https://doi.org/10.2352/EI.2023.35.9.IPAS-298

DEEP LEARNING BASED SPEECH EMOTION RECOGNITION FOR PARKINSON PATIENT

Habib Khan[1] *Mohib Ullah*[2] *Fadi Al-Machot*[3] *Faouzi Alaya Cheikh*[2] *Muhammad Sajjad* [2]

[1] Islamia College University Peshawar, 25000 Peshawar, Pakistan
[2] Norwegian University of Science and Technology, 2815 Gjøvik, Norway
[3] Faculty of Science and Technology (REALTEK), Norwegian University of Life Sciences (NMBU),
1430 Ås, Norway

ABSTRACT

Speech emotions (SEs) are an essential component of human interactions and an efficient way of persuading human behavior. The recognition of emotions from the speech is an emergent but challenging area of digital signal processing (DSP). Healthcare professionals are always looking for the best ways to understand patient voices for better diagnosis and treatment. Speech emotion recognition (SER) from the human voice, particularly in a person with neurological disorders like Parkinson's disease (PD), can expedite the diagnostic process. Patients with PD are primarily passed through diagnosis via expensive tests and continuous monitoring that is time-consuming and costly. This research aims to develop a system that can accurately identify common SEs which are important for PD patients, such as anger, happiness, normal, and sadness. We proposed a novel lightweight deep model to predict common SEs. The adaptive wavelet thresholding method is employed for pre-processing the audio data. Furthermore, we generated spectrograms from the speech data instead of directly processing voice data to extract more discriminative features. The proposed method is trained on generated spectrograms of the IEMOCAP dataset. The suggested deep learning method contains convolution layers for learning discriminative features from spectrograms. The performance of the proposed framework is evaluated on standard performance metrics, which show promising real-time results for PD patients.

***Index Terms*—** Parkinsonian, Speech emotions Recognition, Deep Learning, Lightweight, Healthcare AI, Classification.

1. INTRODUCTION

Speech Emotion is a psychological condition linked to the neurological system. It is what a person feels inside as a response to his environment. The study of emotions in speech is becoming increasingly popular due to recent advancements in artificial intelligence (AI), which have made applications for the automatic identification of SEs [1, 2]. A person's emotion can be recognized differently [3]; one is the tonal properties of the voice. Analyzing speech signals using conventional signal processing techniques is challenging due to various attributes and frequencies that variate according to SEs. Contrarily, AI-based techniques play a significant role in developing intelligent systems to understand emotions. Likewise, smart healthcare centers recognize and analyze patients' emotional states using AI [4]. Healthcare professionals are looking for the best ways to understand patient voices for better diagnosis and treatment. State-of-the-art healthcare organizations seek to engage their patients during discussions in different phases of the patient journey [5]. Patient voice can help healthcare providers better understand the patients' conditions. SER from the human voice, particularly for a person with neurological disorders like PD, is a difficult task due to unexpected conditions of the voice of patients.

PD refers to the malfunction and cell death of the dopamine-producing area of the brain. Dopamine is a neurotransmitter that sends signals from the brain to nerve cells. It is responsible for controlling many of the body's motions. The oversight of the larynx, as well as the respiratory system [6], are both negatively affected by PD pathology, which has an impact on the voice production system. The motor speech disorder hypokinetic dysarthria is present in about 90 percent of persons with PD [7]. This condition is marked by monotone pitch and volume, sudden bursts of speech, and an expedited and varied verbal tempo. These characteristics of Parkinsonian voice alter listeners' perceptions, and as a consequence, PD voices are often regarded as unusual and observed as ”sad” [8]. After the emergence of DL and Machine learning in various domains like [9–13]

Several studies have discovered that parkinsonians are not only in their sense of emotions given from healthcare sessions but also in their perception of emotions by speech expressions [14]. In this regard, many researchers have deployed different approaches for the enhancement of this field. For instance, Almeida et al. [15] proposed a machine learning (ML) based approach for PD SER. They used a dataset created from two vocal tasks and then split the data into two modalities: speech and phonation. The raw data were preprocessed, and features were extracted by deploying eighteen conventional feature ex-

