Parkinson's Disease Identification using Deep Neural Network with RESNET50

Anila M[1]
Research Scholar, Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

Pradeepini Gera[2]
Professor, Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

***Abstract*—Recent Parkinson's disease (PD) research has focused on recognizing vocal defects from people's prolonged vowel phonations or running speech since 90% of Parkinson's patients demonstrate vocal dysfunction in the early stages of the illness. This research provides a hybrid analysis of time and frequency and deep learning techniques for PD signal categorization based on ResNet50. The recommended strategy eliminates manual procedures to perform feature extraction in machine learning. 2D time-frequency graphs give frequency and energy information while retaining PD morphology. The method transforms 1D PD recordings into 2D time-frequency diagrams using hybrid HT/Wigner-Ville distribution (WVD). We obtained 91.04% accuracy in five-fold cross-validation and 86.86% in testing using RESNET50. F1-score achieved 0.89186. The suggested approach is more accurate than state-of-the-art models.**

Keywords—Parkinson's disease; speech impairment; artificial intelligence; RESNET50; deep learning; ht/wigner-ville distribution; 2D time-frequency

I. INTRODUCTION

Parkinson's disease (PD) affects 1% of persons over 60 [1]. Parkinson's disease is the second most prevalent neurological condition [2]. According to European and American epidemiological studies [3], PD affects men 1.5 times more than women. Parkinson's disease causes increasing disability that affects everyday life. Dopamine neurodegeneration induces motor and non-motor symptoms in the midbrain's substantia nigra pars compacta [3].

Early on with Parkinson's, voice problems affect 90% of individuals. Recent PD telemedicine research [4-6] focused on speech problems. These studies employed speech signal processing to evaluate Parkinson's disease. The generated characteristics were used to build reliable decision support systems. Parkinson's disease causes rigidity, resting tremors, bradykinesia, and postural instability [7, 8].

Many research that compares PD and healthy persons utilize a publicly accessible dataset [9] with 195 sound measures from 23 PD patients and eight healthy controls. Another publicly accessible PD telediagnosis dataset [6] utilized in the linked research includes 20 PD and 20 healthy participants' voice recordings. In all datasets, each voice recording contains fundamental vocal frequency, amplitude variation measurements, noise-to-tone component ratio, nonlinear dynamical complexity measures, and nonlinear fundamental frequency variation measures. With these parameters achieved accuracy is less than 85%, which is too less. Using deep learning methods, it can be improved.

The remainder of this research is organized as follows: the research analyzes existing automated speech analysis techniques in Parkinson's disease patients in the section "Related Work." It discusses the 1D to 2D transformation methods and the deep learning classification model in the section "Methodology." The section "Results and Discussion" addresses the employed dataset, classification performance, and statistical analysis results, while the section "Conclusions and Future Study" propose conclusions and future work.

II. RELATED WORK

PD speech analysis research includes recording activities. Sustained vowel phonation is prevalent since it is a regular work [10]. Other research focuses on continuous speech recordings of sentences, reading texts, and spontaneous speech to study prosody [11]. Few papers discussed single-word invention. Early research focused on Parkinson's patients [12]. This study examined spectral and cepstral properties taken from 24 isolated words and five vowels spoken by Colombian Spanish speakers. Scientists used an SVM with a Gaussian kernel to classify Parkinson's patients. They compared each set of attributes individually and all coefficients together. When all utterances and attributes were combined, word and vowel accuracy was 92% and 79%. Despite positive findings, the approach lacked pre-processing.

In [13] mentioned lonely words as a speaking difficulty. The database includes Spanish (50 HC, 50 PD), German (88 PD, 88 HC), and Czech speakers (20 PD, 16 HC). Different languages and feature sets were automatically used to classify HC and PD speakers. Each corpus was modeled using four factors to identify linguistic impairment. Using a radial basis SVM, the authors claimed the latter technique was robust, with 85 to 95% classification accuracy. Both papers used PC-GITA. None studied model generalization on a single dataset. The test models in [13] were optimized, resulting in unduly optimistic findings.

In [14], to explain speech signal non-linearities, the authors proposed isolated word modeling based on Hilbert Spectrum properties. The recommended coefficients, Instantaneous Energy Deviation Coefficients (IEDCC), exceeded the traditional acoustic characteristics when dealing with isolated individual syllables, outperforming them with accuracy ranging from 81 to 91%. The accuracy was 82%. However, the authors did not give the findings of the combined features. They used 20 Parkinson's patients and 20 healthy controls.

Little et al. [15] reported that each patient made six phonations. A soundproof audiology booth created by an industrial acoustics business caught the phonations. Human sound signals were recorded in a computerized speech laboratory. After recording phonations, they analyzed entropy and conventional and nonstandard measures. A kernel SVM classifier distinguished between healthy and Parkinson's patients. Researchers then used the Little et al. dataset to enhance their classifier. Bhattacharya et al. employed Weka and SVM [16]. Das employed neural networks, DMneural, regression, and decision trees to classify features [17]. Sakar et al. [18] employed SVM and feature selection. Polat identified Parkinson's disease patients using FCMFW and k-NN [19].

Sakar and Kursun [18] suggested an SVM model using mutual knowledge for PD feature selection. Speech characteristics were filtered and sorted using the highest relevance lowest redundancy approach (mRMR). Using the leave-one-individual-out strategy, their test on the UCI dataset obtained 81.53% accuracy and 92.75% using the bootstrap resampling method. It discovered four critical parameters associated with the voice impairment test for Parkinson's disease diagnosis, which additional studies may corroborate.

