Cognitive Computation (2024) 16:1198–1209
https://doi.org/10.1007/s12559-024-10254-8

Investigation of Scalograms with a Deep Feature Fusion Approach for Detection of Parkinson's Disease

İsmail Cantürk[1,2] · Osman Günay[1]

Received: 23 May 2023 / Accepted: 9 January 2024 / Published online: 2 February 2024

Abstract

Parkinson's disease (PD) is a neurological condition that millions of people worldwide suffer from. Early symptoms include a slight sense of weakness and a propensity for involuntary tremulous motion in body limbs, particularly in the arms, hands, and head. PD is diagnosed based on motor symptoms. Additionally, scholars have proposed various remote monitoring tests that offer benefits such as early diagnosis, ease of application, and cost-effectiveness. PD patients often exhibit voice disorders. Speech signals of the patients can be used for early diagnosis of the disease. This study proposed an artificial intelligence–based approach for PD diagnosis using speech signals. Scalogram images, generated through the Continuous Wavelet Transform of the speech signals, were employed in deep learning techniques to detect PD. The scalograms were tested with various deep learning techniques. In the first part of the experiment, AlexNet, GoogleNet, ResNet50, and a majority voting-based hybrid system were used as classifiers. Secondly, a deep feature fusion method based on DenseNet and NasNet was investigated. Several evaluation metrics were employed to assess the performance. The deep feature fusion system achieved an accuracy of 0.95 and an F1 score with stratified 10-fold cross-validation, improving accuracy by 38% over the ablation study. The key contributions of this study include the investigation of scalogram images with a comprehensive analysis of deep learning models and deep feature fusion for PD detection.

Keywords Parkinson's disease · Speech signals · Time-frequency plot · Decision support systems · Machine learning systems

Introduction

Parkinson's disease (PD) is a prevalent neurodegenerative condition [1]. In addition, a rapid increase in the incidence and prevalence of the disease has been observed in the last 20 years [2]. Diagnosis of PD primarily relies on the presence of motor symptoms such as bradykinesia, rest tremor, and rigidity. However, there may be a long latency between the onset of the disease and the clinical diagnosis due to the slow progression of the disorder [3]. PD patients may also suffer from non-motor symptoms like depression and anxiety [3]. The etiology of PD encompasses various factors, including aging, genetics, and environmental influences [4]. Although symptomatic treatments exist, the progression rate of PD cannot be reversed or decelerated [3].

Apart from the clinical diagnosis, extensive research has been conducted by scholars to explore different remote monitoring tests for early-stage PD detection [5–9]. These innovative tests offer several advantages, including early diagnosis, ease of application, cost-effectiveness, and remote monitoring.

PD patients commonly exhibit speech impairments attributed to the involuntary motion of the jaws and lips, as well as impaired mobility and stability of articulatory organs [10]. These impairments result in vocalization abnormalities. Dysphonia and dysarthria, which are related to problems in voice and articulation, contribute to speech impairments in PD. A study conducted by Roberts and Post [11] demonstrated that PD patients have reduced discourse informativeness compared to healthy subjects. Therefore, early-stage PD detection from speech signals has become a popular research topic [12].

✉ İsmail Cantürk
icanturk@yildiz.edu.tr

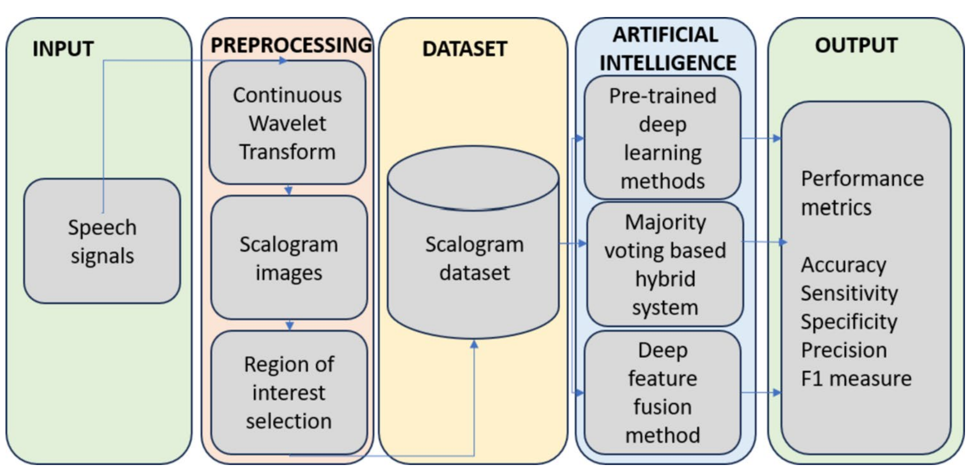
1 Department of Biomedical Engineering, Yildiz Technical University, Istanbul, Turkey

2 Department of Biomedical Engineering, Faculty of Electrical and Electronics Engineering, Medical Intelligence Innovation Center, Yildiz Technical University, 34220 Esenler, Istanbul, Turkey

