinetic rate (DIDK) and Sustained phoneme production (PHON). Experiments are conducted using speech data recorded from 60 ALS, 60 PD, and 60 HC subjects. Classifications using SVM and DNN are considered as baseline schemes. Classification accuracy of ALS and HC (indicated by ALS/HC) using CNN-LSTM has shown an improvement of 10.40%, 4.22% and 0.08% for PHON, SPON and DIDK tasks, respectively over the best of the baseline schemes. Furthermore, the CNN-LSTM network achieves the highest PD/HC classification accuracy of 88.5% for the SPON task and the highest 3-class (ALS/PD/HC) classification accuracy of 85.24% for the DIDK task. Experiments using transfer learning at low resource training data show that data from ALS benefits PD/HC classification and vice-versa. Experiments with fine-tuning weights of 3-class (ALS/PD/HC) classifier for 2-class classification (PD/HC or ALS/HC) gives an absolute improvement of 2% classification accuracy in SPON task when compared with randomly initialized 2-class classifier.

***Index Terms*—** Amyotrophic Lateral Sclerosis, Parkinson's Disease, CNN-LSTM

1. INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) and Parkinson's Disease (PD) are some of the most prevalent neuro-degenerative movement disorders. Early detection can help to prolong survival and quality of life. ALS, a progressive motor neuron disease, affects nerve cells in the brain and spinal cord, causing loss of muscle control. Life expectancy of a person with ALS averages from 2 to 5 years from the time of diagnosis. About 5-10% of all ALS patients live to survive beyond 10 years [1]. The annual worldwide incidence is about 1.9/100,000 [2] while in India, the prevalence of ALS is 4/100,000 with an annual incidence of 1/100,000 and a male to female ratio of 5:7 [2]. Diagnosis of ALS is based on the revised El Escorial criteria [3], with a median diagnosis time of 14 months [4]. ALS Functional Rating Scale-Revised (ALSFRS-R) is used for monitoring the progress of the disease [5]. Symptoms of ALS include progressive muscle atrophy and weakness, leading to problems that include dysphagia, dyspnoea, orthopnea and dysarthria [1]. About 30% of all ALS patients experience dysarthria as the first symptom [6]. Dysarthria is a frequently occurring symptom as the disease progresses [7]. There is currently no single definitive test that can be used to accurately diagnose ALS [5]. Despite all the benefits of the ALSFRS-R, it remains a subjective score where the judgements may be inconsistent [8].

PD, on the other hand, is a brain disorder that occurs when the dopamine generating neurons (neurotransmitters) in the brain are damaged or die. Subjects begin to experience difficulty in speaking, writing, walking, or completing other simple tasks. Patients suffering from PD have an average survival expectancy of 5 to 10 years. For individuals between the ages of 30 and 39, the incidence of the disease is about 41/100,000 while it goes beyond 1,900/100,000 among those who are 80 and older. It is observed that men are 1.5 times more likely to have PD when compared to women. About 40% to 95% of PD patients experience dysphagia. Different structures of the articulatory subsystem (e.g., the lips, tongue, and jaw) are affected at different times during the progression of the disease. Presumably, this non-uniform rate of deterioration leads to compensatory interactions between the articulators (e.g., tongue and jaw). Early movement studies revealed evidence supporting this notion and showed a decrease in the size of tongue movements but exaggerated jaw movements during speech tasks [9]. Similar to ALS, there are currently no blood or laboratory tests to diagnose non-genetic cases of PD. Diagnosis is based on medical history and a neurological examination.

Early detection of ALS and PD through automated methods can avoid clinicians' subjectivity in the diagnosis. It could also reduce diagnosis time. The speech impairment due to ALS is caused by muscle disorder while in PD, it is due to reduced levels of dopamine which, in turn, affect speech articulators. There have been attempts to use Electromyography (EMG) to assess neuromuscular disorder [10] and perform automatic classification using features extracted from EMG signal [11]. The rate of articulatory movement of ALS patients have been studied [12–14] and are found to be lower than those with healthy control (HC). Gomez et al. [15] have used running speech segments to infer articulation kinematics to detect early symptoms and monitor the evolution of the disease. Yamini et al. [16] observed a reduction in the vowel space area in case of bulbar ALS patients compared to that of HCs. Using syllable rate and maximum phonation duration, Taylor et al. [17] attempted automatic classification of ALS patients based on fractal analysis and using diadochokinetic (DIDK) rates as speech tasks. Suhas B N et al. [18] studied the performance of three different speech tasks namely, Sponta-