This paper proposes an approach for analyzing solitary words based on several signal processing and pattern recognition methods. We began with a pre-processing phase, followed by a 1D to 2D time-frequency conversion approach and a classification step. The PD sound series were converted into 2D time-frequency distribution diagrams to display additional information such as time, frequency, and energy. To translate 1D to 2D time-frequency distribution diagrams, a technique combining HT and WVD was presented. The hybrid analysis of the time and frequency flowchart is shown in Fig. 2. The approach first employs the HT to transform the original signal into analytic signals, followed by the analysis of time and frequency using the traditional WVD to generate the relevant time-frequency diagrams.

III. METHODOLOGY

The importance of implementation is to improve the recognition rate.

A. Method Outline

The 1D sound signal was reconstructed as 2D time-frequency diagrams in this work to gather information on the signal's time, frequency, and energy. The suggested method's flowchart is shown in Fig. 1. During data pre-processing, noises like baseline wander (BW) and 50/60 Hz power line noise are initially removed. To convert the 1D time series into 2D time and frequency diagrams, the HT-WVD technique is used. The dataset was randomly partitioned into three sets: training, validation, and testing. 70% of each category was chosen randomly for the training, 10% for the validation, and 20% for the testing.

The validation data is used to validate the performance of the learned deep learning models using a five-fold cross-validation technique. The training data is used to train multiple deep-learning models with a learning rate of 0.001. The ResNet50 was utilized for deep learning to classify the 2D time and frequency diagrams of the PD for classification tasks. The classification outcomes for the test data for the optimal neural network model are shown.

B. Data Pre-processing

1) Removing noises: Our database of PD recordings comprises familiar sounds such as 50/60 Hz powerline noise and BW. As a result, noise should be deleted from the recordings to ensure accuracy. This research used a median filter approach to remove BW from PD recordings first, followed by a wavelet transform algorithm to remove powerline noise. The PD recordings' baseline curves were retrieved using the median filtering technique. This entails running a median filtering algorithm [20], then subtracting the original PD recordings from the derived baseline maps to generate the new PD recordings without BW. This research employed the wavelet transform to remove additional noise. The Daubechies db5 wavelet is used to create a three-level wavelet decomposition.

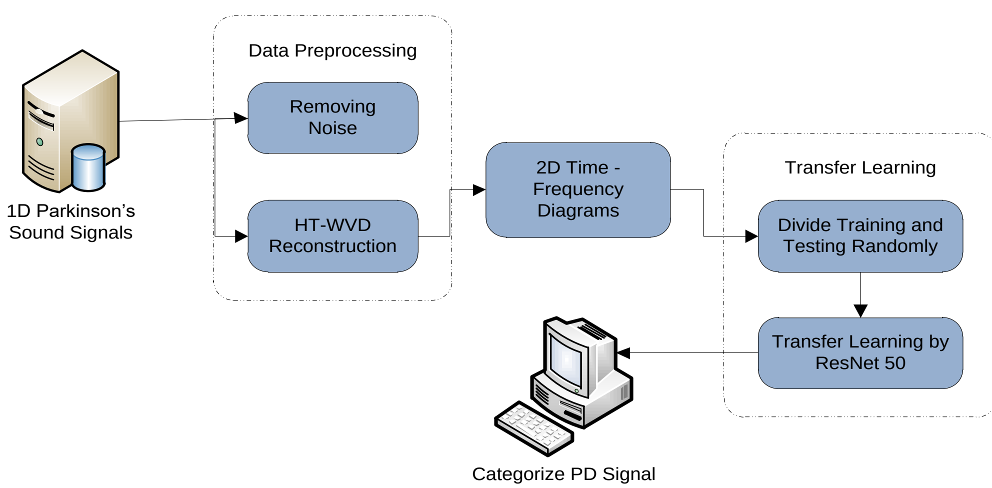

Fig. 1. The Proposed Method of Architecture.

2) Hybrid analysis of time and frequency: Because PD sound signals are 1D non-stationary time series, they cannot display the connection between time, frequency, and energy within the time series. Analysis of time and frequency techniques may convert a one-dimensional signal into a two-dimensional density function of time and frequency, revealing how many frequency components are present in the signal and how each component changes over time. Temporal frequency approaches such as the WVD and HT are widely employed for signal analysis [24].

3) One can utilize the bilinear Wigner-Ville distribution to evaluate nonlinear signals and reflect energy distributions in both the time and frequency domains. The definition of WVD is as follows:

$$W(t,f) = \int_{+\infty}^{-\infty} x\left(t+\frac{\tau}{2}\right) x^{*}\left(t-\frac{\tau}{2}\right) d\tau \qquad (1)$$

4) X* is a complex conjugate, and x(t) is the signal. The study [21] explains every step of the procedure. An actual signal can be transformed into an analytical signal using the HT. The HT is described as follows for a discrete time series x(k):

$$H(k) = xH(k) = FFT\text{-}1(f(k) * h(i)) \qquad (2)$$

where the element-wise product of f and h is calculated using the symbol *. The inverse fast Fourier transform, or FFT-1, is represented by vector f, which is the FFT of the y(k), and h is defined as follows:

$$h = \begin{cases} 0\ for\ i\ =\ (N/2)\ +\ 2,\ldots,N \\ l\ for\ i\ =\ 1,\ (N/2)\ +\ 1 \\ 2\ for\ i\ =\ 2,3,\ldots,\ (N/2) \end{cases} \qquad (3)$$

Thus, the analytic signal z(k) of the discrete time series x(k) can be represented as

$$z(k) = x(k) + jxH(k) \qquad (4)$$

5) Where j is the complex imaginary number, i.e., j = (-1)1/2, the HT is detailed in [22].

6) Because the dimensional difference of 1D signals does not affect the energy distribution of 2D time and frequency diagrams, 1D PD signals do not need to be normalized when translated into 2D time-frequency diagrams.

7) Fig. 3 to Fig. 6 shows the signal in the time domain, the signal spectrum, and the signal's spectrogram, which is the 2D time frequency of Parkinson's audio signal for vowels' a' and 'i' for the healthy and Parkinson's disease signal.