Various artificial intelligence–based techniques have been investigated to assess the speech signals of PD patients. In a study by Sakar et al. [13], multiple types of voice signals from PD patients were collected and analyzed using machine learning methods. Cantürk and Karabiber [14] achieved a higher accuracy on the same dataset by incorporating feature selection methods. Parisi et al. [15] developed a hybrid machine learning system for feature reduction and high-accuracy PD classification from speech. Moro-Velazquez et al. [16] employed a speaker recognition system for PD detection. Additionally, Arias-Vergara et al. [17] collected a dataset comprising clinical voice recordings and mobile phone recordings. They analyzed the progression of PD using speech signals. In a study by Grover et al. [18], voice signals were utilized for binary classification of PD severity using a deep learning approach. Aggarwal et al. [19] analyzed the incorporation of artificial intelligence into neuroimaging techniques for the assessment of PD. Anter et al. [20] proposed an artificial intelligence–based system to monitor progress of PD from speech signals. The system selected effective voice features and predicted the severity. Quan et al. [21] detected PD from speech signals by extracting and classifying time series features with the aid of a deep learning method. Ngo et al. [22] reviewed studies of sound-based computational PD detection and concluded that in addition to the importance of studies to assess PD, a physiological understanding of the methods used is crucial for clinicians. Pauline et al. [23] proposed a denoising system to clear speech signals of PD patients. Celik and Başaran [24] extracted deep features with SkipConNet from sound signals and classified them with random forest algorithm for PD detection. Khaskhoussy and Ayed [25] also employed deep features for PD classification from sound signals. Warule et al. [26] extracted time-frequency features by employing chirplet transform, and they detected PD with high accuracy. Campi et al. [27] investigated statistical time series methods with machine learning methods for PD diagnosis. Reddy and Alku [28] detected PD from voice signal with an exemplar-based sparse representation classification method. Wen et al. [29] proposed a personalized speech-based PD detection system that included the use of different patient-based classifiers. Guatelli et al. [30] utilized speech signal–based spectrograms to detect PD with extreme learning machines. Ali et al. [31] proposed a two-stage deep neural network–based scheme to classify PD. Pramanik et al. [32] proposed a decision tree ensemble method–based system for sound-based PD detection. Hemmerling et al. [33] investigated the regression method on speech features to predict the severity of PD. Li et al. [34] studied an ensemble algorithm system that employed multi-type transformation for speech-based detection of PD. Laganas et al. [35] studied the detection of PD from running speech in telephone conversations with the aid of machine learning methods. Rana et al. [36] compared the performances of various classifiers by utilizing voice features for PD detection. Alalayah et al. [37] applied recursive feature elimination method to reduce the dimension of PD voice features and classified them with machine learning methods.

In this study, we proposed an artificial intelligence–based system that detects PD from speech signal based scalograms. Speech signals were transformed into scalograms, and they were fed into deep learning models. Firstly, AlexNet, GoogleNet, ResNet50, and a majority voting-based hybrid system were used to classify scalograms. Then, deep feature fusion system was employed with current benchmark methods: DenseNet and NasNet. The performances were assessed with several metrics. The key contributions of the paper are in the investigation of scalogram images for PD detection with a comprehensive analysis of deep learning models and deep feature fusion method.

Fig. 1 The proposed system's block diagram



Materials and Methods

The Material

The dataset utilized in this study was obtained by Sakar et al. [13] with the ethical approval of the Istanbul University Research Ethics Committee. The recordings were captured using a Trust MC-1500 microphone, which has a frequency response ranging from 50 Hz to 13 kHz and a sampling rate of 96 kHz. The dataset consists of various types of phonations performed by both PD patients and individuals in the control group. These include sustained vowels, numbers, short sentences, and words. The dataset comprises a total of 20 individuals diagnosed with PD and an equal number of 20 healthy individuals in the control group.

The Methods

The proposed system's block diagram is presented in Fig. 1. Speech signals were preprocessed, and scalogram dataset was generated. The dataset was fed into artificial intelligence methods. Performance of the system was evaluated with different metrics. All computations were conducted using MATLAB.

Preprocessing

Based on the superior performance of sustained phonations over running speech in PD detection [38], this study focused on sustained phonations. Three types of sustained vowels, "a," "o," and "u," were incorporated into the study. The sustained phonations were extracted from the speech signals with 20 k samples for each

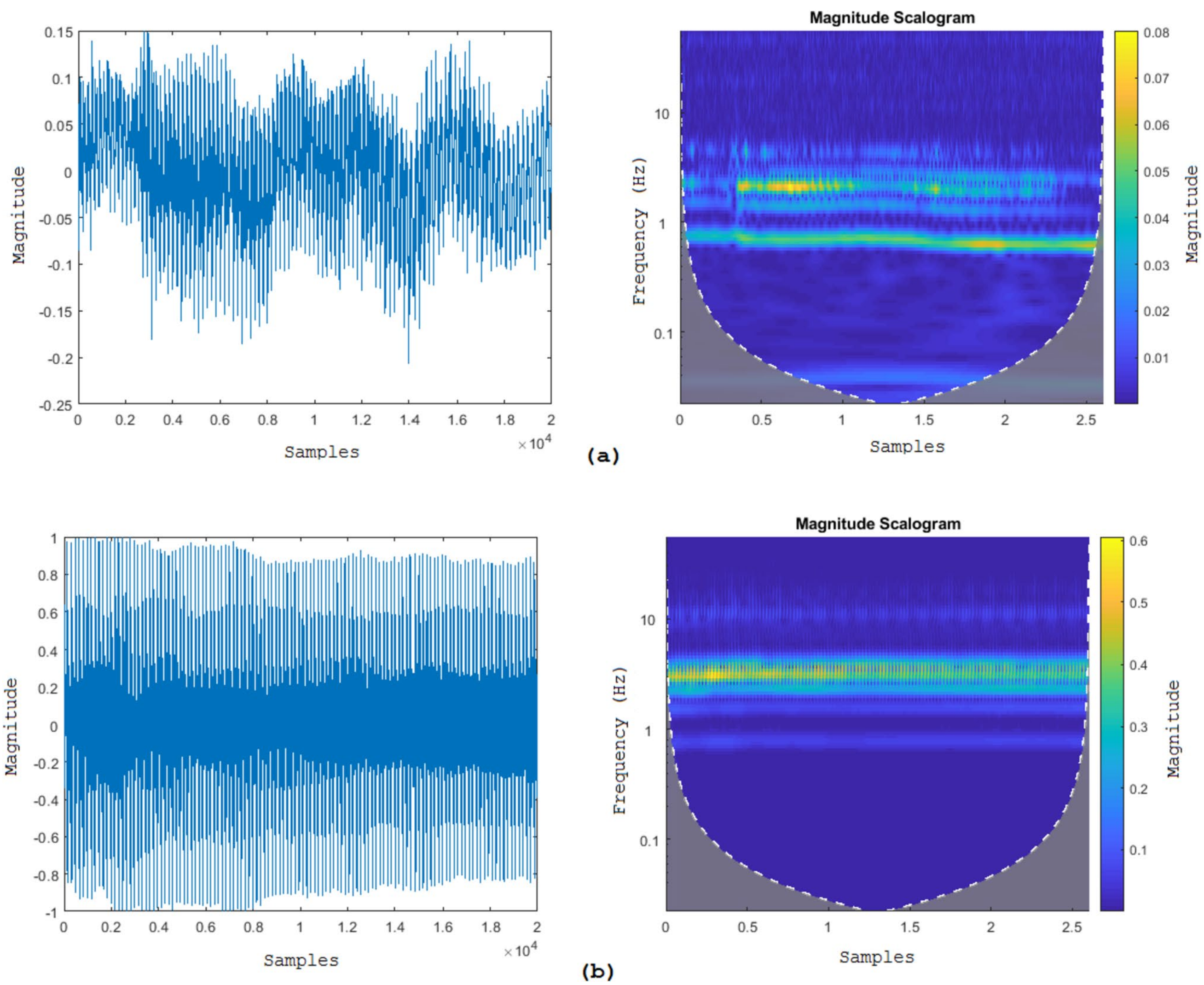


Fig. 2 Sustained "a" phonations of **a** a PD patient and **b** a healthy subject