Authors thank the Department of Science and Technology, Govt. of India for their support in this work.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on December 31,2020 at 09:55:51 UTC from IEEE Xplore. Restrictions apply.

neous speech (SPON), Diadochokinetic rate (DIDK), and Sustained phoneme production (PHON) in automatic classification of ALS/HC using SVM and DNN.

Data collection from ALS and PD patients is often tedious making a large corpus development a challenging task. In order to handle such a low resource condition in the classification task, in this work Transfer Learning (TL) approach is explored. Even though the speech characteristics in PD and ALS could be different from speech of a HC, there could be similarities in the way speech disorder manifests in these two diseases. Such similarities could be exploited in the classification of ALS/HC and PD/HC separately when a large amount of data is not available. For this purpose, we follow a TL approach for classification tasks of ALS/HC and PD/HC. To the best of our knowledge this is the first work that utilizes the ALS (PD) data for classifying PD (ALS) using transfer learning approach. Further in this work, we propose a CNN-LSTM based classifier for 2-class classification between ALS and HC (ALS/HC), PD and HC (PD/HC) and a 3-class classification between ALS, PD, and HC (ALS/PD/HC). Experiments with CNN-LSTM, in comparison to the baseline SVM and DNN classifiers, show an improvement in the classification accuracy. Under low resource training data, initializing the CNN-LSTM classifier from other classification tasks shows an improvement compared to randomly initialized weights, which could imply that the data from PD could help in ALS/HC classification and vice-versa. Also, experiments with the 3-class classifier (ALS/PD/HC) reveal that further fine-tuning of the weights with only ALS (or PD) data improves the 2-class classification of ALS/HC (or PD/HC).The rest of the paper is organised as follows. We will begin with the description of the data collection process in Section 2, followed by the proposed approach in Section 3. In Section 4, we present the experiments and results.

2. DATA COLLECTION

In this work, the speech data was collected from 60 ALS, 60 PD and 60 HC subjects at the National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India. Ethics committee of NIMHANS approved for data collection and consent forms have been signed by the patients prior to data collection. The data used in this study was collected from patients, diagnosed to have ALS (El Escorial criteria) or PD by Neurologists at NIMHANS. Details of the severity, age, and gender of subjects used in this study are provided in Table 1 and 2. Severity ratings for patients as per ALSFRS-R (5-point scale, 0:Loss of useful speech to 4:Normal), UPDRS-III (5-point scale, 0:Normal to 4:Unintelligible speech) severity scales have been confirmed and rated by the five speech language pathologists (SLP) from the Speech Pathology and Audiology Department, NIMHANS. Although the UPDRS-III severity ratings are based on a 5-point scale (0 to 4), we consider scores from (0 to 2) due to lack of PD data for severities over UPDRS-III (2).

The subjects in this work come from six different native languages, namely, Bengali, Hindi, Odiya, Tamil, Telugu, Kannada in an approximately equal proportion.

The speech data used in this study was recorded by using Zoom H-6 recorder with XYH-6 X/Y capsule high quality unidirectional microphone [19] from a distance of 2 feet from the subject at 44.1 kHz. The data was then downsampled to 16 kHz for classification experiments. We use three tasks in this work: 1) Spontaneous speech task, 2) Diadochokinetic task and 3) Sustained Phoneme task.

Spontaneous speech task (SPON): Two parts are considered, where in the first part, the subject was asked to spontaneously talk about *"a festival they celebrate"* and in the second part, they had to describe *"a place that they have recently visited"* for one minute each in their native language. A few minutes of preparation time was given to the subjects before they started. The total duration of recordings from all subjects for this task is **5.62 hours**. Even, SPON is an informal assessment measure, as it has a good representation of the natural speech of a subject, thus making it a useful task for assessing a subject's articulation [20]. It is a task where we can evaluate integrated functioning of all components in speech production (respiration, phonation, articulation, resonance, and prosody) [20].

