A Comparative Study of Parkinson Disease Diagnosis in Machine Learning

Mohammed Y. Thanoun
Department of Electrical Engineering, University of Mosul, Mosul, Nineveh, 41002, Iraq

Mohammad T. Yaseen
Department of Electrical Engineering, University of Mosul, Mosul, Nineveh, 41002, Iraq

ABSTRACT

Parkinson's disease (PD) is a cumulative disorder in the nervous system. PD patients may experience difficulty in movement and speaking due to damages in certain parts in the brain. In this study, we propose using two types of Ensemble learning methods Stacking Classifier and voting classifier, which are potential methods of PD detection using machine learning. Then, we compared between the results of both of them. Stacking Classifier method outperformed voting classifier and the obtained accuracy was 92.2% and 83.57%, respectively. This comparative study would help come out with higher detection accuracy for medical applications such as this chronic disease.

CCS CONCEPTS

• **Machine Learning**;

KEYWORDS

Parkinson's Disease, Stacking Classifier, Voting Classifier

ACM Reference Format:
Mohammed Y. Thanoun and Mohammad T. Yaseen. 2020. A Comparative Study of Parkinson Disease Diagnosis in Machine Learning. In *2020 The 4th International Conference on Advances in Artificial Intelligence (ICAAI 2020), October 09–11, 2020, London, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3441417.3441425

1 INTRODUCTION

PD is one of the diseases that causes human body functions disorder such as mental functions and speaking ability disorder. Although a lot of researchers have studied this disease, there is no cure yet. Mostly old people get infected by PD disease. And the PD patients' number is expected to increase in the future. In fact, there are different clinical indicators for PD patients. The voice signal disorder, for instance, is one of the early signs of PD patients [1-3]. To date, the reason behind the PD disease is not completely understood. However, based on statistics, males tend to be more than females to get infected by PD disease [4]. The major issues with PD-speaking are lack of control, loudness and pitch monotony, decreased tension, inappropriate silence, fast speech rushes, variable speeds, low consonant articulation and rough, breathable sound (dysphonia). A future detective technique is appealing because the recording of voice-related effects is non-invasive and can conveniently be achieved using mobile devices.

There are various methods for the PD disease assessment such as Unified Parkinson's Disease Rating Scale (UPDRS). However, assessment accuracy is still one of the challenges in this field. Also, the PD disease detection in the early stage is very important [5, 6]. The researchers put tremendous efforts to overcome the detection methods difficulties such as utilizing machine learning technique. Machine learning has been exploited in many different fields. The medical applications field is one of the recent fields deployed machine learning in different medical applications. Machine leaning involves many methods and techniques. However, machine learning is based on dataset. The dataset needs to be acquired and arranged prior to the assessment. Assessment accuracy and consumed time are a tradeoff issue in this field. PD diagnosis is one of the important medical applications [7-10].

Several methods and techniques were suggested for the PD detection. For example, Random Forest algorithm, Support Vector Machine, Neural Networks, Particle swarm optimization, fuzzy based nonlinear transformation, and k-Nearest Neighbor method [11-14]. The working principle of these methods is based on different parameters to be recognized for PD patients. The speech signals disorder is one of the early signs of PD disease [15, 16]. Hence, various studies focused on speech signals. Collecting voice signals dataset from the PD patients and the healthy people are a vital process. Then classification of dataset is needed. The classified data then is studied to figure out the PD patients out of healthy people. Hybrid Artificial Intelligence was deployed for voice dataset [17-20] .

Other techniques were used for PD diagnosis such as artificial neural networks (ANN), which can provide good performance [15], Multi-Layer Perceptron (MLP) is good for features selection and classification [21, 22], and support vector machines (SVM) exhibited better performance as a regression technique [23], and Synthetic Minority Over-Sampling Technique (SMOTE) technique is good for transforming imbalanced dataset to balanced dataset. Tunable Q-factor wavelettransform (TQWT) was deployed for features extraction from the voice data. It showed good ability to improve the performance of feature extraction process for the PD detection [17, 18].