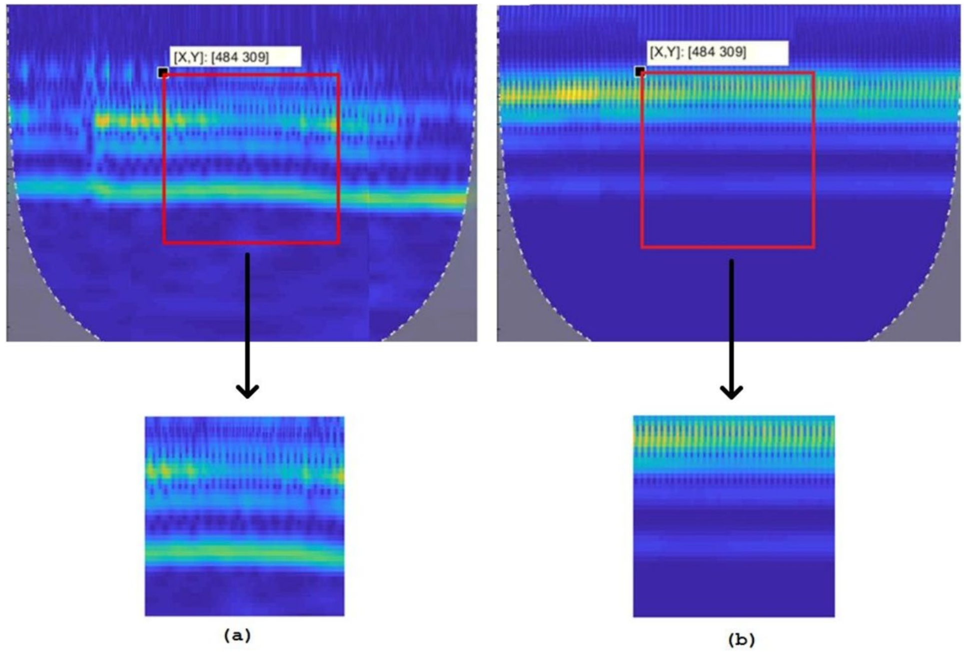


Fig. 3 RoI selection for **a** a PD patient and for **b** a healthy subject

vowel type. The samples were acquired around the midpoints of the recordings to ensure representative data.

To generate the scalograms, the Continuous Wavelet Transform (CWT) was applied to the speech signals. The CWT function in MATLAB was used for this purpose. An analytic Morse wavelet with a gamma parameter of 3 and a time-bandwidth product parameter of 60 was employed for CWT implementation. Alternatively, the short-time Fourier transform (STFT) method (Eq. 1) can also be utilized to obtain a time-frequency plot.

$$X_F(\omega, \tau) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}w(t - \tau)dt \tag{1}$$

where $X_F(\omega, \tau)$ is the STFT of the continuous time signal $x(t)$ and $\omega \, \epsilon \{-\infty, \ \infty\}$. The frequency information of the signal is computed with Gaussian-like windows $w(t - \tau)$. The windowing function is iterated by τ in every step over the signal. The major drawback of STFT is that it uses the same windowing function for all frequencies, which limits its ability to provide a multiresolution view of the signal in the time-frequency plot. In contrast, the wavelet transform (Eq. 2) overcomes this limitation by allowing for variable windowing sizes. By stretching and shrinking the basis functions (Eq. 3), the wavelet transform can provide a multiresolution view of the signal, which allows a more comprehensive representation of the signal's time-frequency characteristics.

$$X_W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h^* \left(\frac{t - b}{a}\right) x(t)dt \tag{2}$$

where

$$h_{a,b}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t - b}{a}\right) \tag{3}$$

$X_W(a, b)$ is CWT of $x(t)$ and * denotes conjugate of $h_{a,b}(t)$. By altering the $a \in R^+$ and $b \in R$, basis function shrinks or stretches [39].

Figure 2 demonstrates the speech signals and their corresponding CWT outputs, calculated with a sampling frequency of 128 Hz, for a PD patient and a control subject, respectively. It can be observed that the time-frequency images of PD patients exhibit more disordered patterns compared to those of control subjects, reflecting the progressive distortion of speech in PD. However, Fig. 2 represents one of the most discriminative cases. When the ages of PD patients and healthy subjects are similar and the severity of PD is mild, visual discrimination becomes challenging. Therefore, the use of computer-assisted techniques becomes essential for accurate detection PD patients.

Regions of interest (RoIs) were extracted from the time-frequency plots to eliminate redundant areas (Fig. 3). The upper left corner coordinates [484, 309] were identified