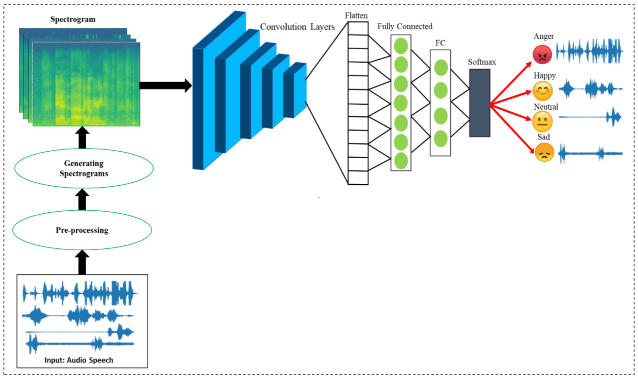

Fig. 1: ISER-Net framework is divided into three sub-sections which include 1). The dataset consists of four classes which include Anger, happy, neutral, and sad. 2). The second one is training in which we load 80% of data from the dataset for training purposes. The training phase contains sub-steps which include pre-processing in which preprocess the data by using resizing and denoising. Lastly, model training in which the ISER-Net is trained on the spectrograms of the speech. 3). Testing, after training models we evaluated the trained model by checking the performance on unseen data. In this step we load the remaining 20% data passed from pre-processing steps as in the training phase and observed the prediction of the model.

traction techniques. They claimed that Phonation audio has provided comparatively better results than the Speech modality. Furthermore, Zisad et al. [16] developed a CNN-based method for SER of neurological disordered persons. They deployed Ryerson Audio-Visual Database (RAVDESS) for the comprehensive experiments. Using the audio cardioid (AC) microphones in the vocal mode, the Yaffe (YA) set of features with a 1-nearest neighbor (K1) classifier performed the best. It is observable that CNN achieved good performances in many domains [17] [18]. However, the deployed features extraction techniques in [16] are conventional and they did not consider audio fusion before moving toward the final prediction. Recently, Sechidis et al. [4] deployed the MOE model by modifying an existing ML architecture that is able to infer the speaker's emotional state. After training, they use it to measure the extent to which people with PD and their matched healthy controls seem sad in previously unseen voice recordings. It is observable that these four classes are important to develop an efficient system as referenced above. However, they targeted only the sad class which is not enough to develop a perfect intelligent system for SER of Parkinsonians.

To bridge the research gap, we proposed a novel lightweight CNN-based framework entitled intelligent speech emotions recognition network of Parkinsonians (ISER-Net) that is able to recognize all necessary emotional states of Parkinsonians. We used the specific number of strides in each convolution layer instead of a pooling layers technique to extract high-level features from spectrograms of voice signals. We find hidden patterns of speech signals in convolution layers. Several experiments were carried out using the IEMOCAP dataset to efficiently train the model. We evaluate the model on the voice data of Parkinsonians acquired from Youtube. The ISER-Net outperforms previous state-of-the-art techniques in terms of unit performance.

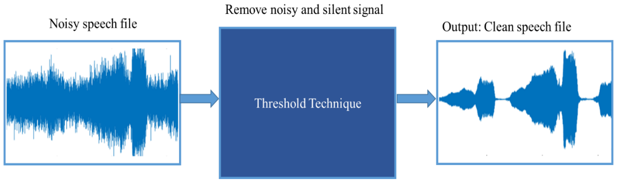

Fig. 2: Block diagram of pre-processing with an adaptive threshold technique

The major contributions of our research work are given below:

- We developed a lightweight CNN-based ISER-Net for Parkinson's patients, which can assist a doctor in treating Parkinson's patients as an automatic speech monitoring system of parkinsonians. To the best of our knowledge, the proposed framework is the first one which focused on the necessary emotional states for PD rehabilitation contrary to work in [4].
- Data Preprocessing before modeling plays a crucial role in the final accuracy of the proposed SER system. We utilized the adaptive thresholding method as a preprocessing that removes silent portions and noises from audio data and leaves only important voices in the input data.
- The proposed ISER-Net contains only five lightweight CNN layers and has only a 5MB size which can be easily displayable on resource constraint devices in the future. Extensive experiments empirically validated the superiority of the proposed framework.