The study of speech disability in general and in the PD case in particular has led to the development, for example [28-31] and the algorithms that are involved in speech signal processing. It was seen [28] that PWP was able to discriminate between secure controls with approximately 90% of the overall classification with four dysphonia functions which are the most common speaker signals

ICAAI 2020, October 09–11, 2020, London, United Kingdom

ACM ISBN 978-1-4503-8784-2/20/10…$15.00
https://doi.org/10.1145/3441417.3441425

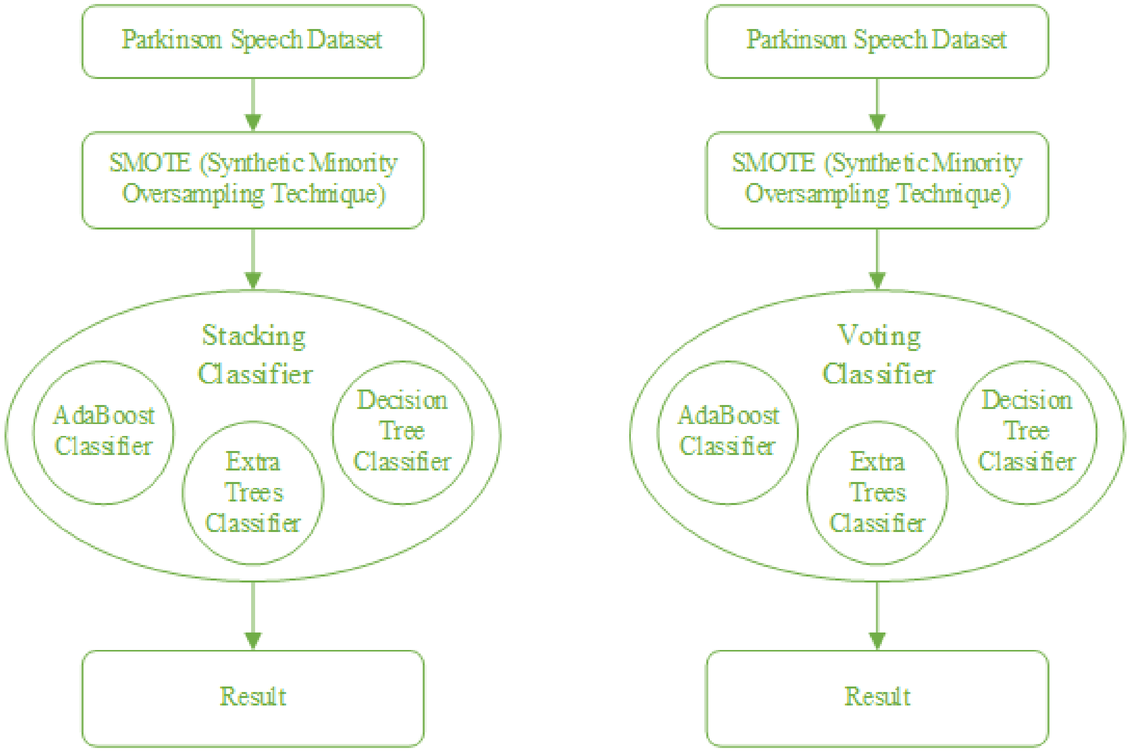

Figure 1: Flowchart architecture for Parkinson classification

algorithms used. This work included traditional calculating algorithms focused on basic frequency (grid measurement) disturbances, amplitude disorders and noise signal ratios (harmonics to noise ratio measures) [32-35]. Three experimental non-linear dysphonic experiments were also used in the study to supplement classical behavior.

In this work, a comparative study of the performance of two classification methods for the PD detection was conducted. The classification process was conducted based on a collection of speech signals from healthy and patient with PD. The proposed method of this comparative study is based on employing SMOTE technique because the available dataset is unbalanced dataset. After the dataset was balanced with SMOTE, a classifier with k-fold technique was used [17]. The paper is arranged as follows. Section 2 is the methodology. Section 3 is the experiment setup that includes dataset, techniques, and the performance evaluation for each technique, and the results. Finally, Section 4 is the conclusion.

2 METHODOLOGY

The proposed two types of Ensemble learning methods are (Stacking Classifier and voting classifier) and comparing between them by using the same types of classifiers algorithms which are (Adaboost classifier, decision tree classifier and Extra Tree Classifier) as a PD classifier as shown in figure 1. Initially, the first phase (Pre-process dataset) Apply data cleaning on the Parkinson speech dataset. In the second phase (Normalization) this phase will normalize the dataset by using the min-max technique. The third phase (balance) PD datasets have then been transformed into a representation of the balance classes with SMOTE [17]. The fourth phase (comparison) Construct the (stacking and voting) classifier with enhanced the dataset from phase two and Training the deep learning models as classifiers with the classifiers algorithms inside each of them are (Adaboost classifier, decision tree classifier, and Extra Tree Classifier). The last phase (Evaluation) Evaluate our proposed two ensemble learning techniques the PD detection accuracy for each type of ensemble learning technique [36, 37]. Whereas, the flow chart for our methodology is shown in figure 1