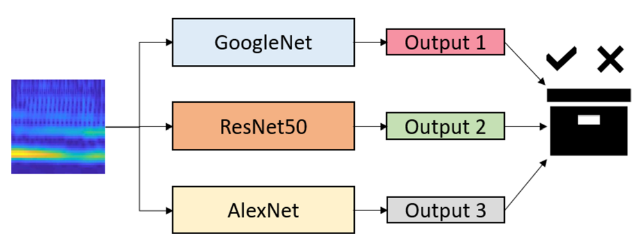


Fig. 4 The majority voting-based hybrid system

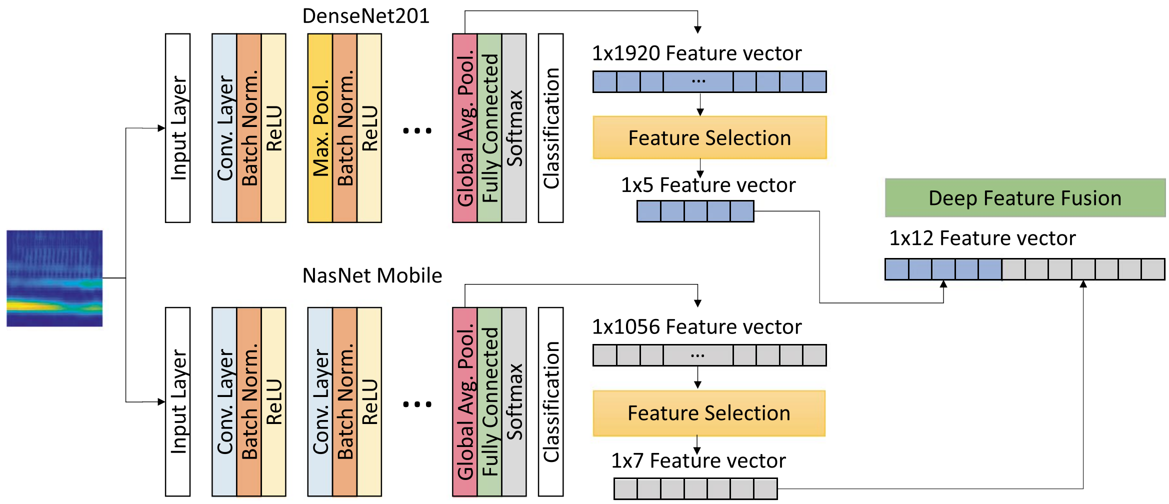


Fig. 5 The deep feature fusion method

as reference points for all images, and square regions of 250*250 pixels were cropped accordingly. The dataset contained a total of 120 images consisting of each subject's three types of scalograms.

Applied Artificial Intelligence Methods

Use of Pre-trained Deep Learning Methods as Classifiers Pre-trained deep learning models, which are GoogleNet [40], AlexNet [41], and ResNet50 [42], were utilized as classifiers. Essential layers of the deep learning models are convolution layer, rectified linear units (ReLU), and pooling layer. The convolution layer refers to the process that input image "I" is convolved with a predetermined function, known as the kernel "K" (Eq. 4).

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \quad (4)$$

where K is an $[mxn]$ matrix and $S(i,j)$ is the value of the matrix at (i_{th}, j_{th}) points.

ReLU is an activation function (Eq. 5). There are also different activation functions.

$$g(z) = \max\{0, z\} \quad (5)$$

where $z \in Z$.

The pooling layer generates a reduced-dimensional feature map by combining similar features into a single feature. This combination can be achieved using various methods such as max pooling and average pooling [43].

GoogleNet and ResNet50 are deep neural network architectures consisting of 144 and 177 layers, respectively. Due to the relatively small size of the dataset used in this study, the weights of the initial 110 layers were kept constant for all experiments, while the remaining layers were updated during the training process. This weight freezing technique was not applied to AlexNet, which has 25 layers. The utilized MATLAB update functions were adam, sgdm, and RMSProp. The minimum batch and the maximum number of epochs were chosen as 7 and 12, respectively. The learning rate was 10^{-4}.

Furthermore, a hybrid machine learning system (Fig. 4) was developed to further enhance the detection accuracy.

Table 1 The confusion matrix and performance metrics

Conf. mat.		**Predicted class**		**Performance metrics**	
		1	**2**		
True class	**1**	**True positive (TP)**	**False negative (FN)**	**Accuracy**	(TP + TN)/(TP + FN + TN + FP)
				Sensitivity	TP/(TP + FN)
				Specificity	TN/(TN + FP)
	2	**False positive (FP)**	**True negative (TN)**	**Precision**	TP/(TP + FP)
				F1 measure	$2 * \frac{\text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$

1, PD patient; 2, healthy subject

Table 2 The average results of the deep learning models for "a" phonation

Model	Update function	Accuracy	Sensitivity	Specificity	Precision	F1 measure
GoogleNet	**Sgdm**	0.81	0.83	0.80	0.82	0.81
	Adam	0.75	0.83	0.68	0.72	0.76
	Rmsprop	0.66	0.83	0.50	0.66	0.71
ResNet50	**Sgdm**	0.64	0.68	0.60	0.64	0.64
	Adam	0.73	0.73	0.73	0.78	0.77
	Rmsprop	0.68	0.55	0.80	0.76	0.61
AlexNet	**Sgdm**	0.70	0.73	0.68	0.73	0.71
	Adam	0.61	0.53	0.70	0.73	0.55
	Rmsprop	0.60	0.45	0.75	0.69	0.50

This system operated on the principle of majority voting. The same scalogram was tested on the deep learning systems separately, and the outputs were counted. The majority answer was selected as the final output (Eq. 6).

$$P = \begin{cases} 1, if\ Count(1, A) + Count(1, B) + Count(1, C) >= 2 \\ 2, otherwise \end{cases} \tag{6}$$

P is the final prediction. Count$(1, X)$ represents the count of positive predictions for X algorithm, where X can be A, B, or C. 1 and 2 refer to PD patient and healhty subject, respectively.

To assess the generalization performance of the systems, 10-fold cross-validation technique was utilized for all experiments. By averaging the evaluation metrics obtained from these iterations, the models' performances were obtained. The training and test sample rates were applied as 75% and 25%, respectively. All samples were randomly distributed in each fold.

Deep Feature Fusion Method Deep feature fusion is a popular approach used to better represent the feature space. DenseNet [44] and NasNet [45] were used as deep feature extractors (Fig. 5). We utilized DenseNet201 and NasNet mobile architectures, which are current benchmark methods. The last global average pooling layers of both algorithms, namely, avg_pool and global_average_pooling2d_1 for DenseNet201 and NasNet mobile, respectively, were investigated as feature extraction layers. DenseNet and NasNet produced 1920 and 1056 features, respectively. The feature dimensions were reduced with Lasso [46]. Five out of 1920 features from DenseNet and seven out of 1056 features from NasNet were selected. The resulting feature vectors were concatenated, and 1×12 length feature vector was obtained for each scalogram.

Performance Metrics

The experimental results were evaluated with five metrics: accuracy, sensitivity, specificity, precision, and F1 measure. The calculation of the metrics according to the confusion matrix is given in Table 1.