The experimental portion of this study contains experiments and discussions of the suggested framework. The rest of the article structure is divided as follows: Section 2 elaborates on the proposed methodology. Experimental results and discussion are given in section 4. Section 5 concludes the paper and gives future directions of research.

2. PROPOSED METHODOLOGY

In the proposed system, the data are refined using adaptive thresholding. After that, feature learning techniques for spectrograms obtained from speech data are comprehensively discussed. SER is based on the automated learning of robust and discriminative features from spectrograms. We have illustrated the suggested CNN-based SER framework and its key components with a detailed explanation.

2.1. Pre-processing

Data preprocessing is a crucial step in preparing data to achieve high accuracy. In the proposed framework, the model starts preprocessing by taking audio data from the dataset. An adaptive wavelet thresholding function is introduced to enhance speech performance and the accuracy of automated speech recognition (ASR). The utilized adaptive threshold based on the wave [19] cleans the audio signals from the background noises, unnecessary information, and silent portions [20]. Here, we use a direct relation approach to determine the correlation between energy and amplitude in voice signals. An energy amplitude connection indicates that there is a correlation between the energy carried by a wave and its amplitude. The amplitude of a wave is a measure of its energy level, with larger values indicating more powerful waves and smaller values indicating weaker ones. The amplitude of a wave is defined as the largest increase by which a central component is moved away from its initial position. The following justification justifies using the energy-amplitude connection to filter out noise from audio transmissions. This procedure consists of three steps:

1 Read the audio file at 16,000 sampling rates step by step.

2 Following the discovery of the energy-amplitude connection in waves, we calculate the maximum amplitude in each using Equation (1) and run it through an appropriate threshold to filter out the noise and retain the salient portions.

3 we recreate it using the same sample rates in order to remove the noise and inaudible signals from the original audio file.

$$D = A \times \sin(2 \times \pi \times f \times t) \tag{1}$$

In Equation 1, D represents the particle's displacements, f is the frequency with respect to time t, and A denoted the signal's peak or amplitude. Figure 2 depicts a block schematic of the pre-processing.

2.2. Generating Spectrograms

One of the difficult challenges of 2D networks for SER is determining the dimensions of the voice signal. The primary goal of this study is to use a lightweight CNN model to learn rising features from voice signals, thus we need to transform the speech signal's original one-dimensional version into a suitable two-dimensional representation for 2D CNN. The spectrogram is the most accurate and useful two-dimensional representation of auditory voice signal, which show how strong a signal is at various frequencies. To visually depict frequency information over time intervals in voice signals, the short-term Fourier transformation (STFT) is employed. By employing STFT to chop up a lengthier voice signal into equal-length segments (frames), and then using fast Fourier transformation (FFT) on those clips to get the Fourier spectrum. The x-axis of a spectrogram depicts time, while the y-axis shows the frequencies over a short interval of time. The spectrograms improve SER performance as spectrograms include a plethora of information and discriminative features; that is why we are using spectrograms instead of using voice data signals directly to 1D CNN. The basic concept is to extract high-level discriminative features from voice signals. In this regard, the spectrograms are suitable to extract visual features of the speeches. In speech signal S (t, f), Spectrogram (S) comprises numerous type frequencies (f) over different times (t). The spectrograms were then fed to the proposed DL model for training. The given Figure. 3 shows samples of extracted spectrograms of each audio file using STFT.