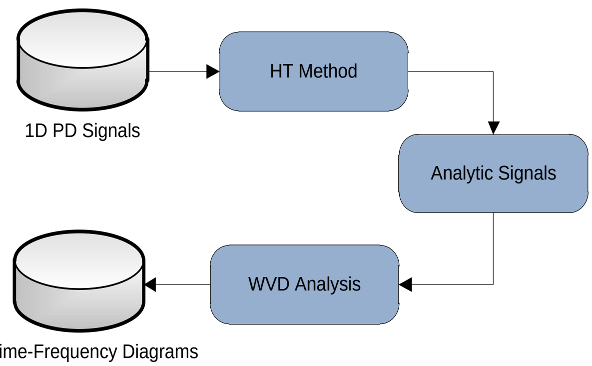

Fig. 2. The Flowchart of the Hybrid Time and Frequency Algorithm.

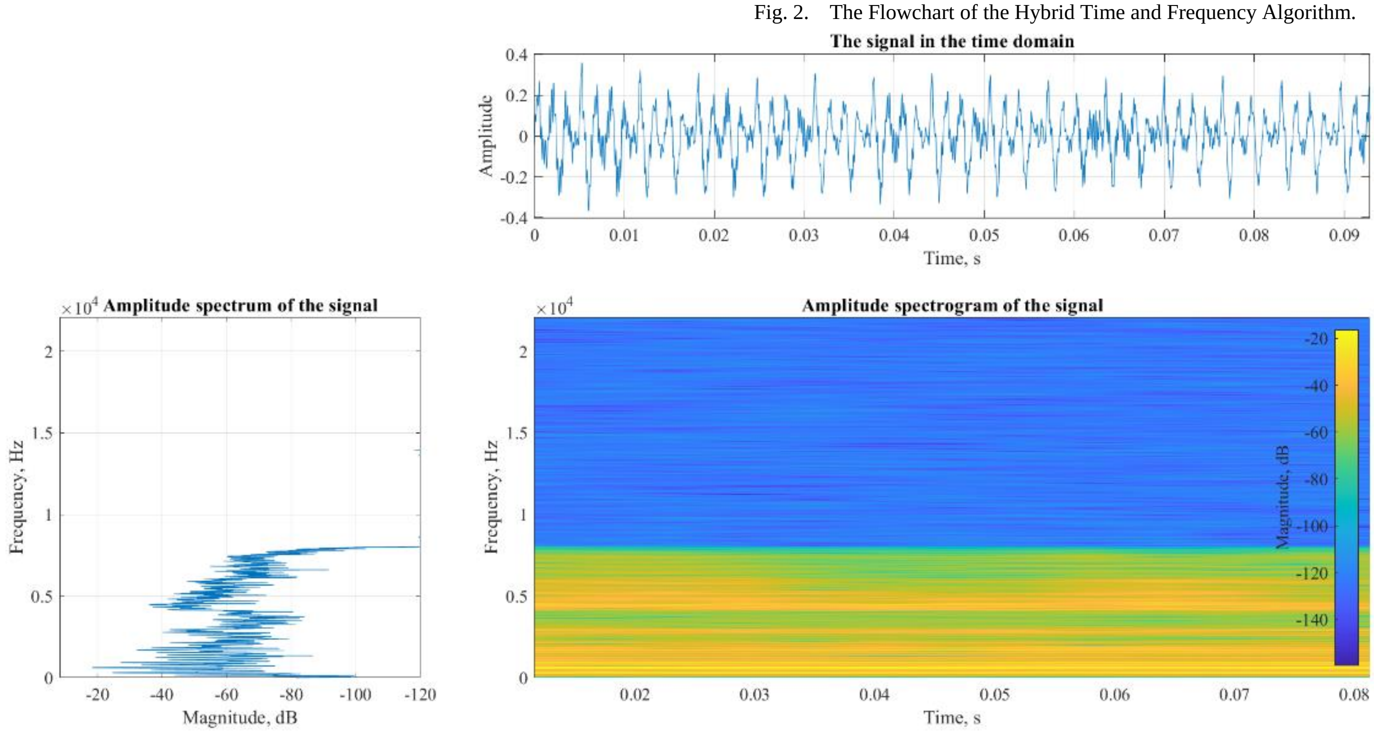

Fig. 3. 2D Time-Frequency Diagrams of PD Signals for Vowels 'A'.

