Service Oriented Computing and Applications (2024) 18:101–107
https://doi.org/10.1007/s11761-023-00372-w

SPECIAL ISSUE PAPER

Parkinson disease prediction using machine learning-based features from speech signal

Linlin Yuan[1] · Yao Liu[2] · Hsuan-Ming Feng[3]

Received: 7 March 2023 / Revised: 5 June 2023 / Accepted: 12 June 2023 / Published online: 27 June 2023

Abstract

Parkinson's disease (PD) is a prevalent neurodegenerative disorder that has prompted the development of telediagnosis and remote monitoring systems. Dysphonia, a common symptom in the early stages of PD, affects approximately 90% of patients. Therefore, testing for persistent pronunciation or dysphonia in continuous speech can aid in the diagnosis of PD. Our study utilized speech signals from 252 subjects as the dataset. In this study, language signal features were used as input to machine learning algorithms, and the resulting classifiers were integrated to improve accuracy in the classification of Parkinson's disease (PD). The experimental results demonstrated a diagnostic accuracy of up to 95% using these machine learning algorithms. Additionally, a method of feature extraction based on clinical experience was presented for analyzing subjects' language signals.

Keywords Parkinson disease · Machine learning · Speech signal

1 Introduction

Parkinson's disease (PD) is a prevalent neurodegenerative disorder that affects the central, peripheral, and intestinal nervous systems of humans [1]. The early stages of the disease are characterized by a range of symptoms, including static tremors, muscle rigidity, dyskinesia, uncoordinated body movements, and abnormal postures [2]. Parkinson's disease (PD) is a neurodegenerative disorder characterized by the death of neurons. This is caused by pathological changes in the substantia nigra cells, which are located in the middle part of the human brain. These changes result in a significant reduction in dopamine synthesis. PD patients often experience symptoms such as depression, anxiety, physical fatigue, pain, night sweats, and excessive salivation [2, 3]. In PD, there is an imbalance between dopamine and acetylcholine in the body. Specifically, there is a decrease in dopamine and a relative increase in acetylcholine. This imbalance leads to physiological symptoms such as tremors and paralysis [4].

According to studies, the incidence of Parkinson's disease (PD) increases as people age, with the highest incidence occurring between the ages of 70 and 79 [1]. These findings suggest that as the population ages, there will be a significant rise in the number of people diagnosed with PD [5]. In fact, some studies have reported a 21.9% increase in incidence rates between 1990 and 2017 [1]. As medical visits and expenses continue to increase, there is a growing need for a reliable telemedicine system. Early diagnosis is crucial in improving patients' quality of life and extending their lifespan, making it imperative to establish such a system [1].

The unified PD rating scale has been traditionally used by doctors to monitor the progress of PD and assess the outcomes of interventions such as surgery and internal medicine [6]. However, in light of the current COVID-19 pandemic [7], telediagnosis has gained importance as a means to reduce hospital visits, limit contact between patients and healthcare providers, and ensure the convenience of patients [2]. According to research, 90% of individuals with Parkinson's disease exhibit speech disorders and resting tremors, among other symptoms [8]. Remote collection of voice information enables patients to self-assess their condition from home, eliminating the need for frequent hospital visits and reducing the risk of viral infections during an epidemic. Additionally,

✉ Hsuan-Ming Feng
hmfenghmfeng@gmail.com

1 College of Physical Education, Huaqiao University, Xiamen 361021, China

2 College of Medicine, Huaqiao University, Quanzhou 362021, Fujian, China

3 Department of Computer Science and Information Engineering, National Quemoy University, Kinmen, Taiwan

remote monitoring by healthcare professionals facilitates early detection and treatment of the disease.