ALS Subjects	**ALSFRS - R Severity Score**				
	0	**1**	**2**	**3**	**4**
Male	5	8	5	8	4
Female	5	8	5	8	4
Mean age	57.30	55.36	56.10	53.90	55.11
Std. Dev	7.18	10.86	11.16	9.74	7.22

Table 1: ALSFRS-R Severity, gender and age details of ALS patients

Diadochokinetic rate task (DIDK): In DIDK, there are two parts: (i) Alternating Motion Rates (AMRs), (ii) Sequential Motion Rates (SMRs). For AMRs, subjects were asked to repeat, without interruption, the sequence of monosyllabic targets *"pa-pa-pa", "ta-ta-ta", "ka-ka-ka"* as fast as possible and without losing articulatory precision. It determines the speed and regularity of reciprocal jaw, lip, and tongue movements and also represents the articulatory accuracy and the respiratory and phonatory support. For SMRs, subjects were asked to repeat *"pataka"* and *"badaga"* for a duration of up to 5 seconds that measure the ability of articulators to move quickly and in a proper sequence from one articulatory position to another [21]. The total duration of recordings from all subjects for this task is **4.65 hours** for all subjects.

Sustained phoneme production task (PHON): In this, there are two parts: (i) Vowel Prolongation (VP), (ii) Fricative Prolongation (FP). In VP part, subjects were asked to produce sustained phonemes corresponding to five vowels, namely, /a/, /i/, /o/, /u/, /æ/. In the FP part, subjects were asked to produce sustained phonemes corresponding to three fricatives, namely, /s/, /sh/, and /f/. The subjects were asked to do this at a comfortable pitch and loudness level, after taking a deep breath. For each vowel and fricative, the process is repeated three times. The total duration of recordings from all subjects for this task is **5.79 hours**. It is a task which isolates the respiratory-phonatory system for speech [21] and depends on the respiratory function that reflects information on respiratory abilities, voice quality and phonatory support. The fricative prolongation requires respiratory-articulatory competence which could be affected by ALS and PD.

Based on the speech recordings of ALS, PD and HC subjects, it is observed that in the DIDK task, the HC is able to utter multiple "pa" in a second while a patient with an ALSFR-S score of 0 (ALS) and patient with UPDRS-III score of 2 (PD) are able to repeat "pa" only one time and four times, respectively. In SPON task it is observed that pauses occur more frequently in HC whereas in ALS and PD patients there are no pauses over one second duration.

3. PROPOSED APPROACH

In this section, we first briefly review convolutional and recurrent neural networks and then describe the proposed CNN-LSTM and transfer learning approach.

Convolutional neural networks (CNN): Recently, convolutional neural networks have shown great success in the field of computer

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on December 31,2020 at 09:55:51 UTC from IEEE Xplore. Restrictions apply.

PD & HC Subjects	UPDRS - III Severity Score			Healthy Control
	0	1	2	5
Male	10	15	9	30
Female	6	12	8	30
Mean age	57.31	56.06	60.04	45.57
Std. Dev	10.61	10.08	8.59	8.68

Table 2: UPDRS-III Severity, gender and age details of PD patients and Healthy Controls

vision [22] and gained attention in speech processing to extract local features. We deploy the convolutional network to perform temporal convolution to extract features from the acoustic features. Let us consider, a CNN layer (m) with number of convolution filters N_m with filter length l_m and denote the collection of filters by $\mathbf{K}^m$=$\{\mathbf{K}_j^m\}_{j=1}^{N_m}$, for $\mathbf{K}_j^m \in \mathbb{R}^{N_{m-1} \times l_m}$ with a bias vector $\mathbf{b}^m \in \mathbb{R}^{N_m}$. Then given an input signal (from the output of previous convolution layer $(m-1)$) $\mathbf{S}^{m-1} \in \mathbb{R}^{N_{m-1} \times T_{m-1}}$, ($T_m$ is mth layer CNN output sequence length) we compute output $\mathbf{S}^m \in \mathbb{R}^{N_m \times (T_{m-1}-l_m+1)}$by

$$\mathbf{S}^m = \sigma(\mathbf{K}^m \star \mathbf{S}^{m-1} + \mathbf{b}^m) \tag{1}$$

where, $\star$ denotes 1-D convolution operator along time axis (t) and σ denotes non-linear activation function.