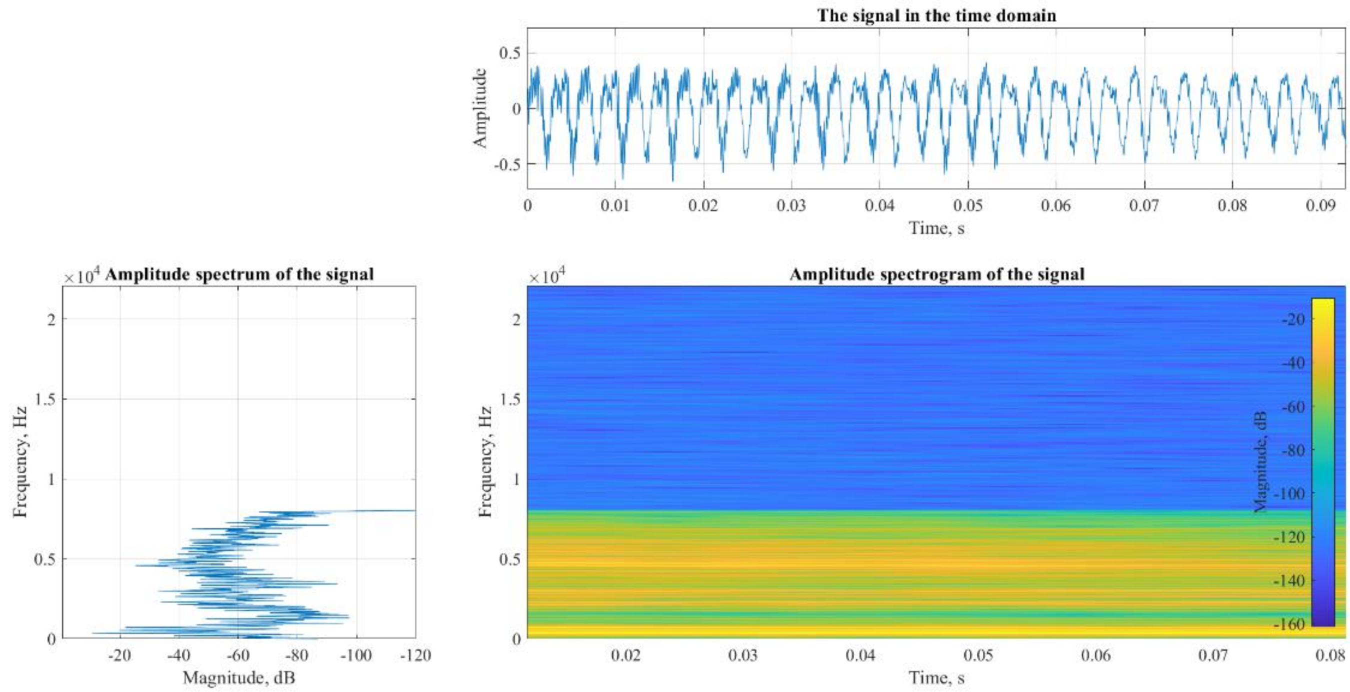

Fig. 4. 2D Time-Frequency Diagrams of PD Signals for Vowels 'I'.

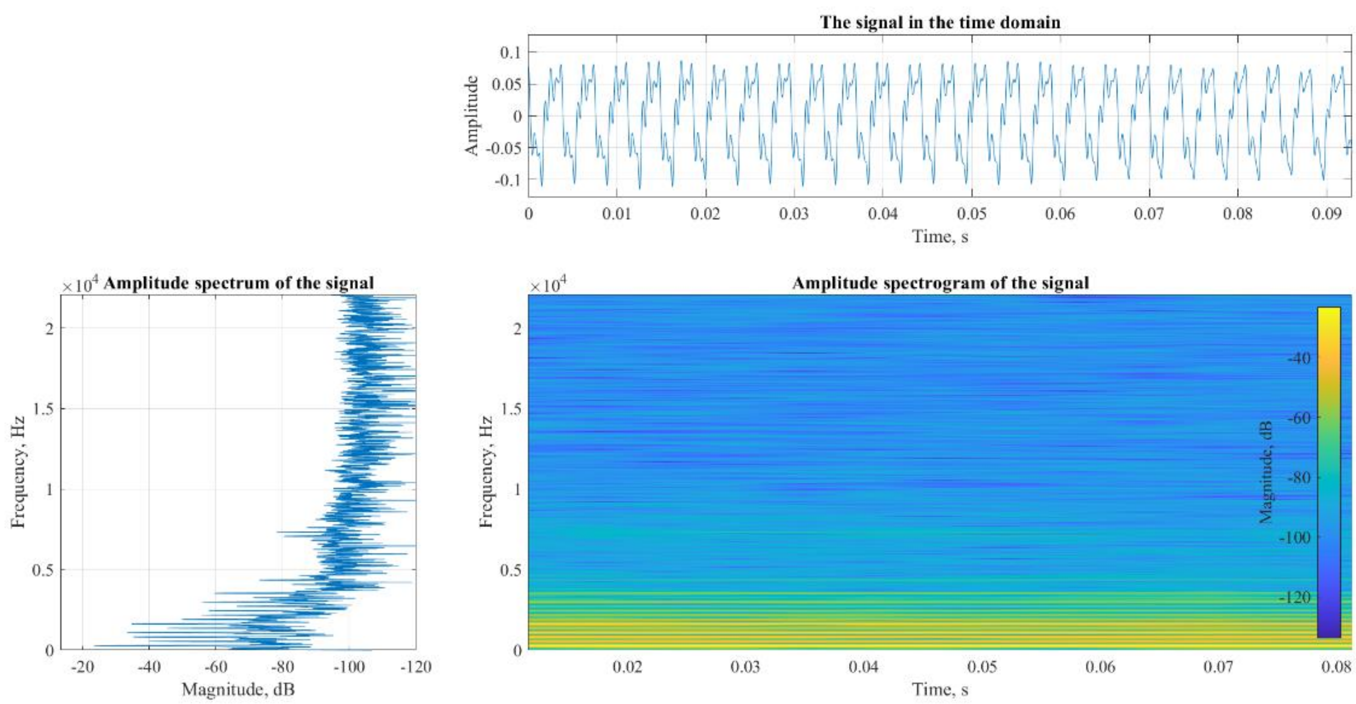

Fig. 5. 2D Time-Frequency Diagrams of HC Signals for Vowels 'A'.