This study aims to examine the precision of machine learning techniques in detecting Parkinson's disease and their suitability for monitoring patients remotely. Although telemedicine presents numerous advantages, it has certain limitations. One of the possible challenges to its implementation is the shortage of technology accessibility, especially in rural or low-income regions where patients may not have the required equipment or internet access [9]. Although some patients may prefer face-to-face interactions with their doctors, telemedicine may not be able to provide the same level of personal connection and trust, as some individuals may feel more comfortable speaking with their healthcare provider in person [10]. Additionally, there may be concerns about the security and privacy of patient data when using telemedicine, which must be addressed to ensure patient confidentiality [11]. However, despite these limitations, telemedicine has the potential to revolutionize healthcare delivery and improve patient outcomes, especially during the current pandemic where remote healthcare services are crucial in reducing the spread of the virus.

The field of machine learning has numerous applications, ranging from network security [12, 13] to software development [14]. Recently, it has gained popularity in biomedical research [15] and has proven to be effective in diagnosing various diseases [1]. Deep neural networks have rapidly evolved and are widely used across different disciplines due to their exceptional predictive performance and robustness to over-fitting [16, 17]. The present study employs a machine learning approach to diagnose PD. The study utilizes a deep neural network to analyze voice data obtained from patients. Deep neural networks are a form of artificial neural network that comprises several interconnected nodes, allowing it to learn intricate patterns and relationships present in the data. The paper introduces a deep neural network architecture that incorporates multiple hidden layers and utilizes the rectified linear unit (ReLU) activation function [18]. This study utilizes a vast dataset of voice recordings from both healthy individuals and PD patients to train the network. The diagnosis is based on multiple acoustic features extracted from the voice recordings such as pitch, jitter, and shimmer [19]. The use of deep neural networks has shown positive outcomes in PD diagnosis, and this research aims to explore their potential in telemedicine applications.

2 Related work

2.1 Machine Learning

In recent years, machine learning (ML) has emerged as a promising tool for diagnosing Parkinson's disease (PD), offering several advantages over traditional diagnostic approaches. One key advantage of ML is its ability to analyze large amounts of data and identify patterns that may not be apparent to human observers. In the case of PD diagnosis, ML algorithms can analyze patient data such as clinical assessments, imaging scans, and genetic information to identify subtle changes that may indicate the presence of PD. For instance, in a study conducted by Arora et al. [20], a ML algorithm was utilized to analyze MRI scans of patients with PD and healthy controls. The study found that the algorithm was able to accurately distinguish between the two groups with high sensitivity and specificity [20].

Machine learning (ML) has an added benefit in Parkinson's disease (PD) diagnosis as it can reduce inter-observer variability. Clinical assessments for PD diagnosis are often subjective and can vary between different clinicians. However, ML algorithms can be trained on a large dataset of patient data and learn to identify consistent patterns across multiple assessments and observers. A study by Zhang et al. (2020) demonstrated the improved accuracy and reliability of PD diagnosis through the use of a ML algorithm to analyze clinical assessments of patients with PD and healthy controls [21].

In recent years, machine learning has shown great potential in diagnosing Parkinson's disease. Compared to traditional diagnostic methods, ML algorithms have the ability to analyze large amounts of patient data, detect subtle changes that may indicate the presence of PD, and account for inter-observer variability [22]. With further advancements in these techniques, they could become an increasingly valuable tool for early diagnosis and treatment of PD.

2.2 Dataset source

Dataset-1 was obtained from DataFountain (https://www.datafountain.cn/datasets/146). Dataset-2 was obtained from UCI (https://archive.ics.uci.edu/ml/datasets/Parkinsons). Dataset-3 was also obtained from UCI (https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings).

2.3 Data description

In this study, data from 188 PD patients (107 males and 81 females) aged between 33 and 87 years were collected from the Department of Neurology in Cerrahpaşa Faculty of Medicine, Istanbul University. The control group comprised of 64 healthy individuals (23 males and 41 females) aged between 41 to 82 years old (61.1 $\pm$ 8.9). The dataset consisted of various features such as baseline features, time–frequency features, mel-frequency cepstral coefficients (MFCCs), wavelet transform-based features, and vocal fold features. Detailed descriptions of these features were previously reported in [8].