Long short-term memory networks (LSTM): For modelling sequential information, recurrent neural networks, especially long short-term memory (LSTM) networks, have been shown to perform well [23]. Let $\mathbf{x}_t$ be a D-dimensional input at time t, and C be the number of memory cells in an LSTM layer with output $\mathbf{y}_t \in \mathbb{R}^C$. There will be weight vectors for each LSTM layer of type: input weights $\mathbf{W}_* \in \mathbb{R}^{C \times D}$, recurrent weights $\mathbf{R}_* \in \mathbb{R}^{C \times C}$ and bias weights $\mathbf{b}_* \in \mathbb{R}^C$, (where $*$ indicates: input gate i, forget gate f, cell memory c, output gate o) . The output of each gate for LSTM layer is obtained as follows [23, 24]:

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{R}_i \mathbf{y}_{t-1} + \mathbf{b}_i) && \text{input gate} \\
\mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{R}_f \mathbf{y}_{t-1} + \mathbf{b}_f) && \text{forget gate} \\
\mathbf{c}_t &= \mathbf{c}_{t-1} \odot \mathbf{f}_t + \mathbf{i}_t \odot \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{R}_c \mathbf{y}_{t-1} + \mathbf{b}_c) && \\
& && \text{cell memory} \\
\mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{R}_o \mathbf{y}_{t-1} + \mathbf{b}_o) && \text{output gate}
\end{aligned} \tag{2}$$

where, σ is an element-wise non-linear activation function, $\odot$ denotes point-wise multiplication of two vectors. The block output of forward pass for LSTM layer is computed by $\mathbf{y}_t = \tanh \mathbf{c}_t \odot \mathbf{o}_t$.

CNN-LSTM for speech disorder classification: The characteristics of speech disorder are known to encode as para-linguistic information, which span over a long duration of time and are known as supra-segmental features. These are computed over a speech segment by computing statistics on the frame level acoustic features like Mel Frequency Cepstral Coefficients (MFCCs). These supra-segmental features are further utilized as input features for SVM/DNNs to perform classification tasks [18]. In this work, instead of computing the statistical features like mean, median, and standard deviation on frame level features, we derive supra-segmental features on a speech segment in a data driven manner. We propose a CNN-LSTM architecture for classifying ALS, PD and HC with different speech tasks. Fig. 1 illustrates the current approach to classification task. First we use 1D-CNN to extract the local temporal structure by performing temporal convolutions followed by a max-pooling layer. Upon the extracted local features

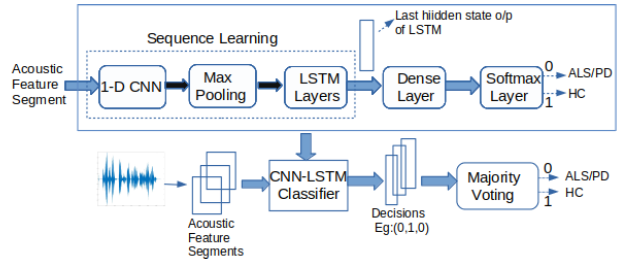

Fig. 1: Illustration of proposed classifier using CNN-LSTM.

computed by CNN, LSTM layers are used to capture the temporal dynamics of the sequence. The last hidden state output of LSTM layer is given to dense layer. Finally at the output of dense layer, we use softmax activation layer to classify HC with patient of ALS or PD. The classification tasks using CNN-LSTM are performed at segment level (each speech utterance is split into segments with overlapping frames). Decisions from all speech segments are combined to perform majority voting and obtain a decision on the speech utterance.

Transfer learning across ALS/HC, PD/HC and ALS/PD/HC: Transfer learning is a machine learning approach where weights of a neural network model trained for a particular task are utilized as the initialization of weights for a model with a different task. These approaches have been shown to benefit classifications, especially in the unavailability of large amounts of training data. In this work, we deploy a transfer learning approach for the classification of ALS/HC, PD/HC. We perform two different training techniques in transfer learning i) train a 2-class ALS/HC classifier and use it as an initialization for PD/HC classifier and vice-versa. ii) train a 3-class classifier (ALS/PD/HC) and fine-tune the weights to two separate classifiers, i.e., ALS/HC and PD/HC using respective data from ALS and PD separately.