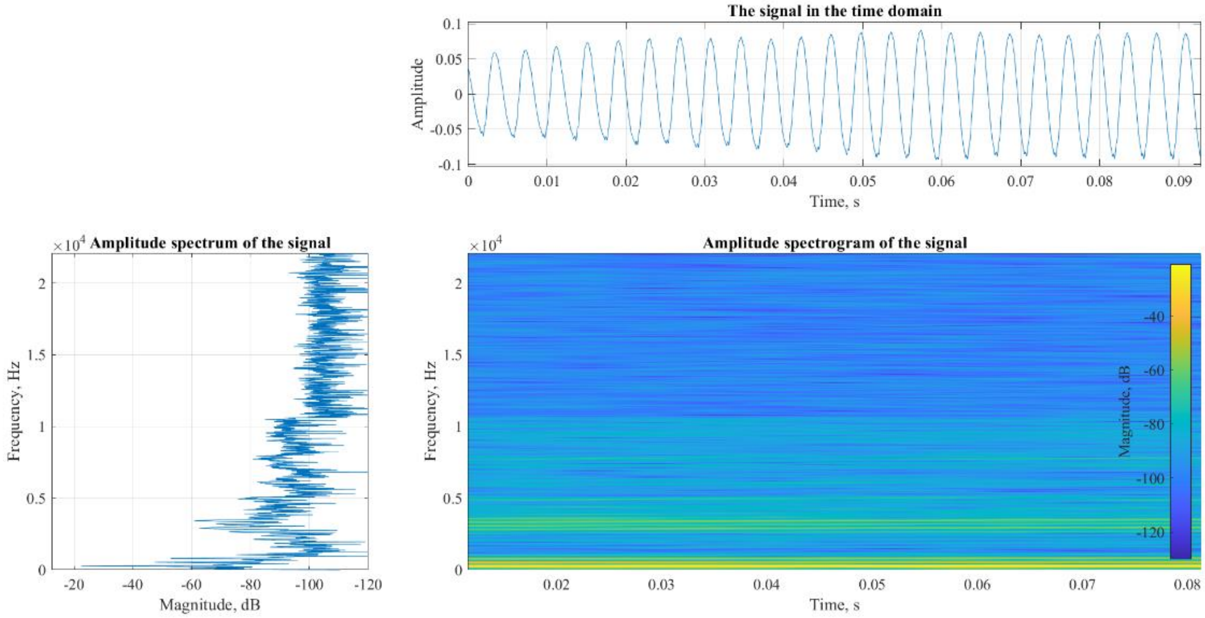

Fig. 6. 2D Time-Frequency Diagrams of HC Signals for Vowels 'I'

C. ResNet50 & Transfer Learning

He et al. [23] presented ResNet at the 2015 ILSVRC competition. ResNet includes a short connection structure that avoids gradient vanishing by immediately bypassing the input information to the output. ResNet is a network-in-network (NIN) architecture comprised of several residual modules stacked on top of one another. These residual units are employed in the construction of deep ResNet architecture. Convolution, pooling, and layering are the residual units. Instead of wholly linked layers, ResNet employs global average pooling.

The ResNet uses residual units to prevent the gradient from disappearing [25]. Let the input and output vectors for the layers be x and y, respectively, and *H(x)* function as nonlinear stacked layers. It is possible to obtain $H(x) = F(x) + x$ by using the formula $F(x) = H(x) - x$, where F(x) and x represent stacked nonlinear layers. The residual block's typical structure is described in (5).

$$y = F(x, \{Wi\}) + x \quad (5)$$

Wi is the weight in the weight matrix, and F is the residual mapping to be learned.

The ImageNet dataset was used to train the conventional pre-trained ResNet50. 2D time and frequency diagrams made from PD recordings were the images to be analyzed in this work; these diagrams differed significantly from the photographs in ImageNet. To enable the modified ResNet50 to classify two PD classes, the last three layers of the pre-trained ResNet50 - the fully connected layer, the Softmax layer, and the classification layer - were altered. We used the same dense layer network after the convolutional layers, as shown in Fig. 7. After the convolutional layer and every block, the rectified linear unit (ReLU) is activated by doing a max-pooling operation. To make the spatial dimension of the activation map half the preceding layer, a max pooling layer with a 2x2 kernel and stride size of two is utilized. The improved ResNet50 was retrained to provide new parameters, and Table I displays the changed ResNet50's structure. The learning rate of deep learning based on ResNet50 was set at 0.001 in this research.

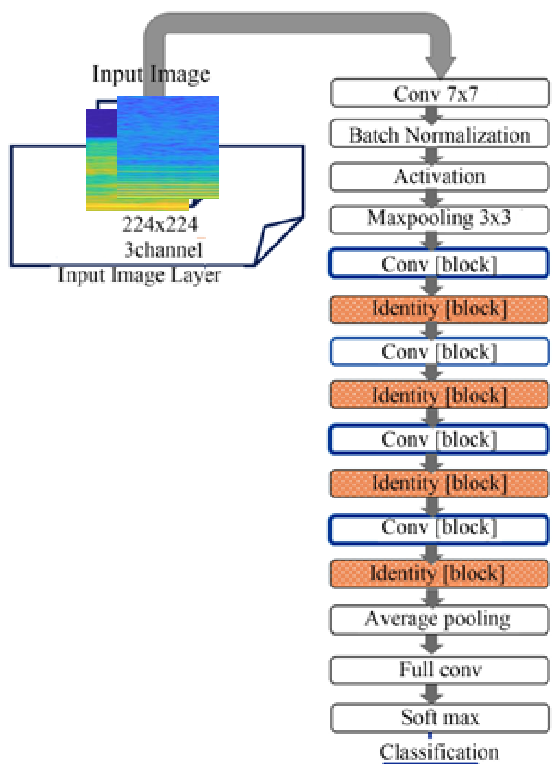

Fig. 7. ResNet50 Architecture for Proposed Model.