3 EXPERIMENT

3.1 Dataset

The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 at the Department of Neurology in CerrahpaÅŸa Faculty of Medicine, Istanbul University. The control group consists of 64 healthy individuals (23 men and 41 women) with ages varying between 41 and 82. During the data collection process, the microphone is set to 44.1 KHz and following the physician examination, the sustained phonation of the vowel /a/ was collected from each subject with three repetitions [1]. This dataset includes 753 features and 756 samples. This dataset is also a problem of two classes: 564 Parkinson patient and 192 healthy trials. This dataset is thus unbalanced. The 2018 PD dataset was developed by Sakar et al [18]. Attributes from

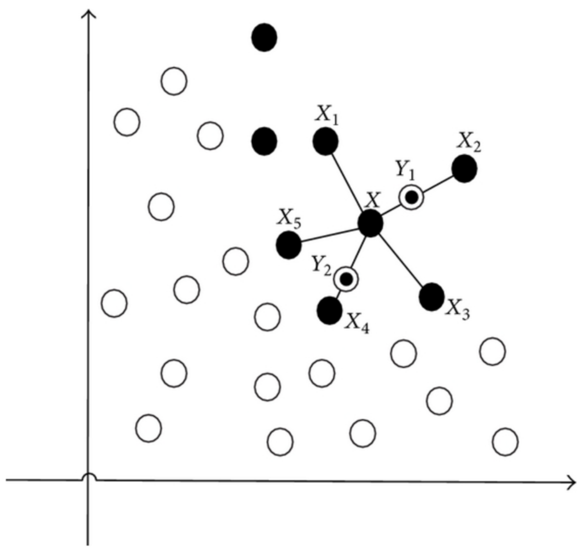

Figure 2: The SMOTE algorithm

voice signal processing techniques have been collected. Data from 188 PD patients (107 males and 81 females) between the ages of 33 to 87 have been taken from the PD dataset [18].

3.2 SMOTE

A problem with imbalanced classification is that there are too few examples of the minority class for a model to effectively learn the decision boundary. One way to solve this problem is to oversample the examples in the minority class. This can be achieved by simply duplicating examples from the minority class in the training dataset prior to fitting a model. This can balance the class distribution but does not provide any additional information to the model.

An improvement on duplicating examples from the minority class is to synthesize new examples from the minority class. This is a type of data augmentation for tabular data and can be very effective. Perhaps the most widely used approach to synthesizing new examples is called the Synthetic Minority Oversampling Technique, or SMOTE for short. This technique was described by [23].

SMOTE works by selecting examples that are close in the feature space, drawing a line between the examples in the feature space and drawing a new sample at a point along that line. Specifically, a random example from the minority class is first chosen. Then k of the nearest neighbors for that example are found (typically k=5). A randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space [17], see figure 2

3.3 EXPERIMENT SETUP

3.3.1 Ensemble Learning Techniques. The process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem. Ensemble learning is primarily used to improve the (classification, prediction, function approximation, etc.) Performance of a model, or reduce the likelihood of an unfortunate selection of a poor one. Other applications of ensemble learning include assigning a confidence to the decision made by the model, selecting optimal (or near optimal) features, data fusion, incremental learning, nonstationary learning and error-correcting [10], see figure 3

3.3.2 Voting Technique. The voting classifier works like an electoral system in which a prediction on a new data point is made based on a voting system of the members of a group of machine learning models. According to the scikit_learn's documentation, one may choose between the hard and the soft voting type. The hard-voting type is applied to predicted class labels for majority rule voting. This uses the idea of "Majority carries the vote" i.e. a decision is made in favor of whoever has more than half of the vote [24].

The soft voting, expects the class label depending on the argmax of the sums of the expected probabilities of the individual estimators that make up the ensemble. The soft voting is often recommended in the case of an ensemble of well-calibrated/fitted classifiers. E.g. if model 1 predicts A, and model 2 predicts B, and model 3 predicts A. The voting classifier (with voting='hard') returns A. In case of a tie, the voting classifier would choose the class depending on the ascending order [24].