Dataset-2 was obtained from the UCI Machine Learning Repository and contains data from 195 individuals, including 48 patients with Parkinson's disease and 147 healthy controls. The data was collected using a microphone while the subjects performed sustained phonation of the vowel sound "/a/". The dataset includes 22 features, including various acoustic measures such as fundamental frequency, jitter, and shimmer.

Dataset-3 was also obtained from the UCI Machine Learning Repository and contains data from 104 individuals, including 47 patients with Parkinson's disease and 57 healthy controls. The data were collected using a microphone while the subjects performed sustained phonation of the vowel sound "/a/" and sustained phonation of the vowel sound "/o/". The dataset includes 26 features, including various acoustic measures such as fundamental frequency, jitter, and shimmer, as well as additional features such as formants and spectral entropy.

We used these three data sets through various speech signal processing algorithms and obtained 8 main features and 747 feature variables, which are explained in Table 1. This study utilizes multiple datasets to conduct a thorough analysis of the potential of using a deep neural network to diagnose Parkinson's disease through voice data. The selection of datasets was based on their availability and the range of features they offer. The references provided contain detailed descriptions of the specific features used in this study.

Table 1 Total table of feature data variables for the three datasets

Variable name	Descriptions	Numbers
Baseline features	The basic parameters of the speech signal, such as speech duration, fundamental period, pitch, and volume	23
Time–frequency features	The frequency characteristics of the speech signal, such as frequency range and spectrogram	4
Mel-frequency cepstral coefficients	Be obtained by transforming the speech signal into a Mel-frequency representation and then calculating the cepstral coefficients	84
Wavelet features	The temporal and spectral characteristics of the speech signal, such as instantaneous frequency and amplitude	182
Vocal fold	The sound source properties of the speech signal, such as the speaker's vocal characteristics	22
Tunable Q-factor features	The frequency characteristics of the speech signal, such as the positions and widths of the resonance peaks	432

3 Method

3.1 Classification

In this study, various machine learning algorithms were used to analyze the standardized speech feature data. These algorithms included logistic regression, random forest, k-nearest neighbors, and deep neural network classifiers, as shown in Fig. 1. The feature extraction module comprised of several subsets, such as baseline features, time–frequency features, mel-frequency cepstral coefficients (MFCCs), wavelet transform-based features, and vocal fold features.

To assess the efficiency of the classifier, we used the metrics of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which were calculated based on the test data. TP denotes a correctly classified patient, TN denotes a correctly classified healthy person, FP denotes a patient misclassified as a healthy person, and FN denotes a healthy person misclassified as a patient.

This study examined Parkinson's speech diagnosis classifiers and evaluated their advantages and disadvantages using metrics such as accuracy, precision, recall, and F1-score. These metrics are commonly used for classification problems and are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}}. \tag{4}$$

Furthermore, to show the overall performance of our classifiers, we calculated the MCC as follows:

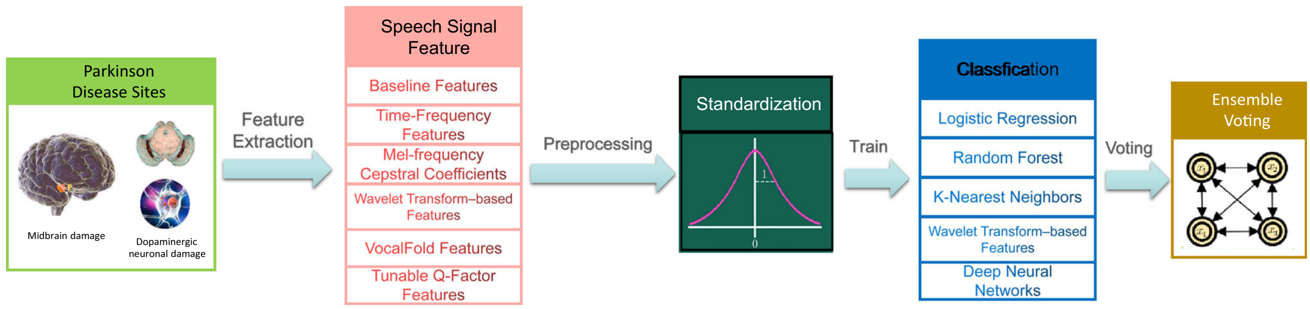

Fig. 1 Overview of the proposed Parkinson's disease (PD) classification system

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} \times \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{5}$$

Following the feature extraction step, the feature vector was standardized to ensure that each feature had a zero mean and unit variance. Subsequently, the feature subset was fed into multiple classifiers to differentiate between healthy subjects and patients with Parkinson's disease.