Experimental Results

Results of Using Pre-trained Methods as Classifiers

Table 2 represents the average results of the deep learning models for the sustained "a" phonation. GoogleNet with the sgdm function showed the best performance, scoring above 80% for all performance metrics. The best and worst accuracies of this experiment out of 10-fold were 1.00 and 0.63, respectively. The most successful results are highlighted in the table.

In Table 3, the average scores of the deep learning models for the sustained "o" phonation were presented. Both GoogleNet and ResNet50 showed similar performances with an F1 measure of 0.66. However, GoogleNet was more successful in accurately identifying PD patients. It achieved a higher

Table 3 The average results of the deep learning models for "o" phonation

Model	Update function	Accuracy	Sensitivity	Specificity	Precision	F1 measure
GoogleNet	**Sgdm**	0.64	0.73	0.55	0.68	0.66
	Adam	0.59	0.60	0.58	0.58	0.58
	Rmsprop	0.49	0.50	0.48	0.48	0.48
ResNet50	**Sgdm**	0.68	0.65	0.70	0.72	0.66
	Adam	0.56	0.50	0.63	0.59	0.53
	Rmsprop	0.56	0.58	0.55	0.58	0.55
AlexNet	**Sgdm**	0.59	0.73	0.45	0.61	0.63
	Adam	0.59	0.60	0.58	0.63	0.59
	Rmsprop	0.49	0.45	0.53	0.51	0.47

Table 4 The average results of the deep learning models for "u" phonation

Model	Update function	Accuracy	Sensitivity	Specificity	Precision	F1 measure
GoogleNet	**Sgdm**	0.66	0.73	0.60	0.64	0.67
	Adam	0.58	0.68	0.48	0.57	0.61
	Rmsprop	0.55	0.68	0.43	0.54	0.58
ResNet50	**Sgdm**	0.54	0.50	0.58	0.56	0.52
	Adam	0.48	0.50	0.45	0.48	0.48
	Rmsprop	0.43	0.60	0.25	0.43	0.49
AlexNet	**Sgdm**	0.51	0.60	0.43	0.54	0.55
	Adam	0.59	0.50	0.68	0.66	0.52
	Rmsprop	0.58	0.60	0.55	0.58	0.58

sensitivity score of 0.73. On the other hand, ResNet50 had a higher specificity score of 0.70, indicating its better performance in correctly classifying healthy subjects. The best accuracy was obtained as 0.68 in ResNet50 experiment.

The performances of the deep learning models for sustained "u" phonation are presented in Table 4. Overall, GoogleNet achieved the highest accuracy score as 0.66 among the models. Notably, the scores obtained for the "o" and "u" phonations were close to each other. The best F1 scores were 0.66 and 0.67 for "o" and "u" phonations, respectively.

In the next experiment, the developed hybrid system was tested. Since obtained scores were better for "a" phonation, it was used as the input to the system. The most successful update functions for each of the deep learning methods were preferred: sgdm for GoogleNet and AlexNet and adam for ResNet50. Table 5 presents the scores of the hybrid system. The system achieved an average accuracy of 0.86 and an average F1 measure of 0.84. Additionally, average specificity and precision metrics were both over 0.90, indicating high performance in correctly identifying healthy subjects. The best and the worst accuracies of this experiment out of 10-fold were 1.00 and 0.75, respectively.

Results of Deep Feature Fusion Method

The deep feature fusion experiment was also conducted on "a" phonation. Extracted deep features from the last global average pooling layers of DenseNet and NasNet were applied to the feature reduction algorithm separately. The algorithm selected the features at index 42 out of 100 feature subsets for both (Fig. 6). The selected features were fused, and they were tested on K nearest neighbor (KNN), support vector machines (SVM), and random forest (RF) classifiers. The results were validated with stratified 10-fold cross validation. We have obtained state-of-the-art performance metrics in the KNN experiment (Table 6). All the scores were obtained as 0.95. The confusion matrix of the experiment is given in Fig. 7. Only one subject was misclassified from each group. The hyperparameter was 11 for k, and we utilized Euclidean as the distance parameter (Fig. 8). Second best scores were obtained in the SVM experiment. Two subjects from each group were mislabeled. The performance score was 0.90 for all metrics. Default hyperparameters were used in SVM and RF experiments.

To demonstrate the robustness of deep feature fusion model, the DenseNet and NasNet were used to classify scalograms as an ablation experiment. Table 7 shows the 10-fold cross validation average scores of the experiment. The average accuracies were 0.69 and 0.65 for DenseNet and NasNet, respectively. The experimental scores clearly indicated the superiority of the deep feature fusion approach.

Discussion

In the first part of the experiments, pre-trained methods were used as classifiers. The best scores for both accuracy and F1 scores were obtained with GoogleNet as 0.81 for "a" phonation (Fig. 9). The experimental results for "o" and "u" phonations were similar. The hybrid system improved

Table 5 The scores of the hybrid system

The hybrid system	Accuracy	Sensitivity	Specificity	Precision	F1 measure
Average	0.86	0.80	0.93	0.94	0.84
The best	1.00	1.00	1.00	1.00	1.00
The worst	0.75	0.75	0.75	0.75	0.75