4. EXPERIMENTAL SETUP AND RESULTS

For acoustic features, we compute 36-dim MFCCs at frame length 20ms and frame shift 10ms. Experiments are performed in a 5-fold cross validation setup comprising five groups, where each group consists of twelve subjects of ALS, HC and (or) PD. Subjects in each group are chosen such that they are balanced in all aspects such as age, gender and ALSFRS-R, UPDRS-III scores, as mentioned in Table 1 and 2. In each fold, three groups are used for training, one group for validation and one for testing in a round robin fashion. We trained the CNN-LSTM by chunking a speech signal into segments of length 2sec with an overlapping of 0.1sec. For this work, we have used 1 layer of CNN with 1D convolution across each dimension of the input with 'relu' activation function. For each dimension, 30 filters of length 20 with stride of 1 were chosen which was followed by a max-pooling of window size 4. The output of CNN is fed to LSTM, comprised of two hidden layers each with 64 units. To evaluate the classification performance, we use classification accuracy as a performance metric.

We presented the experiments and results with three main objectives – i) Comparing the performance of CNN-LSTM to ALS classification with those of baseline schemes based on SVM and DNN, ii) Transfer learning approach for 2-class classifications (ALS/HC; PD/HC) in low resource data condition, iii) 3-class classification (ALS/PD/HC) with CNN-LSTM.

Comparison of CNN-LSTM with SVM and DNN for ALS/HC classification task: We compare the results of the proposed CNN-LSTM approach with the baseline (SVM, DNN) reported in [18].

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on December 31,2020 at 09:55:51 UTC from IEEE Xplore. Restrictions apply.

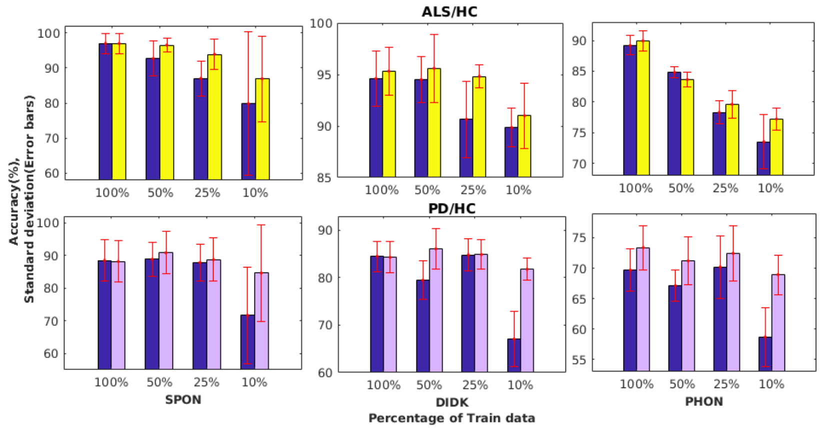

Fig. 2: Classification accuracy by varying percentage of training data.● RI, ● TL from PD/HC, ● TL from ALS/HC. Top (bottom) row corresponds to ALS/HC (PD/HC) classification task for SPON, DIDK, and PHON tasks in three columns.

ALS/HC	SPON	DIDK	PHON
SVM	89.99(3.2)	94.52(4.3)	78.52(5.1)
DNN	92.44(3.1)	93.43(3.2)	78.80(4.3)
CNN-LSTM	**96.96(2.8)**	**94.60(2.7)**	**89.20(1.5)**

Table 3: Average (SD) accuracy of ALS/HC classification

		SPON	DIDK	PHON
3-class	ALS/PD/HC	83.04(2.17)	85.24(4.25)	77.20(1.95)
2-class	ALS/HC	95.2(2.33)	87(2.36)	86.1(1.0)
	PD/HC	90.7(3.94)	87(2.36)	74.1(5.67)
	ALS/HC (FT)	**98.22(1.87)**	95.5(1.72)	89.28(1.10)
	PD/HC (FT)	**90.98(4.03)**	84.62(5.69)	73.52(4.61)

Table 4: Average accuracy (SD) of ALS/PD/HC model and pair-wise accuracy's of ALS/HC and PD/HC.