The three recorded sets of data were averaged, and the median was taken to balance the data. The standardized data were then randomly split into training and testing sets with a 4:1 ratio. To compare the results for different feature subsets and classifiers, we used overall accuracy, F1-score, and Matthews correlation. Figure 1 provides an overview of the entire PD classification process. The input data undergoes feature extraction where various features such as baseline, time–frequency, MFCCs, wavelet transform-based and vocal fold features are filtered. Standardization is used for pre-processing. The classification module uses logistic regression, random forest, KNN, and DNN classifiers. The classification results are combined using an ensemble voting mechanism.

3.2 Ensemble of classifiers

In this study, a combined method was used for the classifier prediction when training multiple classifiers for the same classification task. The learning algorithms as described above were utilized. In cases where the classification results were uncertain, the classifier would prioritize the previous better prediction outcomes.

3.3 Feature ranking

To select the most effective features and obtain more robust and accurate PD classification features, we used a method based on the minimum redundancy–maximum correlation (MRMR) [1]. The mRMR method is based on the correlation coefficient of the features with the predicted results and aims to reduce the redundancy between numbers [1]. As a preprocessing step, the mRMR algorithm has been successfully applied to a variety of machine learning problems, including gene expression analysis, protein interactions, protein structure prediction, and biomedical decision support systems [1]. Therefore, we used the mRMR to obtain the characteristics of the most critical language signals of PD.

The study utilized various feature subsets and classifiers in its machine learning algorithms, and employed an ensemble of classifiers to enhance classification performance. Additionally, a feature ranking method was implemented to identify the most impactful features for PD classification. These details contribute to a clearer comprehension of the methodology and facilitate potential replication of the study.

4 Experimental results

4.1 Classification performances for all features

This study utilized a training set consisting of 437 samples and a validation set consisting of 50 samples to train our model. The training accuracy and validation accuracy were recorded for each training epoch. Figure 2 shows a rapid increase in both training accuracy and validation accuracy in the first 100 epochs of training, followed by a gradual slowing of improvement in subsequent epochs. The final training results showed a training accuracy of 0.95 and a verification accuracy of 0.89. It is noteworthy that the training accuracy and validation accuracy increased in a similar trend.

In further studies, we utilized the full data, average, and median of three sets of data for a single subject in dataset-1 to examine various characteristics. Table 2 displays the accuracy for the test set achieved by three different processing methods that employed the same dataset. The term 'mean' denotes the average, while 'med' denotes the median. Additionally, we have included the voting performance. The ensemble of classifiers, referred to as 'vote' in Fig. 3, achieved an accuracy of 0.95, an F1-score of 0.97, and an MCC of 0.87. After balancing the data using averages and

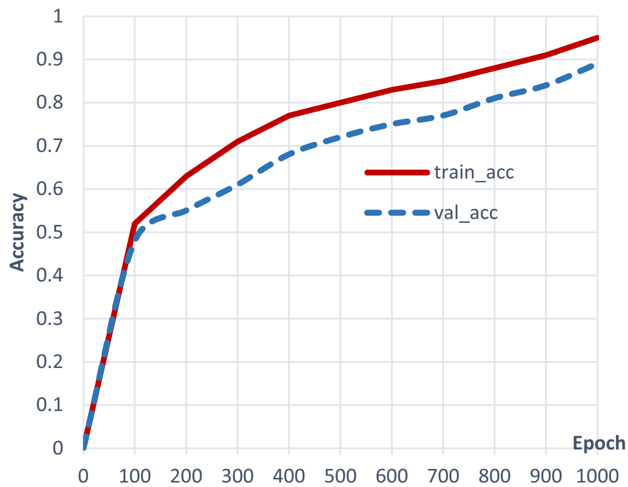

Fig. 2 The accuracy curve of our method in the training dataset and the validation dataset