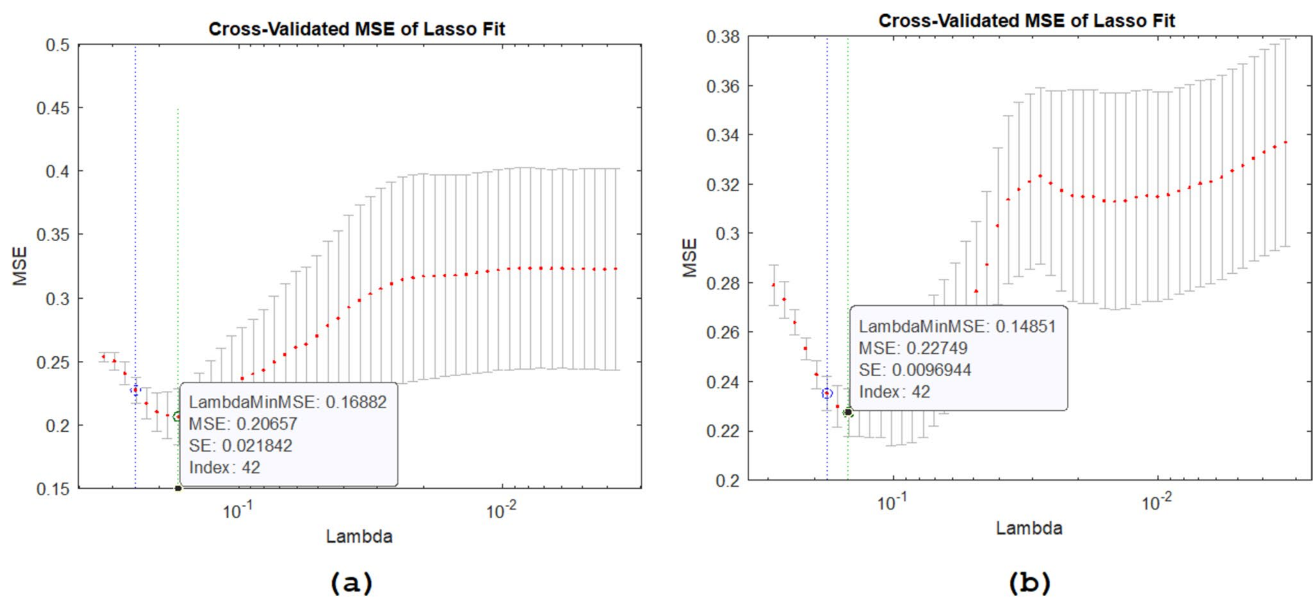


Fig. 6 Deep feature selection with lasso from **a** DenseNet and **b** NasNet

Table 6 The scores of the deep feature fusion experiments

	Accuracy	Sensitivity	Specificity	Precision	F1 measure
KNN	0.95	0.95	0.95	0.95	0.95
SVM	0.90	0.90	0.90	0.90	0.90
RF	0.88	0.90	0.85	0.86	0.88

scores. The accuracy and F1 scores of the hybrid system were 0.86 and 0.84, respectively. State-of-the-art scores were acquired in the second part, the deep feature fusion method. The fused deep features were tested in several classifiers. The best scores for accuracy and F1 measure were obtained as 0.95 in the KNN experiment.

This study represented an extended version of a previous work [47] where additional phonation types and classifiers were investigated. Time-frequency plot representations were also employed to detect PD in various languages. Karan et al. [48] and Wodzinski et al. [49] studied on a Spanish

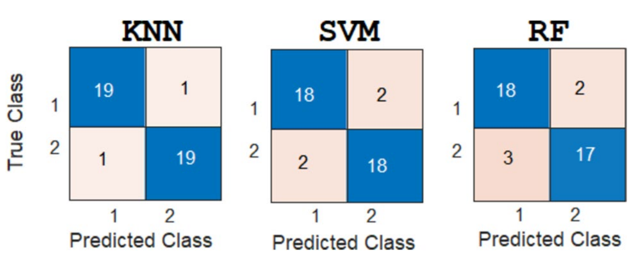


Fig. 7 Confusion matrixes for deep feature fusion experiments (1, PD patient; 2, healthy subject)

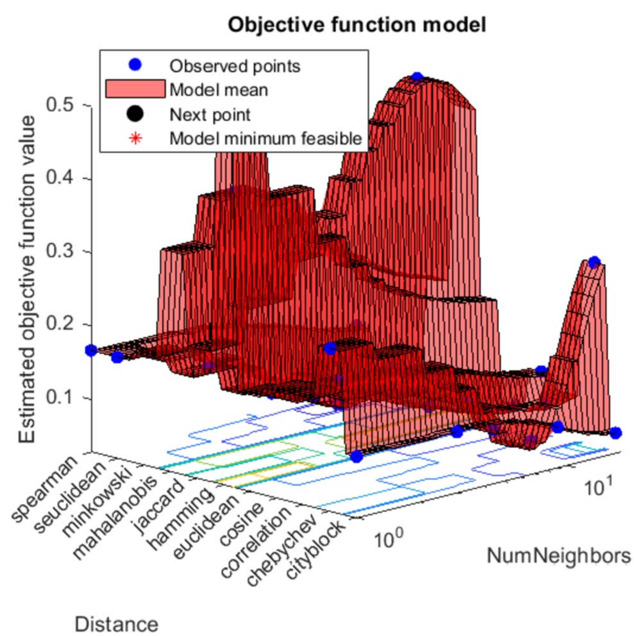


Fig. 8 Hyperparameter tuning of KNN

Table 7 Average scores of the ablation experiment

	Accuracy	Sensitivity	Specificity	Precision	F1 measure
DenseNet	0.69	0.66	0.72	0.74	0.67
NasNet	0.65	0.62	0.68	0.69	0.62