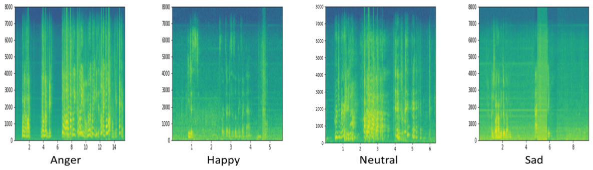

Fig. 3: Samples of the 2D generated spectrograms from the speech data

Table 1: Hierarchical Layering structure of the proposed CNN model.

Layers	Kernel Size	No. of Neurons	Activation $f(.)$	Dropout rate
$Conv2D_1$	9 × 1	10	ReLU	-
$Conv2D_2$	3 × 1	10	ReLU	-
$BatchNorm_1$	-	-	-	-
$MaxPool_1$	2 × 1	-	-	-
$Conv2D_3$	3 × 1	20	ReLU	-
$Conv2D_4$	3 × 1	20	ReLU	-
$MaxPool_2$	2 × 1	-	-	-
$BatchNorm_2$	-	-	-	-
$Conv2D_5$	3 × 1	20	ReLU	-
$MaxPool_3$	2 × 1	-	-	-
$BatchNorm_3$	-	-	-	-
Flatten	-	-	-	-
$Dense_1$	-	-	-	-
$BatchNorm_4$	-	-	-	-
$Dense_2$	-	-	-	-

3. MODEL ARCHITECTURE

CNN is a hierarchical neural network with a series of layers for extracting high-level features from low-level raw pixel data. It extracts high-level features from images using the number of kernels, and these features are then used to train a

Table 2: The suggested CNN model's training performance on raw spectrograms.

Emotion Nature	Precision	Recall	F1 Score
Anger	0.81	0.62	0.70
Happy	0.91	1.00	0.95
Neutral	0.67	0.88	0.76
Sad	0.98	0.84	0.90
Weighted Avg.	0.84	0.83	0.83
Unweighted Avg.	0.84	0.82	0.82

Table 3: The suggested CNN model's training performance on clean spectrograms.

Emotion Nature	Precision	Recall	F1 Score
Anger	0.98	0.99	0.99
Happy	0.98	1.00	0.99
Neutral	0.99	0.96	0.97
Sad	1.00	1.0	1.0
Weighted Avg.	0.99	0.99	0.99
Unweighted Avg.	0.99	0.99	0.99

CNN model for classification. The first layer which extracts features from an input image is convolution. This layer executes a process known as ”convolution.” By learning image features with small squares of input data, convolution retains the connection between pixels. It is an operation with two inputs: an image matrix and a kernel or filter. The ReLu function is employed as an activation function in a convolution layer to create nonlinearity and maintain feature values within a specific range. After convolutional layers, pooling layers (PL) are applied to extract the more activated features from feature maps and reduce computational model complexity. Moreover, dropout and batch normalization layers are employed to avoid overfitting and normalizing input values. CNN layers are mostly connected after feature extraction with fully connected (FC) layers, in which every neuron in the input layer is linked to every neuron in the next layer. FC layers are primarily used to extract global features, which are then fed into a SoftMax classifier to determine the probability for each class. CNN arranged all these CLs and PLs followed by FC layers in a hierarchical order. Finally, a Softmax layer applies this representation to give the probability of the final classification task. We have suggested the lightweight CNN model for SER of PD patients, shown in Figure 2. The network receives input of size 100 x 100 spectrogram images produced from emotional speech signals. The proposed framework mainly consists of five convolution layers and two dense layers. Technical details of the proposed model are given in Table 1.

4. EXPERIMENTAL RESULTS

The evaluation performance of the suggested approach is presented in this section. First, we will go through the experimental setup, then the dataset utilized in the model's training and assessment, and the evaluation matrix, ablation analysis, and real-time testing. All of these stages are detailed in the sections below.