medians for individual subjects, the deep neural network demonstrated improved performance with an accuracy, F1-score, and MCC mean of 0.92, 0.94, and 0.53, respectively. The median accuracy showed even better results with values of 0.94, 0.97, and 0.81 for accuracy, F1-score, and MCC, respectively.

4.2 Dataset analysis

In dataset-1, the same subject provided a voice recording three times in a row. However, due to the varying statuses of patients with Parkinson's disease (PD), the first voice recording may sound similar to that of a healthy subject, while subsequent recordings may exhibit significant differences. It is our hypothesis that the original dataset did not fully utilize this feature in speech data feature extraction. This finding could be used as a potential feature for diagnosing Parkinson's disease through speech analysis or for optimizing the dataset.

In our study, we analyzed 443 features and found significant differences ($P < 0.05$) between subjects with and without disease through analysis of variance. We then used mRMR to screen the most important 50 features and identified an intersection of 21 features. These 21 features were not only statistically significant but also key features for machine learning. We calculated the standard deviation of the features in this intersection and found that 18 features had larger standard deviations in Parkinson's patients than in healthy subjects. In dataset-2 and dataset-3, we also obtained similar results. In dataset-2, of the 22 significantly different ($P < 0.05$) PD characteristics, the standard deviations of 19 features in the Parkinson's patients was larger than those in healthy subjects. In dataset-3, of the 16 significantly different ($P < 0.05$) PD characteristics, the standard deviations of 14 features in Parkinson's patients were larger than those in healthy subjects.

This showed that the triplicate language recording data of Parkinson's patients varied significantly. It is speculated that Parkinson's patients can pronounce words similarly to healthy people at first, but as the pronunciation continues, the phonetic differences begin to be more significant.

5 Result

The ensemble of classifiers (as shown in Table 1) achieved an accuracy of 0.95, F1-score of 0.97, and MCC of 0.87 when using all the data. While some reports in the literature have reported higher accuracy, it is important to note that many of these studies used datasets with multiple voice recordings of the same subject or smaller datasets. The aim of this study was to evaluate the reliability of datasets based on clinical experience and apply machine learning to Parkinson's speech diagnosis to achieve improved classification results.

6 Discussion

In the field of PD tele-diagnosis, the accuracy and reliability of the results are highly dependent on the selection of features to extract and the learning algorithm used. Previous

Table 2 Results obtained with all feature subsets

	All feature subsets			All feature subsets (mean)			All feature subsets (med)		
	Accuracy	F1-Score	MCC	Accuracy	F1-Score	MCC	Accuracy	F1-Score	MCC
Random forest	0.84	0.90	0.57	0.84	0.91	0.31	0.84	0.91	0.31
Logistic regression	0.82	0.89	0.51	0.86	0.92	0.44	0.86	0.92	0.44
KNN	0.91	0.94	0.76	0.82	0.90	0.25	0.82	0.90	0.25
DNN	0.92	0.95	0.76	**0.92**	**0.94**	**0.83**	**0.94**	**0.97**	**0.81**
Vote	**0.95**	**0.97**	**0.87**	0.86	0.92	0.44	**0.94**	**0.97**	**0.81**

The bold values indicate the highest values for these methods in each metric (such as accuracy)