The suprasegmental features which are mean, median, and standard deviation (SD) of each MFCC are computed in a manner similar to that in [18]. The SVM classifier with the radial basis kernel has been trained using the *libsvm* package [25]. Optimal values of C and γ have been selected by maximizing the performance of the validation set. While for DNN, we use 2 hidden layers with 128 units [18]. Like CNN-LSTM, majority voting was also used for SVM and DNN based classification schemes to obtain utterance level decision. Table 3, reports the performance results of CNN-LSTM in comparison with SVM and DNN with different speech tasks in terms of average (and standard deviation (SD) in brackets) classification accuracy across five folds. In the SPON task, we observe an absolute improvement of 7% and 5% over SVM and DNN, respectively. While the performance of the DIDK task remains the same compared to SVM, there is a 2% improvement over DNN. In the PHON task, CNN-LSTM achieves 9-10% improvement compared to SVM and DNN. The improvements with the CNN-LSTM could be due to better modelling of temporal pattern from the acoustic features, by learning the supra-segmental features in a data driven manner, unlike fixed statistical features used for classification in SVM and DNN.

Transfer learning for ALS/HC and PD/HC with CNN-LSTM: The performance of the neural network based models depends on the amount of data used for training. In this experiment, we investigate to determine whether PD data will benefit the ALS/HC classification and vice-versa, especially under low amount of training data. For ALS/HC classification, we utilize the data available from PD to train a CNN-LSTM model to classify PD/HC; we further use this model as an initialization for training model to classify ALS/HC and vice-versa. In Fig. 2, bar height indicates the average (SD as error bar) classification accuracy across five folds, for ALS/HC and PD/HC classification, with random initialization (RI) and TL approach. We observe that at 100% of training data, the performance with TL in ALS/HC and PD/HC for all the speech tasks remains similar to those of RI. While the amount of training data is varied from 10, 25, and 50%, we observe that there is a reduction in accuracy's for both RI and TL based approaches. But interestingly, TL, on average, performed consistently better than the RI in most cases (23 out of 24).

Three class classification ALS/PD/HC using CNN-LSTM: We also perform experiments with 3-class classification to classify ALS/PD/HC. We train CNN-LSTM to perform a 3-class classification with a similar architecture and an additional node at the output. Table 4 reports the accuracy values with ALS/PD/HC in the first row and ALS/HC, and PD/HC accuracy values computed from the 3-class classifier in the second and third row. While comparing the two class accuracy with respect to Fig. 2, we observe a drop in the individual two class performances both in ALS/HC and PD/HC. So, we further Fine-Tune (FT) the weights of ALS/PD/HC model with ALS/HC data and PD/HC data separately. The accuracy values of these FT models have been reported in the last two rows in Table 4. When these are compared with 100% training data accuracy in Fig. 2, interestingly we observe an absolute improvement of $\sim$2% on SPON task using FT for both ALS/HC and PD/HC tasks, while the performance did not change much in the case of DIDK and PHON.

5. CONCLUSION

In this work, we have considered three classification problems (ALS/HC, PD/HC, ALS/PD/HC) using three different speech tasks (SPON, DIDK, PHON) with CNN-LSTM. In the case of ALS/HC, the experiments showed that the classification with CNN-LSTM consistently performs better than baselines (SVM and DNN) in all speech tasks. Experiments with low resource of training data showed that PD and ALS data are beneficial for cross classification tasks and showed improvements compared to random initialization. Further, fine-tuning the weights from 3-class to 2-class classification has also shown improvements in the SPON task of both ALS/HC and PD/HC. In future, we would like to examine the scientific rationale behind such benefits due to PD in classification for ALS and vice-versa. We also plan to investigate CNN-LSTM and transfer learning techniques on severity estimation of ALS and PD patients.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on December 31,2020 at 09:55:51 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Adriano Chiò, Giancarlo Logroscino, Orla Hardiman, Robert Swingler, Douglas Mitchell, Ettore Beghi, Bryan G Traynor, Eurals Consortium, et al., "Prognostic factors in ALS: a critical review," *Amyotrophic Lateral Sclerosis*, vol. 10, no. 5-6, pp. 310–323, 2009.

