CNN Classification of Parkinson's Disease using STFT Spectrum of User's Running Speech

Rani Kumari
School of Electronics Engineering
Vellore Institute of Technology, Vellore
Tamil Nadu, India
rani.kumari2020@vitstudent.ac.in

Prakash Ramachandran
School of Electronics Engineering
Vellore Institute of Technology, Vellore
Tamil Nadu, India
prakash.r@vit.ac.in

2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT) | 978-1-6654-9360-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECCT56650.2023.10179774

***Abstract*— Speech deterioration is a well-established indicator for early detection of Parkinson disease (PD). The remote monitoring of PD symptoms is practicable by analyzing the person's speech. This study demonstrates that deep learning of speech spectrum images of Running speech from Parkinson's disease (PD) patients and healthy controls (HC) results in a convolution neural network model with high accuracy for classifying PD. Earlier work for PD detection was totally based on the computation of speech features. Complicated speech feature computations take time and require a large amount of memory access. Using a spectrum image representation of speech in a deep convolution neural network gives very high accuracy without the need for feature computation, and selection in this work. This is possible because the convolution layer learns the features from the spectrum automatically and in this work, it is proposed to work with PD speech database for Running speech. In this case, the short time Fourier transform (STFT) is a good choice for representing speech as images. The training accuracy achieved in the experiment is 99.48% and validation accuracy are 79.10%. This method performs best and is a well-founded method for classifying PD from running speech signals.**

Keywords*— *Parkinson Disease (PD), Convolution Neural Network (CNN), Speech, Healthy Control (HC), Short Time Fourier Transform (STFT), Running Speech.

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative ailment. There are numerous symptoms like speech disorder, tremor, sleep disorder, dysarthria, and depression. Speech problem is one of the first symptoms of Parkinson's disease [1]. It has a negative effect on the lives of PD patients because the disease worsens with age [2]. In terms of clinic, the early detection of Parkinson's disease necessitates subjective assessment. Current subjective assessments indicate that there is an urgent requirement for the development of diagnostic support instruments for the detection and careful monitoring of Parkinson's disease [3]. Slight speech deterioration has been shown to be a sensitive marker of early degeneration [4]. Work [2] demonstrates that early diagnosis of PD. The statistical features were extracted and observed, and the statistical Kruskal-Wallis test was used to select the best feature set. These characteristics were used to differentiate between PD and healthy controls (HC). In this paper, it is attempted to analyse the PD speech pattern using a machine learning approach. These machine learning methods rely on spectrum-based parameters. Artificial Intelligence (AI) is used to train the spectrum-based parameters to classify speech input as PD positive or negative [5-16]. Recent PD classification research emphasises the significance of machine learning. Machine learning can handle large amounts of data for PD classification while maintaining high accuracy. Z. Karapinar Senturk (2020) considered feature selection task and used feature selection and classification processes for the detection of PD. Support Vector Machines (SVMs) and Artificial Neural Networks (ANN) were employed for classification during the experiment. Support Vector Machines (SVMs) with Recursive Feature Elimination performed better, with an accuracy of 93.84% [5]. This paper is arranged as follows: in Section II we introduce relevant state-of-the art research. In section III presents the proposed research approach. Section IV demonstrate the Parkinson's disease speech database, Section V on spectrum representation of Speech. Section VI inaugurates Machine learning experiment, Section VII elaborates results and discussion. Finally, concluding remarks on section VIII.

II. RELATED WORK

Abhishek M.S. et al (2020), investigated the feature extraction and prediction algorithm employed to predict PD. They used the k-nearest neighbors (KNN) classifier and the support vector machine (SVM) classifier. Jittering, shimmering, and voice signal frequency were considered for analysis. They achieved accuracy from algorithms ranging from 85% to 95%, demonstrating that machine learning algorithms can be employed to predict PD through voice analysis [9].