TABLE I. NETWORK PARAMETER CONFIGURATION OF RESNET50

Layer Name	**Output size**	**Layers**	
Conv1	224x224	7x7, 64, stride 2	
Conv2	112x112	3x3 Max pool, stride 2	
		1x1, 64	X3
		3x3, 64	
		1x1, 256	
Conv3	56x56	1x1, 128	X4
		3x3, 128	
		1x1, 512	
Conv4	28x28	1x1, 256	X6
		3x3, 256	
		1x1, 1024	
Conv5	14x14	1x1, 512	X3
		3x3, 512	
		1x1, 2048	
	1x1	Average Pool	
		2-d Fully Connected	
		SoftMax	

IV. RESULTS AND DISCUSSION

A. Dataset Description

This research gathered speech recordings from 20 people (10 Parkinson's patients and 10 healthy controls). The data for the research was acquired from 188 individuals with the disease (5 men and 5 women) aged 33 to 87 (65.110.9). The control group comprises 10 healthy people (5 men and 5 women) aged 41 to 82 (61.18). The microphone was adjusted to 44.1 KHz throughout the data collection procedure. After the physician's examination, the sustained phonation of the vowels/a/and/i/was gathered from each participant three times. Fig. 8 shows the sample 2D time-frequency diagrams used in the analysis.

B. Experimental Setup

A 64-bit Windows 10 with an Intel Core i7 (2.80 GHz, 2808 MHz, 4 Cores, 8 Logical processors), 16 GB of RAM, an NVIDIA GeForce GTX, and CUDA 9.0 is utilized. The experiments were carried out using the Matlab (R2019a) programming language.

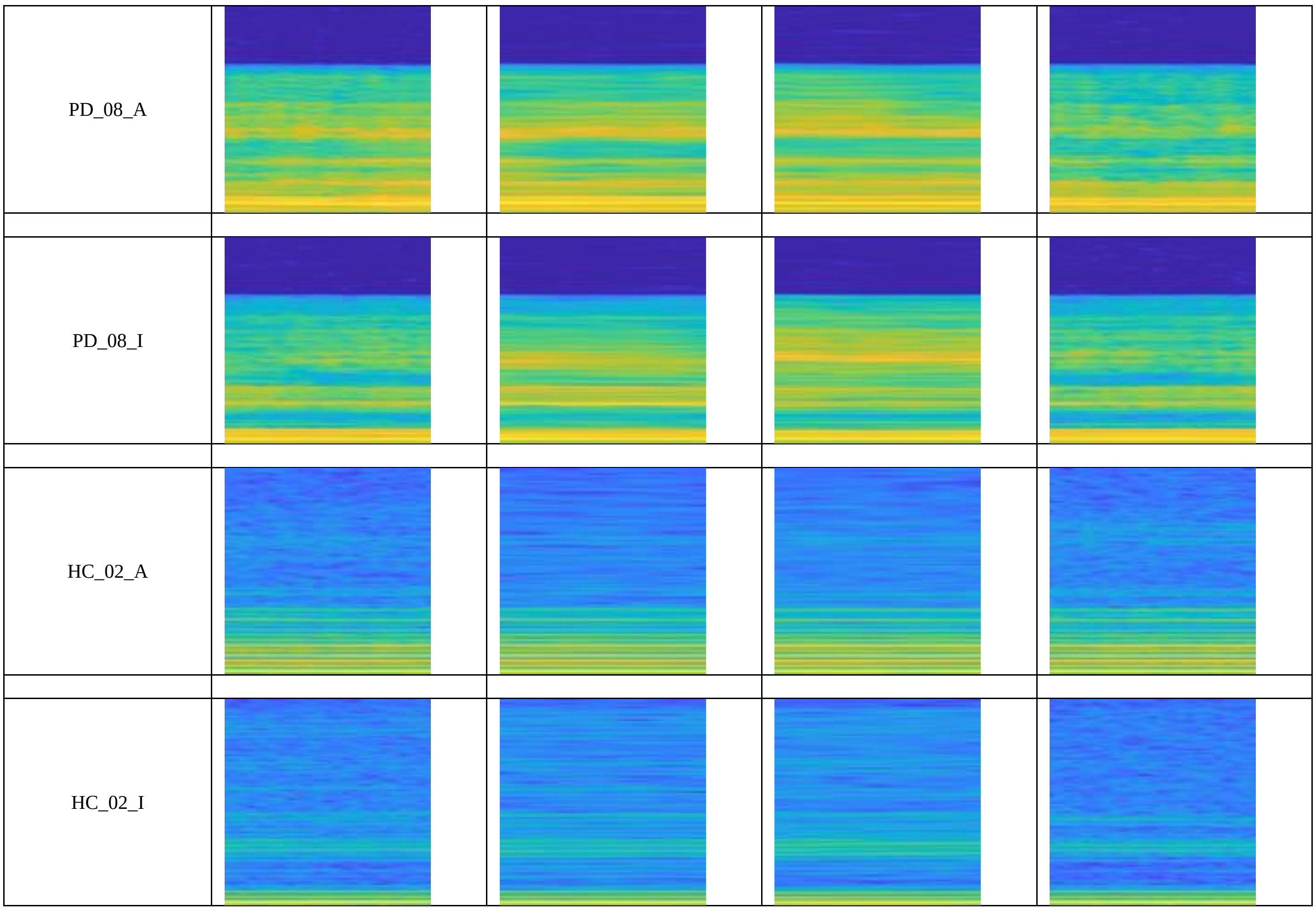

Fig. 8. Sample 2D Time-Frequency Diagrams used for Analysis.

C. Evaluation Metrics

A confusion matrix was utilized in this research to summarize the categorization findings based on the actual and predicted categories. The construction of the confusion matrix is seen in Fig. 9. As a consequence, the numbers on the diagonal display the accurate categorization results, while the other values display the wrong findings.

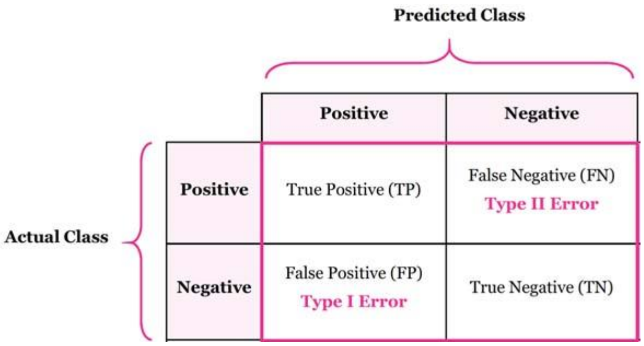

Fig. 9. Confusion Matrix.