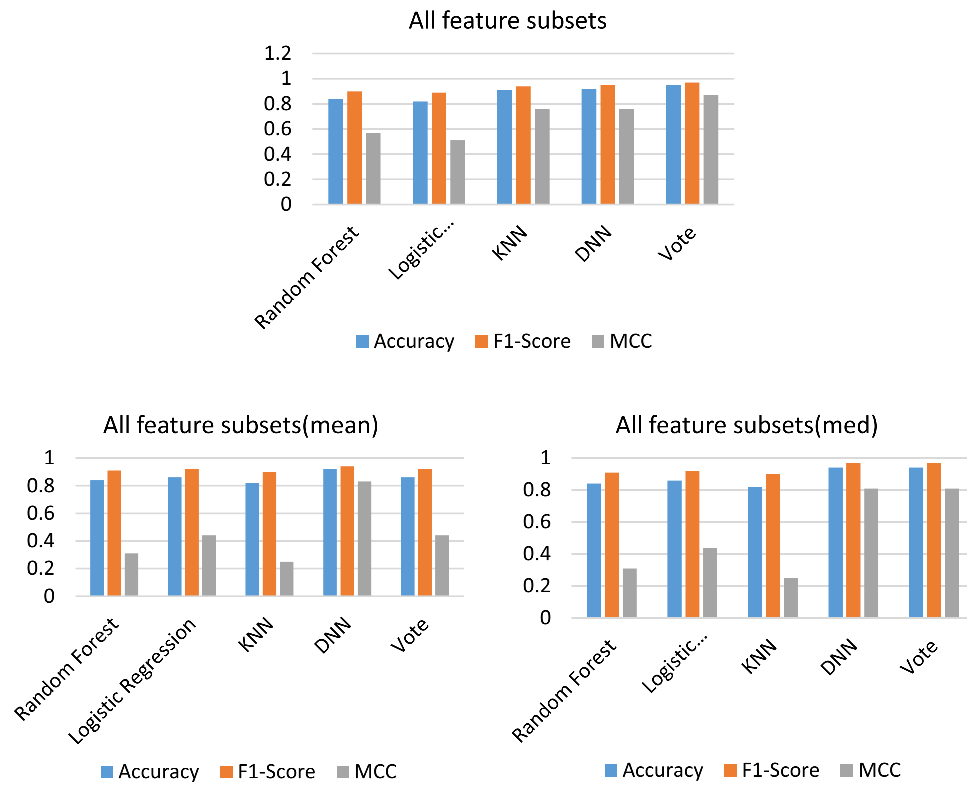

Fig. 3 The resulting performance of the three evaluation metrics in all feature subsets, (mean), and (med)

studies have primarily used a public dataset [19] consisting of 23 PD patients and 8 healthy subjects, which contains 195 sound measurements, to distinguish between PD patients and healthy individuals. However, some studies have used small datasets [1] or repeatedly recorded data. It is worth noting that certain classification models achieved almost 100% accuracy in distinguishing PD patients from healthy subjects on small datasets [23].

This study differs from previous research [24] in that it utilized a larger dataset and employed a deep neural network classifier. The study collected recordings from 252 subjects, consisting of 188 patients with PD and 64 healthy controls, and extracted various feature subsets from the recordings. The study utilized multiple machine learning classifiers and deep neural networks, and combined the predictions of each classifier into a voting classifier. The study compared and analyzed the accuracies of different datasets to evaluate the classifiers. The mean and median of subjects' data were also selected to test the classifier's effect on balanced numbers. The deep neural network showed better classification results with balanced data, while other machine learning classifiers tended to classify PD patients as healthy subjects.

Based on clinical experience, patients with Parkinson's disease tend to exhibit greater differences in voice recordings compared to healthy individuals, particularly with longer pronunciation times. This variability in data can lead to misclassification by the classifier. Furthermore, differences in language among individuals of different nationalities can also result in poor classification results. It is recommended that multiple measurements be taken during the second half of speech to obtain a more accurate classifier model.

Remote monitoring of Parkinson's disease (PD) using wearable devices has shown promising results in providing objective and continuous data for clinicians to monitor patients' symptoms and adjust treatment plans accordingly. However, there are limitations to this approach, such as the need for patients to have access to and be comfortable using the technology, as well as potential issues with data accuracy and privacy. In clinical practice, remote monitoring of PD could potentially improve patient outcomes by allowing for more personalized and timely adjustments to treatment plans, but it will require careful consideration of implementation and integration into existing healthcare systems. These limitations and potential applications should be further explored in future research.