Radha N et al (2021), focused on the development of PD detection by acoustic features such as chroma STFT, root mean square (RMS), spectral centroid and bandwidth, roll-off and zero crossing rate. They used convolution neural network (CNN), artificial neural network (ANN), and hidden Markov model (HMM) to differentiate PD from HC. The spectrogram of speaker information and acoustic features are used to train the CNN. Acoustic features are the only ones used to train ANN and HMM. CNN has an 88% accuracy with spectrograms and a 93.5% accuracy with acoustic features. The accuracy of ANN was 96%, while the accuracy of HMM was 95.2%. ANN outperformed the rest [10].

F.M. Javed Mehedi Shamrat et al (2019), For the detection of PD, three AI methods were used. SVM, KNN, and logistic regression (LR) are the names of the methods. The classifiers' performance was evaluated using recall, precision, and f1 score. The preliminary results show that the SVM performed better than the other two classifiers. It got 100% accuracy [11].

Iqra Nissar et al (2019), The outcome of feature type selection was studied, and the feature names are Mel-frequency cepstral coefficients (MFCC) and Tunable Q wavelet transform (TQWT). They considered several machine learning models in their work, containing Logistic Regression, Nave Bayes, KNN, Random Forest, Decision Tree, SVM, Multi- layer perceptron (MLP), and Extreme

Gradient Boosting (XGBoost), Minimum Redundancy Maximum Relevance (mRMR) and Recursive feature elimination (RFE) techniques were also used for feature selection. When both MFCC and TQWT were employed, the XGBoost model with mRMR feature selection performed better than the other models with an accuracy of 95.39% and precision, recall, and F1 score values of 0.95 each [12].

Alex Frid Edmond J. Safra et al (2014), With the help of machine learning techniques an automatic method was developed for determination and classification of PD directly from natural speech. In the voice analysis, the automatic system did not need any human intervention. This technique demonstrates that deep human expertise in speech signal selection, choice and combination can be eliminated by an automatic task on the auditory signal itself [15].

III. PROPOSED WORK

In this paper it is proposed to investigate the method of PD classification by using Running speech as spectrum images. The spectrum image contains all frequency variations in time plane. In this work, convolution neural network (CNN) is proposed to learn useful features from these images by itself using convolution filter layers and fully connected neural network with hidden layers. This proposed method eliminates the need for time-consuming mathematical computation to determine various speech parameters while getting training accuracy and validation accuracy. The proposed spectrum method is Short Time Fourier Transforms (STFT).

IV. PARKINSON'S DISEASE SPEECH DATABASE

In developing the proposed CNN model, King's College London (KCL) hospital, Denmark Hill, Bixton dataset is used. It is a running speech dataset that was available for both PD and HC patients [17]. They collected data by recording voices in an examination room with a reverberation time of 500ms. They used a smartphone to record voice and created one app. The app's name was "Toggle Recording App." They gathered data in "WAVE" format. Table I contains dataset details, such as the total number of voice samples in training and testing. As a result, we can predict its classes. There are a total of 1009 voice samples, with 765 for training and 244 for testing.

TABLE I. DATASET OF RUNNING SPEECH (TRAINING AND TESTING)

Samples	*PD*	*HC*	*Total*
Total Voice Samples	394	615	1009
Voice Samples (Training)	300	465	765
Voice Samples (Testing)	94	150	244

V. SPECTRUM REPRESENTATION OF SPEECH

It is proposed in this paper to use Deep convolution neural network (DCNN) to obtain results in terms of images and to represent running speech as a 2D image in the time-frequency plane. These images will be used for PD classification later on. The short time Fourier transform (STFT) will be used for this purpose.

A. Images of STFT

The voice sample files in Table I are processed to generate an image representation with the help of short time Fourier transform (STFT), these voice sample files are down sampled at 8kHz sampling rate, and the sliding frame size is set to 128 with a 75% overlap with the neighbour frame. The fast fourier transform used in STFT has a length of 128, and the magnitude spectrum of the frame is in the form of column wise, and this process is repeated for all frames to produce a spectrum image of size N x 128, where N is the number of frames. Fig. 1 and fig. 2 show the STFT image obtained for a sample PD file and HC file, respectively. The procedure is repeated for all of the voice sample files in the dataset, yielding 1009 images.