The F1 score was also used in this research to evaluate the performance of the suggested technique. The F1 score may be viewed as a weighted average of precision and recall, with a most incredible score of 1 and the worst score of 0.

$$F1_{Score} = \frac{Tp}{Tp+\frac{1}{2}(Fp+Fn)} \tag{6}$$

In addition, four statistical indices were produced to assess the performance of the proposed classifier in this work, namely accuracy, Precision, sensitivity, and specificity were defined as follows;

$$Accuracy = \frac{Tp+Tn}{Tp+Fp+Tn+Fn} \tag{7}$$

$$Precision = \frac{Tp}{Tp+ Fp} \tag{8}$$

$$Sensitivity = \frac{Tp}{Tp+Fn} \tag{9}$$

$$Specificity = \frac{Tn}{Tn+ Fp} \tag{10}$$

Where the (Tp) is the number of PD signals correctly classified as the PD category, the (Fn) is the number of signals of the PD category incorrectly classified as the HC category. The (Tn) is the number of signals of the HC categories not classified as the PD category. The (Fp) is the number of signals incorrectly classified as the PD category in the HC categories.

D. Classification Results

Table II displays the classification results for the test data using the proposed classifier trained at a learning rate of 0.001. Table II demonstrates that the classifier can get the best results: accuracy 86.86%, precision 89.19%, sensitivity 89.17%, and specificity 84.67%. Moreover, the F1-Score is 89.18%.

Fig. 10 shows the system-generated confusion matrix. Moreover, Fig. 11 shows the receiver operating curve of the system. ROC is the plot of TPR vs. FPR. The Area under the Curve for the PD and HC categories is 93.7%, which is comparatively reasonable.

Table III shows the state-of-the-art results compared with the proposed classification results. From the above comparison, we proved that the proposed method is robust in PD recognition using vowels.

TABLE II. CLASSIFICATION RESULTS

Metrics	Results
Accuracy	86.86
Precision	89.19
Sensitivity	89.17
Specificity	84.67
F1 Score	89.18

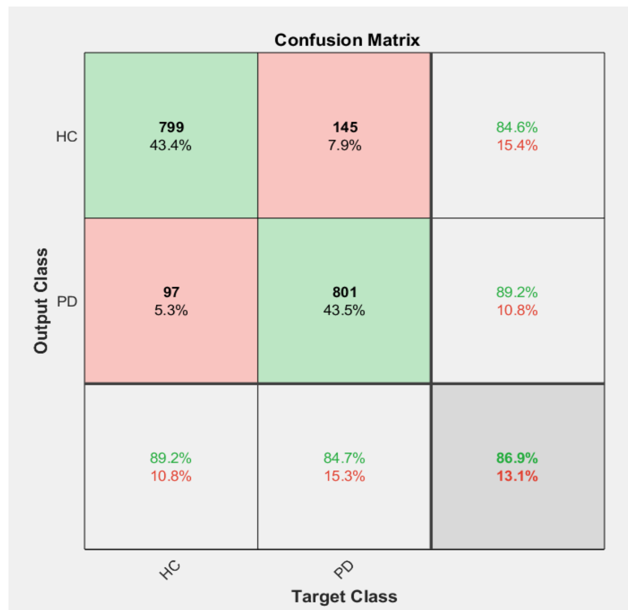

Fig. 10. Confusion Matrix shows 86.9% of Accuracy.

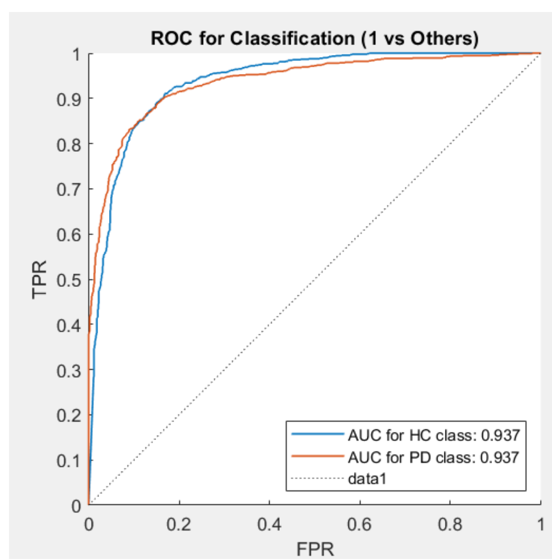

Fig. 11. ROC for PD and HC Categories.

TABLE III. COMPARISON RESULTS WITH PREVIOUS METHODS

Previous Methods	**Accuracy (%)**
Existing Method (Ref 12)	79.00
Existing Method (Ref 13)	85.00
Existing Method (Ref 14)	82.00
Existing Method (Ref 18)	81.53
Proposed Method	88.86

V. CONCLUSIONS AND FUTURE WORK

The HT-WVD and the deep learning based on ResNet50 are both modules that make up the recommended approach in this study. The HT-WVD removes cross terms to depict the relationship between time, frequency, and energy while keeping time-domain qualities. It directs energy towards the peak (i.e., amplitude and phase within the PD records). In addition, the depth of deep learning based on ResNet50 is adequate to leverage deep neural network performance more extensively than in past studies. Conversely, the suggested technique is more advanced than the existing 1D CNN, and the 2D ResNet model is computationally expensive.

This study proved the feasibility of a speech-based PD classification while proposing new potential approaches for vocal feature analysis. The highest, 86.86% accuracy, was achieved using 5-fold cross-validation. In a clinic, new patients with Parkinson's disease may be identified from healthy persons using the blind test interface. The suggested technique performs well in categorizing PD voice signals compared to earlier research.