4.1. Experimental Setup

All experiments were carried out in a Python 3.7 virtual environment installed on a PC with the specifications of Windows 10 OS, having GTX GeForce TITAN 1070 graphic processing unit (GPU) with a memory of GB, the processor of intel ® X5560, and clock speed of 2.80GH. Further, different frameworks and libraries are utilized during training; the proposed framework is utilized for training as a backend TensorFlow-GPU and a frontend Keras-GPU of versions 2.0.0 and 2.34, respectively. The categorical cross-entropy loss function and Adam optimizer with an initial learning rate of 0.0001 are used to calculate the loss of the model and update its weights while training, respectively. Moreover, we trained the proposed model on a mini-batch of size 30 for 100 iterations of epochs, which took almost two hours to complete the training of our proposed framework. Fig.4 shows the training and loss graph of the proposed model.

4.2. Dataset

It is generally believed that a lack of training data causes poor prediction. The primary barrier is the availability and accessibility of well-suited datasets for DL tasks. There are databases with millions of samples in domains like image or speech recognition, such as ImageNet with 14 million samples and Google AudioSet with 2.1 million samples. However, several SER databases have a limited amount of samples. Furthermore, the datasets are often made up of discrete emotional speech utterances, this is not always the case in real-life circumstances since overlaps often occur across speaker speech flows. As a result, models built on discrete utterances would perform poorly in continuous speech environments. We can also create a superset by merging some dataset classes, e.g., excited and happy features are almost the same in some acted datasets. Extensive research works are conducted to develop intelligent SER-based systems for PD but the datasets are not available due to patients' privacy. However, the IEMOCAP dataset is a multimodal and multi-speaker database collected by the Speech Analysis and Interpretation Laboratory (SAIL) at the University of Southern California (USC) [21]. The dataset is divided into five sessions, and each session has two actors who record a script in different emotions like anger, happy, disgust, sadness, fear, neutral, and excitement. Since our target is four desired classes, so used only four emotions (anger, neutral, happy, and sad) for the model's training. Our research aims to design a system that can accurately recognize common SEs of PD patients such as anger, happiness, normal, and sad. The given four classes contain 4180 total speech files, and each class consists of 1045 voice files. We have taken 500 samples from the excited class of the same dataset and merged them into the happy class. This is due to the similar characteristics of both classes. A happy class is essential for PD patients because the only psychological treatment for PD patients is to make them happy as they are always sad.

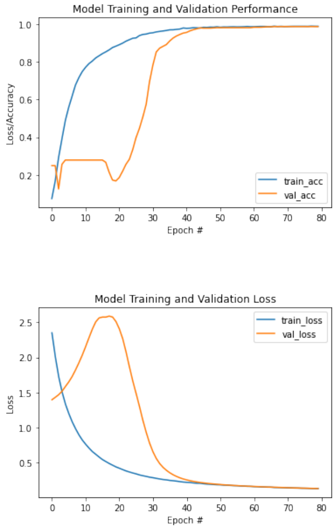

Fig. 4: Accuracies and losses of the proposed model.

4.3. Evaluation

Experiments were conducted using the IEMOCAP dataset to evaluate the performance of the proposed technique for the emotion recognition of PD patients. The two sets of experiments were conducted based on raw and clean spectrograms and image-based validation of the proposed CNN. The results in detail are discussed in the following sub-sections.

4.4. Raw and Clean Spectrograms

We have compared the results of raw spectrograms and clean spectrograms using the data of the IEMOCAP dataset. The proposed model achieved an accuracy of 82% and 99% on the raw spectrogram and clean spectrogram, respectively. Table 2 illustrates the results based on raw and clean spectrogram in detail.

4.5. Validation on image-based mode

We found an improvement in the model's performance during the experiment when the image size was increased. By doing many experiments, we gained that our proposed model achieved good accuracy on the image of size 100×100. Table 4. Shows the statistical details of experiments based on the size of input spectrogram images.