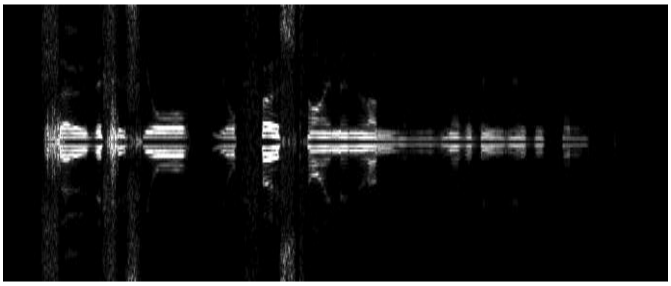

Fig. 1. Image of Running Speech for STFT (PD).

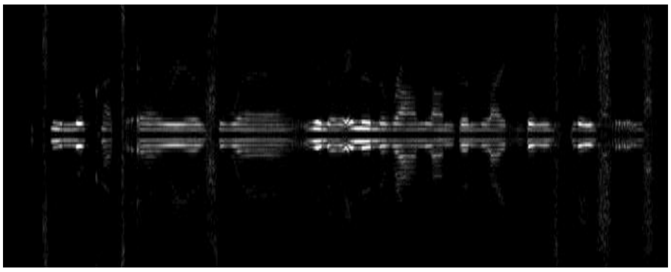

Fig. 2. Image of Running Speech for STFT (HC).

VI. MACHINE LEARNING EXPERIMENT

Convolutional neural networks are artificial neural networks that are used in speech recognition, image classification, natural language processing, face recognition and in healthcare sectors. It appears to be similar to a basic neural network. CNN has parameters that can be learned, such as weights and biases. It has an input layer, an output layer, a convolution layer, and numerous hidden layers. The input layer is the first layer, the output layer is the last layer, and the layers in between are pooling and convolutional. The convolutional layer reduces image dimensionality without sacrificing information [18]. CNN is used in our research to classify speech-based spectrum images.

It provides good accuracy and validation accuracy. Keras and Tensor flow are used to create models and layers for this classification process. Several sub-steps are involved in the creation of models and layers, such as convolution, pooling, and flattening, then with the step of fully connected layer. In this case, four hidden dense layers of different lengths are combined with the ReLU activation function. The target size is 64, the batch size is 32, the class mode is binary, and the adam optimizer is used. The method needs one output to predict two classes and used the sigmoid activation function. Because the sigmoid activation function is employed in binary classification problems and its values are between 0 and 1.

VII. RESULTS AND DISCUSSION

CNN is used for training and testing of proposed speech-based spectrum images. Using the STFT method for the final epoch, 99.48% training accuracy and 79.10% validation accuracy are achieved. The accuracies of the model and model loss are plotted against epochs. Fig. 3, fig. 4, fig. 5 & fig. 6 for STFT shows plot of training accuracy, validation accuracy, model accuracies and model loss. The graph shows that the representation performs well, and to get the performance of classification model to further analyze it, such as dimension of images, images of PD voice (Training), images of HC voice (Training), images of PD voice (Testing), images of HC voice (Testing), Accuracy and validation accuracy. Table II contains overall performance of the classification model.

TABLE II. OVERALL PERFORMANCE OF CLASSIFICATION MODEL

Technique	**STFT**
Dimension of Images	Nx128
Images of PD voice (Training)	300
Images of HC voice (Training)	465
Images of PD voice (Testing)	94
Images of HC voice (Testing)	150
Accuracy	99.48%
Validation Accuracy	79.10%

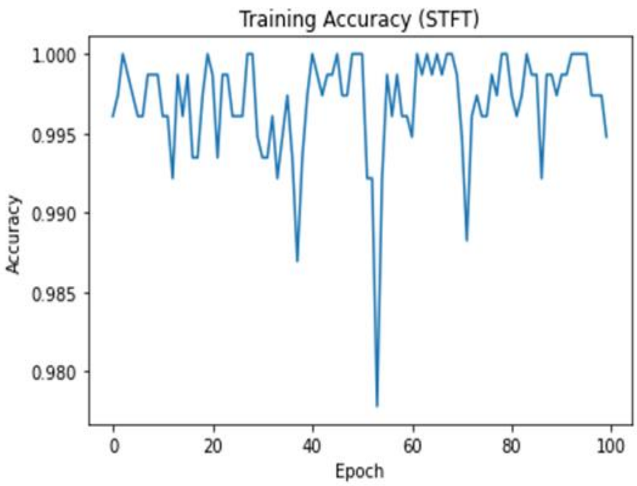

Fig. 3. Training Accuracy of STFT.