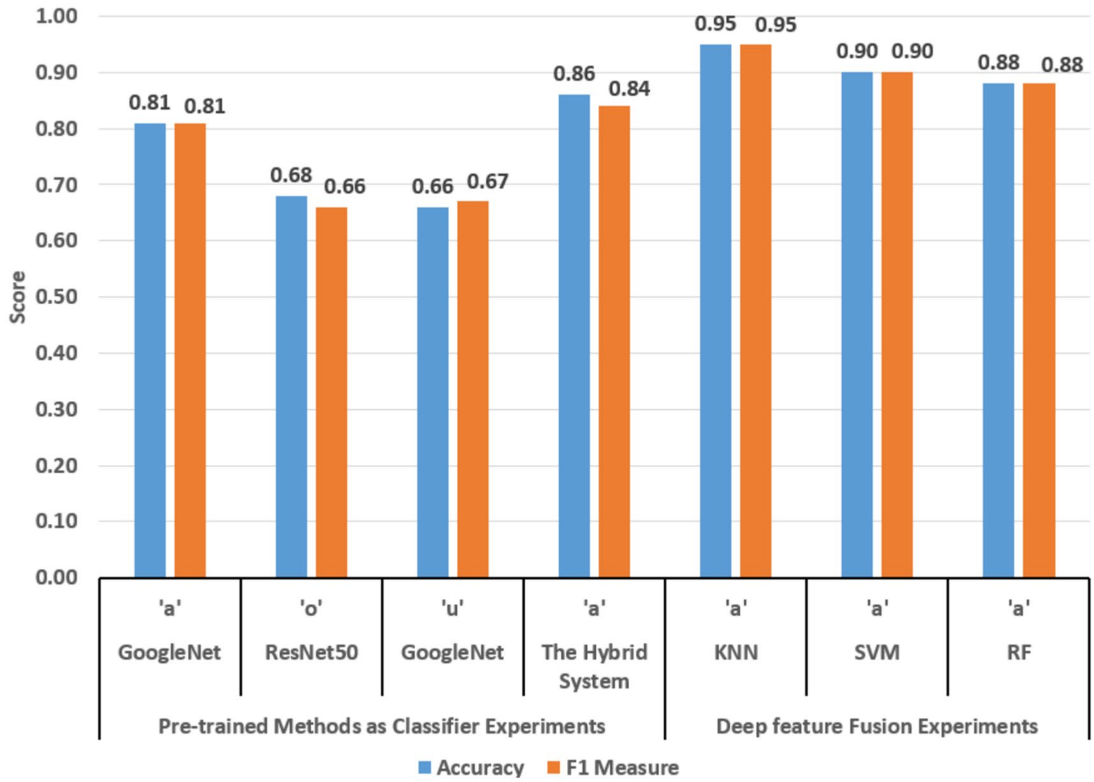


Fig. 9 Performance comparison of the systems

Table 8 Performance comparison with similar works

Study	Input	Method	Accuracy
Karan et al. [48]	Scalograms and spectrograms (i.e., obtained with STFT) Language: Spanish	Autoencoder-based features were classified with SVM and softmax	0.87 for spectrograms and 0.83 scalograms
Vásquez-Correa et al. [50]	Spectograms Language: Spanish, German, and Czech	Convolutional neural networks with transfer learning approach	0.72 for Spanish, 0.77 for German, and 0.73 for Czech
Wodzinski et al. [49]	Spectograms Language: Spanish	ResNet	0.92
Guatelli et al. [30]	Spectograms	AlexNet, VGG-16, SqueezeNet, Inception V3, and ResNet-50	0.84 average
Proposed method	Scalograms Language: Turkish	Deep feature fusion with DenseNet and NasNet	0.95

speech database. Vásquez-Correa et al. [50] worked on Spanish, German, and Czech voice recordings. Performance comparisons with similar works are given in Table 8. These studies used time-frequency representations with different methods to detect PD. The proposed method obtained state-of-the-art scores.

Additionally, Celik and Başaran [24] developed a deep learning–based system with random forest implementation and obtained an accuracy of 0.99. Madruga et al. [51] studied on effective remote detection of PD and proposed multicondition training system to eliminate recording device mismatch problem. Reddy and Alku [28] utilized exemplar-based sparse representation to detect PD from sound signals and obtained an accuracy of 0.83.

Conclusion

In this study, speech signals of PD patients were analyzed by using artificial intelligence methods. The speech signals were converted into scalograms using CWT. The scalograms were classified to detect PD with different deep learning methods: use of pre-trained methods as classifiers, the hybrid system, and the deep feature fusion method. The best accuracy score was obtained in the deep feature fusion experiment as 0.95, improving accuracy by 38% over the baseline study. The novelties of the paper are the evaluation of scalograms in the Turkish speech corpus for the detection of PD with a deep feature fusion approach and state-of-the-art results. In addition, the performances of the three phonation types were also examined.

Despite the potential limitations posed by the dataset size, the study contributes valuable findings regarding the application of deep feature fusion approach and the analysis of scalograms for PD detection. It highlights the importance of further research with larger datasets to validate and extend the findings obtained in this study. Different techniques for generating time-frequency plots can be employed to enhance accuracy in PD detection, as a future work. Additionally, sampling frequency rate in CWT may significantly impact the performance of the classifiers, although higher sampling frequencies can increase computational complexity. The proposed method can also be applied to detect the PD severity.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). Authors state no funding involved.

Data Availability The datasets generated during and/or analyzed during the current study are confidential.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. De Rijk MD, Tzourio C, Breteler MM, Dartigues JF, Amaducci L, López-Pousa S, et al. Prevalence of parkinsonism and Parkinson's disease in Europe: the EUROPARKINSON Collaborative Study. European Community Concerted Action on the Epidemiology of Parkinson's disease. J Neurol Neurosurg Psychiatry. 1997;62(1):10–5.
2. Feigin VL, Nichols E, Alam T, Bannick MS, Beghi E, Blake N, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(5):459–80.
3. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397(10291):2284–303.
4. Funayama M, Nishioka K, Li Y, Hattori N. Molecular genetics of Parkinson's disease: contributions and global trends. J Hum Genet. 2023;68(3):125–30.
5. Khatamino P, Cantürk İ, Özyılmaz L. A deep learning-CNN based system for medical diagnosis: an application on Parkinson's disease handwriting drawings. In: 2018 6th International Conference on Control Engineering & Information Technology (CEIT). IEEE; 2018.
6. Cantürk İ. Fuzzy recurrence plot-based analysis of dynamic and static spiral tests of Parkinson's disease patients. Neural Comput Appl. 2021;33(1):349–60.
7. Cantürk İ. A computerized method to assess Parkinson's disease severity from gait variability based on gender. Biomed Signal Process Control. 2021;66:102497.
8. Cantürk İ. A feature driven ıntelligent system for neurodegenerative disorder detection: an application on speech dataset for diagnosis of parkinson's disease. Int J Artif Intell Tools. 2021;30(03):2150011.
9. Cantürk İ. Parkinson Hastalığının Derecesi ile Yürüyüş Değişkenliği Arasındaki İlişkinin Bulanık Tekrarlılık Grafiğine Göre Araştırılması. Avrupa Bilim ve Teknoloji Dergisi. 2020;19:410–9.
10. Pawlukowska W, Skonieczna-Żydecka K, Rotter I, Honczarenko K, Nowacki P. The analysis of relationship between selected sociodemografic factors and disorders of speech organs in Parkinsons patients. BMC Neurol. 2017;17(1):1–9.
11. Roberts A, Post D. Information content and efficiency in the spoken discourse of individuals with Parkinson's disease. J Speech Lang Hear Res. 2018;61(9):2259–74.
12. Sakar CO, Kursun O. Telediagnosis of Parkinson's disease using measurements of dysphonia. J Med Syst. 2010;34:591–9.
13. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform. 2013;17(4):828–34.