Table 4: The performance of the proposed CNN model on spectrograms experiments

Image size	Training Accuracy	Validation Accuracy	Model size
50 × 50	88.99	87.25	2.5 MB
60 × 60	94.93	94.40	3.0 MB
70 × 70	96.45	95.83	3.5 MB
80 × 80	97.29	97.50	4.0 MB
90 × 90	97.64	98.09	4.5 MB
100 × 100	98.84	98.69	5.0 MB

4.6. Real-time testing of PD patients speeches

As our ultimate goal is to develop an SER system for PD patients. Therefore, we downloaded the interview sessions from a youtube channel entitled The Michael J. Fox Foundation for Parkinson's Research. We extracted audio modalities from the video sessions. The dimensions of the data are settled according to the input requirements of the proposed ISER-Net. The model achieved convincing results on the real-time data of the patients. One plausible explanation for the suggested model's impressive results is that it was trained on a dataset that is more analogous to the YouTube data.

5. CONCLUSION

This study aimed to develop an SER system for PD patients. The SER literature confronts so many problems in improving recognition accuracy while also reducing the entire model's computing complexity. We developed a lightweight CNN-based ISER-Net with certain salient feature extraction to increase the accuracy and minimize the overall computing cost of the SER model in response to these problems. To reduce noise and silence signals from voice, we employed a dynamic adaptive threshold approach which improved the performance of the network. The enhanced speech signals are then transformed into spectrograms to improve the accuracy of the proposed model while reducing its computing complexity. The suggested model's performance is assessed using the IEMOCAP dataset. In the future, we will explore attention-based networks by employing more datasets to make more robust networks. The model will be deployed on edge devices for real-time use to assist doctors in clinical treatments of parkinsonians.

Acknowledgement

The European Union funded this research through the Horizon 2020 Research and Innovation Programme, in the context

Table 5: Class wise accuracy report of the proposed lightweight CNN model

Emotion Nature	Precision	Recall	F1-score	Support
Anger	0.98	0.99	0.99	234
Happy	0.98	1.00	0.99	196
Neutral	0.99	0.96	0.90	209
Sad	1.00	1.00	1.00	200
Macro Avg.	0.99	0.99	0.99	839
Unweighted Avg.	0.99	0.99	0.99	839

of the ALAMEDA (Bridging the Early Diagnosis and Treatment Gap of Brain Diseases via Smart, Connected, Proactive and Evidence-based Technological Interventions) project under grant agreement No GA 101017558.

6. REFERENCES

[1] Soonil Kwon et al., “A cnn-assisted enhanced audio signal processing for speech emotion recognition,” *Sensors*, vol. 20, no. 1, pp. 183, 2020.

[2] Xiangjun Chen, Zhaohui Wang, Yuefu Zhan, Faouzi Alaya Cheikh, and Mohib Ullah, “Interpretable learning approaches in structural mri: 3d-resnet fused attention for autism spectrum disorder classification,” in *Medical Imaging 2022: Computer-Aided Diagnosis*. SPIE, 2022, vol. 12033, pp. 611–618.

[3] Muhammad Munsif, Mohib Ullah, Bilal Ahmad, Muhammad Sajjad, and Faouzi Alaya Cheikh, “Monitoring neurological disorder patients via deep learning based facial expressions analysis,” in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2022, pp. 412–423.

[4] Konstantinos Sechidis, Riccardo Fusaroli, Juan Rafael Orozco-Arroyave, Detlef Wolf, and Yan-Ping Zhang, “A machine learning perspective on the emotional content of parkinsonian speech,” *Artificial Intelligence in Medicine*, vol. 115, pp. 102061, 2021.

[5] Pip Hardy, “An investigation into the application of the patient voices digital stories in healthcare education: quality of learning, policy impact and practice-based value,” *Belfast: University of Ulster*, 2007.