Acknowledgements This work was partly supported by the Quanzhou Science and Technology Major Project under Grant No. 2021GZ1; the National Natural Science Foundation of Fujian under Grant No. 2021J011404; and the Quanzhou scientific and technological planning projects under Grant Nos. 2021C037R and 2019C028R.

Declarations

Conflict of interest None of the authors have any conflicts of interest that would affect the validity or integrity of the results reported in this work.

References

1. Baloch S, Baloch MA, Zheng T, Pei X (2020) The coronavirus disease 2019 (COVID-19) pandemic. Tohoku J Exp Med 250(4):271–278. https://doi.org/10.1620/tjem.250.271
2. Jankovic J (2008) Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79(40):368–376
3. Menza M, Dobkin RD (2005) Anxiety and Parkinson's disease. J. Neuropsychiatry 8(4):383–92
4. Dashtipour K, Tafreshi A, Lee J, Crawley B (2018) Speech disorders in Parkinson's disease: pathophysiology, medical management and surgical approaches. Neurodegener Dis Manag 8(5):337–348
5. Twelves D, Perkins KS, Counsell C (2003) Systematic review of incidence studies of Parkinson's disease. Mov Disord 18(1):19–31
6. Hsia CH, Liu CH (2022) New hierarchical finger-vein feature extraction method for iVehicles. IEEE Sens J 22(13):13612–13621
7. Baloch S, Baloch MA, Zheng T, Pei X (2020) The coronavirus disease 2019 (COVID-19) pandemic. Tohoku J Exp Med 250(4):271–278
8. Sakar CO et al (2018) A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Appl Soft Comput 74:255–263
9. Bashshur RL, Shannon GW, Krupinski EA (2019) The empirical foundations of telemedicine interventions for chronic disease management. Telemed e-Health 25(3):191–210
10. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M (2017) Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 7(8):e016242
11. El-Masri MM, Ali SA (2019) Privacy and security in telemedicine: a serious concern. Int J Adv Comput Sci Appl 10(2):544–549
12. Alwageed HS (2022) Detection of cyber attacks in smart grids using SVM-boosted machine learning models. SOCA 16:313–326. https://doi.org/10.1007/s11761-022-00349-1
13. Alshammari FH (2023) Design of capability maturity model integration with cybersecurity risk severity complex prediction using bayesian-based machine learning models. SOCA 17:59–72. https://doi.org/10.1007/s11761-022-00354-4
14. Pahl C (2023) Research challenges for machine learning-constructed software. SOCA 17:1–4. https://doi.org/10.1007/s11761-022-00352-6
15. Goecks J, Jalili V, Heiser LM, Gray JW (2020) How machine learning will transform biomedicine. Cell 181(1):92–101
16. Jhong SY, Yang PY, Hsia CH (2022) An expert smart scalp inspection system using deep learning. Sens Mater 34(4):1265–1274
17. Explainable AI (2021) A multispectral palm vein identification system with new augmentation features. ACM Transact Multimed Comput Commun Appl 17(35):1–21
18. Kriegeskorte N, Golan T (2019) Neural network models and deep learning. Curr Biol 29(7):R231–R236
19. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 56(4):1015
20. Arora S, Tiwari P, Sharma M, Madabhushi A (2018) A deep learning based radiomics approach for diagnosis of Parkinson's disease. In: Medical Imaging 2018: computer-aided diagnosis 10575: 105752K
21. Zhang J, Shi K, Huang K, Shen D (2020) Multimodal classification of Parkinson's disease based on comprehensive feature fusion and selection. Front Neurosci 14:309
22. Yang Y, Wei L, Hu Y, Wu Y, Hu L, Nie S (2021) Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. J Neurosci Methods 350:109019
23. Gürüler H (2017) A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Comput Appl 28:1657–1666
24. Karan B, Sahu SS, Mahto K (2020) Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybern Biomed Eng 40(1):249–264