[2] A Nalini, K Thennarasu, M Gourie-Devi, Sandhya Shenoy, and Dinkar Kulshreshtha, "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.

[3] Benjamin Rix Brooks, Robert G Miller, Michael Swash, and Theodore 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.

[4] Matthew C Kiernan, Steve Vucic, Benjamin C Cheah, Martin R Turner, Andrew Eisen, Orla Hardiman, James R Burrell, and Margaret C Zoing, "Amyotrophic Lateral Sclerosis," *The Lancet*, vol. 377, no. 9769, pp. 942–955, 2011.

[5] Jesse M Cedarbaum, Nancy Stambler, Errol Malta, Cynthia Fuller, Dana Hilt, Barbara Thurmond, Arline Nakanishi, Bdnf ALS Study Group, 1A 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.

[6] Barbara Tomik and Roberto J Guiloff, "Dysarthria in Amyotrophic Lateral Sclerosis: a review," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 1-2, pp. 4–15, 2010.

[7] Raymond D Kent, Robert L Sufit, John C Rosenbek, Jane F Kent, Gary Weismer, Ruth E Martin, and Benjamin R Brooks, "Speech deterioration in Amyotrophic Lateral Sclerosis: a case study," *Journal of Speech, Language, and Hearing Research*, vol. 34, no. 6, pp. 1269–1275, 1991.

[8] Ray D Kent, "Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders," *American Journal of Speech-Language Pathology*, vol. 5, no. 3, pp. 7–23, 1996.

[9] H Hirose, S Kiritani, and IF Tatsumi, "On the nature of bradylalia: A preliminary study of the acoustic characteristics of dysarthric speech," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, vol. 16, pp. 229–234, 1982.

[10] Fritz Buchthal, Christian Guld, and Poul 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.

[11] Vipin K Mishra, Varun Bajaj, and Anil Kumar, "Classification of normal, ALS, and myopathy EMG signals using ELM classifier," in *2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*. IEEE, 2016, pp. 455–459.

[12] Hajime Hirose, Shigeru Kiritani, and Masayuki Sawashima, "Patterns of dysarthric movement in patients with Amyotrophic Lateral Sclerosis and pseudobulbar palsy," *Folia Phoniatrica et Logopaedica*, vol. 34, no. 2, pp. 106–112, 1982.

[13] Jordan R Green, Christopher A Moore, and Kevin 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.

[14] Aravind Illa, Deep Patel, BK Yamini, N Shivashankar, Preethish-Kumar Veeramani, Kiran Polavarapui, Saraswati Nashi, Atchayaram Nalini, Prasanta Kumar Ghosh, et al., "Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6014–6018.

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

[16] B. Yamini, N. Shivashankar, and A. Nalini, "Vowel space area in patients with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 9, no. 1, pp. 118–119, 2008.

[17] Antje S Mefferd, Gary L Pattee, and Jordan R Green, "Speaking rate effects on articulatory pattern consistency in talkers with mild als," *Clinical linguistics & phonetics*, vol. 28, no. 11, pp. 799–811, 2014.

[18] Suhas B.N., Deep Patel, Nithin Rao, Yamini Belur, Pradeep Reddy, Nalini Atchayaram, Ravi Yadav, Dipanjan Gope, and Prasanta Kumar 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.

[19] "Zoom high-quality unidirectional microphone, available online: https://www.zoom-na.com/, last accessed:10/15/2019," .

[20] Cyndi Stein-Rubin and Renee Fabus, *A guide to clinical assessment and professional report writing in speech-language pathology*, Nelson Education, 2011.

[21] Joseph R Duffy and Keith A Josephs, "The diagnosis and understanding of apraxia of speech: Why including neurodegenerative etiologies may be important," *Journal of Speech, Language, and Hearing Research*, 2012.

[22] Stéphane Mallat, "Understanding deep convolutional networks," *Phil. Trans. R. Soc. A*, vol. 374, no. 2065, pp. 20150203, 2016.

[23] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber, "LSTM: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[24] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on December 31,2020 at 09:55:51 UTC from IEEE Xplore. Restrictions apply.