[6] KM Torsney and D Forsyth, “Respiratory dysfunction in parkinson’s disease,” *Journal of the Royal College of Physicians of Edinburgh*, vol. 47, no. 1, pp. 35–39, 2017.

[7] Juan Rafael Orozco-Arroyave, *Analysis of speech of people with Parkinson’s disease*, vol. 41, Logos Verlag Berlin GmbH, 2016.

[8] Abhishek Jaywant and Marc D Pell, “Listener impressions of speakers with parkinson’s disease,” *Journal of the International Neuropsychological Society*, vol. 16, no. 1, pp. 49–57, 2010.

[9] Altaf Hussain, Tanveer Hussain, Waseem Ullah, and Sung Wook Baik, “Vision transformer and deep sequence learning for human activity recognition in surveillance videos,” *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[10] Hikmat Yar, Tanveer Hussain, Mohit Agarwal, Zulfiqar Ahmad Khan, Suneet Kumar Gupta, and Sung Wook Baik, “Optimized dual fire attention network and medium-scale fire classification benchmark,” *IEEE Transactions on Image Processing*, vol. 31, pp. 6331–6343, 2022.

[11] Mohib Ullah, Sareer Ul Amin, Muhammad Munsif, Utkurbek Safaev, Habib Khan, Salman Khan, and Habib Ullah, “Serious games in science education. a systematic literature review,” *Virtual Reality & Intelligent Hardware*, vol. 4, no. 3, pp. 189–209, 2022.

[12] Muhammad Munsif, Habib Khan, Zulfiqar Ahmad Khan, Altaf Hussain, Fath U Min Ullah, Mi Young Lee, and Sung Wook Baik, “Pv-anet: Attention-based network for short-term photovoltaic power forecasting,” , pp. 133–135, 2022.

[13] Habib Khan, Ijaz Ul Haq, Muhammad Munsif, Shafi Ullah Khan, and Mi Young Lee, “Automated wheat diseases classification framework using advanced machine learning technique,” *Agriculture*, vol. 12, no. 8, pp. 1226, 2022.

[14] Lorinda C Kwan and Tara L Whitehill, “Perception of speech by individuals with parkinson’s disease: a review,” *Parkinson’s Disease*, vol. 2011, 2011.

[15] Jefferson S Almeida, Pedro P Rebouças Filho, Tiago Carneiro, Wei Wei, Robertas Damaševičius, Rytis Maskeliūnas, and Victor Hugo C de Albuquerque, “Detecting parkinson’s disease with sustained phonation and speech signals using machine learning techniques,” *Pattern Recognition Letters*, vol. 125, pp. 55–62, 2019.

[16] Sharif Noor Zisad, Mohammad Shahadat Hossain, and Karl Andersson, “Speech emotion recognition in neurological disorders using convolutional neural network,” in *International Conference on Brain Informatics*. Springer, 2020, pp. 287–296.

[17] Noman Khan, Amin Ullah, Ijaz Ul Haq, Varun G Menon, and Sung Wook Baik, “Sd-net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network,” *Journal of Real-Time Image Processing*, vol. 18, no. 5, pp. 1729–1743, 2021.

[18] Muhammad Munsif, Hina Afridi, Mohib Ullah, Sultan Daud Khan, Faouzi Alaya Cheikh, and Muhammad Sajjad, “A lightweight convolution neural network for automatic disasters recognition,” in *2022 10th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2022, pp. 1–6.

[19] Feixiang Yan, Hong Zhang, and C Ronald Kube, “A multistage adaptive thresholding method,” *Pattern recognition letters*, vol. 26, no. 8, pp. 1183–1191, 2005.

[20] Yasser Ghanbari and Mohammad Reza Karami-Mollaei, “A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets,” *Speech communication*, vol. 48, no. 8, pp. 927–940, 2006.

[21] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan, “Iemocap: Interactive emotional dyadic motion capture database,” *Language resources and evaluation*, vol. 42, no. 4, pp. 335–359, 2008.