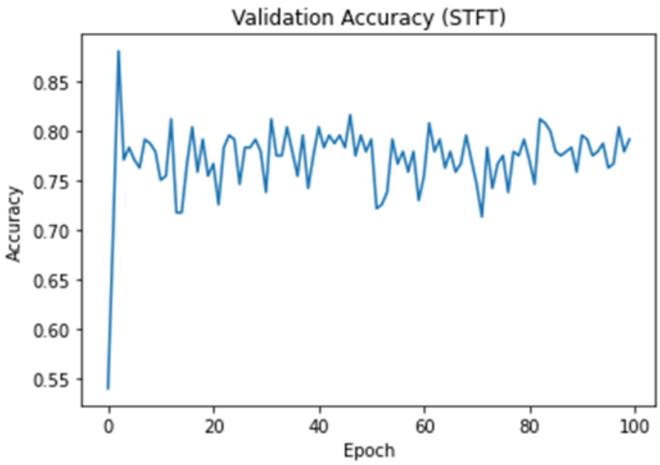

Fig. 4. Validation Accuracy of STFT.

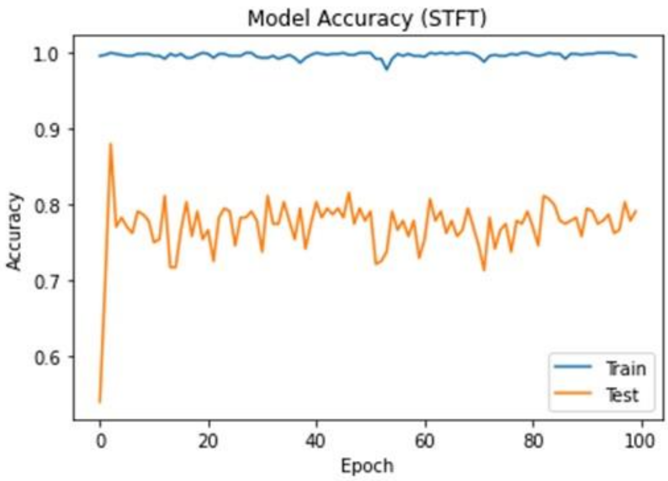

Fig. 5. Model Accuracies of STFT

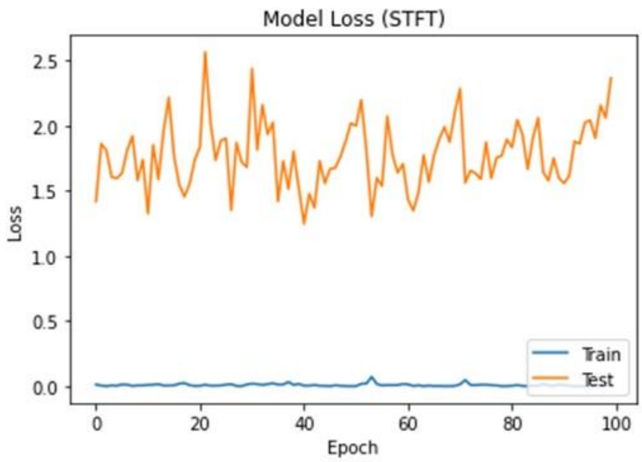

Fig. 6. Model Loss of STFT.

VIII. CONCLUSION

The automatic feature selection capability of convolution layer of DCNN is utilized in a numerical experiment to classify PD from running speech spectrum images without the need of separate feature computation stage. It is demonstrated that the spectral speech representation method outperforms the feature-based learning method. The DCNN-based method has many advantages over the feature-based learning method such as time consumption, memory access requirements, and bias in feature selection. STFT is used as a spectral imaging method in this case. In terms of training accuracy and validation accuracy, the spectral imaging method performed well with 99.48% training accuracy and 79.10% validation accuracy. This paper presents a powerful machine learning approach to classifying PD from speech. Complications can be reduced in the near future for patients living in remote areas with poor medical facilities and benefit the vulnerable people.