REFERENCES

[1] Samii A, Nutt JG, Ransom BR. Parkinson’s disease. Lancet. 2004;363:1783‑93. https:// doi. org/ 10. 1016/ S0140-6736(04) 16305-8.

[2] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni,A. Lobo, J. Martinez-Lage,C.Trenkwalder, and A. Hofman, “Prevalence of Parkinson’s disease in Europe: A collaborative study of population-based cohorts,” Neurology, vol. 54, pp. 21–23, 2000.

[3] Massano J, Bhatia KP. Clinical approach to Parkinson’s disease: features, diagnosis, and management principles. Cold Spring Harbor Perspect Med. 2012;2(6):8870. https:// doi. Org/ 10. 1101/ cshpe rspect. a0088 70.

[4] B. E. Sakar, M. Isenkul, C. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828-834, 2013.

[5] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinson’s Disease Using Measurements of Dysphonia," Journal of Medical Systems, vol. 34, no. 4, pp. 591-599, 2009.

[6] A. Tsanas, M. Little, P. McSharry and L. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," Journal of the Royal Society Interface, vol. 8, no. 59, pp. 842-855, 2010.

[7] Gray P, Hildebrand K. Fall risk factors in Parkinson’s disease. J Neurosci Nurs. 2000;32:222. https:// doi. org/ 10. 1097/ 01376 517-20000 8000-00006.

[8] Gunduz H. Deep learning-based Parkinson’s disease classification using vocal feature sets. IEEE Access. 2019;7:115540‑51. https:// doi. org/ 10. 1109/ access. 2019. 29365 64.

[9] M. A. Little, P. Mcsharry, E. Hunter, J. Spielman, and L. Ramig, “Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease,” IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015–1022, 2009.

[10] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease. IEEE Trans Biomed Eng. 2009;56(4):1015‑22. https:// doi. org/ 10. 1109/ TBME. 2008. 20059 54.

[11] Hlavnika J, Cmejla R, Tykalova T, Šonka K, Ruzicka E, Rusz J. Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behavior disorder. Sci Rep. 2017;7(1):1‑13. https:// doi. org/ 10. 1038/ s41598-017-00047-5.

[12] Orozco-Arroyave JR, Honig F, Arias-Londono JD, Vargas-Bonilla JF, Noth E. Spectral and cepstral analyses for Parkinson’s disease detection in Spanish vowels and words. Expert Syst. 2015;32(6):688‑97. https:// doi. org/ 10. 1111/ easy. 12106.

[13] Orozco-Arroyave JR, Honig F, Arias-Londono JD, Vargas-Bonilla JF, Daqrouq K, Skodda S, Rusz J, Noth E. Automatic detection of Parkinson’s disease in running speech spoken in three different languages. J Acoust Soc Am. 2016;138:481‑500. https:// doi. org/ 10. 1121/1. 49397 39.

[14] Karan B, Sahu SS, Orozco-Arroyave JR, Mahto K. Hilbert spectrum analysis for automatic detection and evaluation of Parkinson’s speech. Biomed Signal Processing Control. 2020;61:102050. https:// doi. org/ 10. 1016/j. bspc. 2020. 102018.

[15] Little, M.A.; McSharry, P.E.; Hunter, E.J.; Ramig, L.O. Suitability of dysphonia measurements for telemonitoring Parkinson’s disease. IEEE Trans. Biomed. Eng. 2009, 56, 1015–1022. [CrossRef] [PubMed].

[16] Bhattacharya, I.; Bhatia, M.P.S. SVM Classification to Distinguish Parkinson Disease Patients. In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing, Coimbatore, India, 16–17 September 2010; pp. 1–6.

[17] Das, R. A Comparison of Multiple Classification Methods for Diagnosis of Parkinson's Disease. Expert Syst. Appl. 2010, 37, 1568–1572. [CrossRef].

[18] Sakar, C.O.; Kursun, O. Telediagnosis of Parkinson’s Disease Using Measurements of Dysphonia. J. Med. Syst. 2010, 34, 591–599. [CrossRef] [PubMed].

[19] Polat, K. Classification of Parkinson’s Disease Using Feature Weighting Method based on Fuzzy C-Means Clustering. Int. J. Syst. Sci. 2011, 43, 597–609. [CrossRef].

[20] E. Ataman, V. Aatre, and K. Wong, “A fast method for real-time median filtering,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 415-421, 1980.

[21] S. S. Qurraie, and R. G. Afkhami, “ECG arrhythmia classification using time-frequency distribution techniques,” Biomed. Eng. Let., vol. 7, no. 4, pp. 325-332, 2017.

[22] A. Ramos, A. Lazaro, D. Girbau, R. Villarino, “Chipless Time-coded UWB RFID: Reader, Signal Processing and Tag Design,” in RFID and Wireless Sensors Using Ultra-Wideband Technology, London, UK: ISTE Press-Elsevier, 2016, ch. 2, sec. 2.4, pp. 19-73.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc: Int. Conf. Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[24] Yatao Zhang, Junyan Li, Shoushui Wei, Fengyu Zhou, Dong Li. "Heartbeats Classification Using Hybrid Time-Frequency Analysis and Transfer Learning Based on ResNet," IEEE Journal of Biomedical and Health Informatics, 2021.

[25] Md. Rashed-Al-Mahfuz, Mohammad Ali Moni, Shahadat Uddin, Salem A. Alyami, Matthew A.Summers, Valsamma Eapen. "A Deep Convolutional Neural Network Method to Detect Seizures and Characteristic Frequencies Using Epileptic Electroencephalogram (EEG) Data," IEEE Journal of Translational Engineering in Health and Medicine, 2021.