14. Cantürk İ, Karabiber F. A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types. Arab J Sci Eng. 2016;41:5049–59.
15. Parisi L, RaviChandran N, Manaog ML. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. Expert Syst Appl. 2018;110:182–90.
16. Moro-Velazquez L, Gómez-García JA, Godino-Llorente JI, Villalba J, Orozco-Arroyave JR, Dehak N. Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's disease. Appl Soft Comput. 2018;62:649–66.
17. Arias-Vergara T, Vásquez-Correa JC, Orozco-Arroyave JR, Nöth E. Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions. Speech Commun. 2018;101:11–25.
18. Grover S, Bhartia S, Yadav A, Seeja K. Predicting severity of Parkinson's disease using deep learning. Procedia Comput Sci. 2018;132:1788–94.
19. Aggarwal N, Saini B, Gupta S. Role of artificial ıntelligence techniques and neuroimaging modalities in detection of Parkinson's disease: a systematic review. Cognit Comput. 2023:1–38.
20. Anter AM, Mohamed AW, Zhang M, Zhang Z. A robust intelligence regression model for monitoring Parkinson's disease based on speech signals. Futur Gener Comput Syst. 2023;147:316–27.
21. Quan C, Ren K, Luo Z, Chen Z, Ling Y. End-to-end deep learning approach for Parkinson's disease detection from speech signals. Biocybern Biomed Eng. 2022;42(2):556–74.
22. Ngo QC, Motin MA, Pah ND, Drotár P, Kempster P, Kumar D. Computerized analysis of speech and voice for Parkinson's disease: a systematic review. Comput Methods Programs Biomed. 2022:107133.
23. Pauline SH, Dhanalakshmi S, Kumar R, Narayanamoorthi R. Noise reduction in speech signal of Parkinson's disease (PD) patients using optimal variable stage cascaded adaptive filter configuration. Biomed Signal Process Control. 2022;77:103802.
24. Celik G, Başaran E. Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals. Appl Acoust. 2023;211:109476.
25. Khaskhoussy R, Ayed YB. Improving Parkinson's disease recognition through voice analysis using deep learning. Pattern Recogn Lett. 2023;168:64–70.
26. Warule P, Mishra SP, Deb S. Time-frequency analysis of speech signal using chirplet transform for automatic diagnosis of Parkinson's disease. Biomed Eng Lett. 2023:1–11.
27. Campi M, Peters GW, Toczydlowska D. Ataxic speech disorders and Parkinson's disease diagnostics via stochastic embedding of empirical mode decomposition. PLoS ONE. 2023;18(4):e0284667.
28. Reddy MK, Alku P. Exemplar-based sparse representations for detection of Parkinson's disease from speech. IEEE/ACM Trans Audio Speech Lang Process. 2023;31:1386–96.
29. Wen P, Zhang Y, Wen G. Intelligent personalized diagnosis modeling in advanced medical system for Parkinson's disease using voice signals. Math Biosci Eng. 2023;20(5):8085–102.
30. Guatelli R, Aubin V, Mora M, Naranjo-Torres J, Mora-Olivari A. Detection of Parkinson's disease based on spectrograms of voice recordings and extreme learning machine random weight neural networks. Eng Appl Artif Intell. 2023;125:106700.
31. Ali L, Chakraborty C, He Z, Cao W, Imrana Y, Rodrigues JJPC. A novel sample and feature dependent ensemble approach for Parkinson's disease detection. Neural Comput Appl. 2023;35(22):15997–6010.
32. Pramanik M, Pradhan R, Nandy P, Bhoi AK, Barsocchi P. The ForEx++ based decision tree ensemble approach for robust detection of Parkinson's disease. J Ambient Intell Humaniz Comput. 2023;14(9):11429–53.
33. Hemmerling D, Wojcik-Pedziwiatr M. Prediction and estimation of Parkinson's disease severity based on voice signal. J Voice. 2022;36(3):439.e9–20.
34. Li Y, Liu C, Wang P, Zhang H, Wei A, Zhang Y. Envelope multi-type transformation ensemble algorithm of Parkinson speech samples. Appl Intell. 2023;53(12):15957–78.
35. Laganas C, Iakovakis D, Hadjidimitriou S, Charisis V, Dias SB, Bostantzopoulou S, et al. Parkinson's disease detection based on running speech data from phone calls. IEEE Trans Biomed Eng. 2022;69(5):1573–84.
36. Rana A, Dumka A, Singh R, Rashid M, Ahmad N, Panda MK. An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. Electronics. 2022;11(22):3782.
37. Alalayah KM, Senan EM, Atlam HF, Ahmed IA, Shatnawi HSA. Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method. Diagnostics. 2023;13(11):1924.
38. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat Preced. 2008:1.
39. Vetterli M, Herley C. Wavelets and filter banks: theory and design. IEEE Trans Signal Process. 1992;40:2207–32.
40. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
41. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
42. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
43. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
44. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
45. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
46. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol. 1996;58(1):267–88.
47. Cantürk İ. Fine tuning approach on scalogram ımages to detect a neurological disorder. In: International Conference on Engineering and Applied Natural Sciences. 2022.
48. Karan B, Sahu SS, Mahto K. Stacked auto-encoder based time-frequency features of speech signal for Parkinson disease prediction. In: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP). IEEE; 2020.
49. Wodzinski M, Skalski A, Hemmerling D, Orozco-Arroyave JR, Nöth E. Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2019.
50. Vásquez-Correa JC, Arias-Vergara T, Rios-Urrego CD, Schuster M, Rusz J, Orozco-Arroyave JR, et al. Convolutional neural networks and a transfer learning strategy to classify Parkinson's disease from speech in three different languages. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019, Havana, Cuba, October 28–31, 2019, Proceedings 24. Springer; 2019.
51. Madruga M, Campos-Roca Y, Pérez CJ. Addressing smartphone mismatch in Parkinson's disease detection aid systems based on speech. Biomed Signal Process Control. 2023;80:104281.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.