REFERENCES

[1] M. Yang *et al.*, "Hierarchical Boosting Dual-Stage Feature Reduction Ensemble Model for Parkinson's disease speech data," *Diagnostics (Basel)*, vol. 11, no. 12, p. 2312, 2021.

[2] S. J. Priya, A. J. Rani, M. S. P. Subathra, M. A. Mohammed, R. Damaševičius, and N. Ubendran, "Local pattern transformation-based feature extraction for recognition of Parkinson's disease based on gait signals," *Diagnostics (Basel)*, vol. 11, no. 8, p. 1395, 2021.

[3] S. Arora and A. Tsanas, "Assessing Parkinson's Disease at Scale Using Telephone- Recorded Speech: Insights from the Parkinson's Voice Initiative," *Diagnostics*, 11, 1892, 2021.

[4] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, p. 2000410, 2019.

[5] Z. Karapinar Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Med. Hypotheses*, vol. 138, no. 109603, p. 109603, 2020.

[6] R. Z. U. Rehman, S. Del Din, Y. Guan, A. J. Yarnall, J. Q. Shi, and L. Rochester, "Selecting clinically relevant gait characteristics for classification of early Parkinson's disease: A comprehensive machine learning approach," *Sci. Rep.*, vol. 9, no. 1, p. 17269, 2019.

[7] H. Hazan, D. Hilu, L. Manevitz, L.O. Ramig, S. Sapir, Early diagnosis of Parkinson's disease via machine learning on speech data, in: 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEE, 2012.

[8] G. Pahuja and T. N. Nagabhushan, "A comparative study of existing machine learning approaches for Parkinson's disease detection," *IETE J. Res.*, vol. 67, no. 1, pp. 4–14, 2021.

[9] Abhishek *et al.*, "Diagnosis of Parkinson's disorder through speech data using machine learning algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 69–72, 2020.

[10] N. Radha and R. M. & Sachin Madhavan, ""Parkinson's Disease Detection using Machine Learning Techniques," *International Journal of Early Childhood Special Education (INT-JECSE)*, vol. 30.

[11] F. J. M. Shamrat, M. Asaduzzaman, A. S. Rahman, R. T. H. Tusher, and Z. Tasnim, "A comparative analysis of parkinson disease prediction using machine learning approaches"," *International Journal of Scientific & Technology Research*, vol. 8, 2019.

[12] I. Nissar, D. Rizvi, S. Masood, and A. Mir, "Voice-based detection of Parkinson's disease through ensemble machine learning approach: A Performance study"," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 5, 2019.

[13] L. Berus, S. Klancnik, M. Brezocnik, and M. Ficko, "Classifying Parkinson's disease based on acoustic measures using artificial neural networks," *Sensors (Basel)*, vol. 19, no. 1, p. 16, 2018.

[14] A. Caliskan, H. Badem, A. Basturk, and M. E. Yuksel, "Diagnosis of the parkinson disease by using deep neural network classifier"," *Istambul University-Journal of Electrical & Electronics Engineering*, vol. 17, pp. 3311–3318, 2017.

[15] A. Frid, E. J. Safra, H. Hazan, L. L. Lokey, D. Hilu, and L. Manevitz, "& Sapir, S, "Computational diagnosis of Parkinson's Disease directly from natural speech using machine learning techniques," in *2014 IEEE International Conference on Software Science*, 2014, pp. 50–53.

[16] M. Anila and G. D. Pradeepini, "A Review on Parkinson's Disease Diagnosis using Machine Learning Techniques"," *International Journal of Engineering Research & Technology*, vol. 9, pp. 330–334, 2020.

[17] H. Jaeger, D. Trivedi, and M. Stadtschnitzer, "Mobile Device Voice Recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls." Zenodo, May-2019.

[18] A. A. Reshi *et al.*, "An efficient CNN model for COVID-19 disease detection based on X-ray image classification"," *Complexity*, 2021.