3.3.3 Stacking Technique. Stacking comprises adding the predictions from various machine learning models in one dataset. We first specify/build some machine learning models called base estimators on our dataset, the results from these base learners then serve as input into our Stacking Classifier. The Stacking Classifier is able to learn when our base estimators can be trusted or not. Stacking enables us to use the strength of every single estimator through using their output as an input for the last estimator [25].

In Stacking Classifier, one can choose to apply the cross-validation at the base learner level or at the on to the final estimator. Using the scikit learn Stacking Classifier, the base learners are fully fit on the full X, meanwhile the last estimator is being trained through cross-validated predictions of the base learners. Multi-Layer stacking is also possible, where one builds layers of base learners before a final estimator is built [25], see figure 4

3.4 Performance Evaluation and Results

The performance of Stacking Classifier and voting classifier techniques were evaluated on PD dataset. This dataset was established by Sakar et al [18]. It was obtained from 188 PD patients and 64 healthy ones with different ages. Also, the dataset involved 753 features and 756 samples. The used dataset is imbalanced due to the numbers of PD patient and healthy individuals. The available dataset was deployed for training and evaluation the used techniques. The approach of 80:20 percent and 10-fold cross validation were deployed for training and evaluation. More training would improve the system evaluation performance. The performance evaluation of Stacking Classifier and voting classifier techniques was carefully analyzed and shown in Table 1. From Table 1, it is clear that Stacking Classifier technique outperforms Voting Classifier technique in differentiating between PD patients and healthy people. As shown in Table 1, using Stacking Classifier technique leads

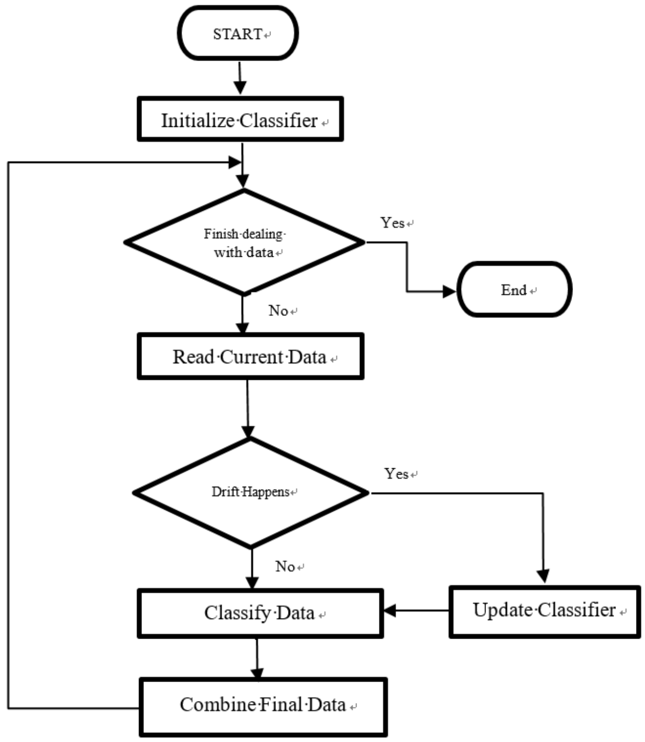

Figure 3: Flowchart of Ensemble Learning Techniques

Table 1: Performance Evaluation

Used Technique	Accuracy%
Stacking Classifier	92.2%
Voting Classifier	83.57%

to obtain better accuracy than that obtained accuracy of using Voting Classifier technique. Stacking Classifier technique was higher by (~1.1%) than Voting Classifier technique.

Although there are significant limits to speech as a specific biomarker for clinical diagnosis, hopefully algorithms that involve many types, including speech and brain testing, or accelerometers can be employed in conjunction to build a powerful clinical instrument to support neurologists identify PD and PD symptoms. As a result, the clinical model and others instruments have succeeded and also showed positive results independently from voice using accelerometer data [26, 27].

4 CONCLUSION

Automated master training architectures with only non-invasive diagnosis and prediction of illnesses as functionality, speech biomarkers were demonstrated. This comparative study showed the usefulness of different machine learning classifications with noisy and high-dimensional data on disease detection. Using two types of Ensemble learning methods Stacking Classifier and voting classifier were proposed in this comparative study. These classifiers are potential methods of PD detection using machine learning. Medical level specificity is achievable after a detailed collection of features. Two types of Ensemble learning techniques, Stacking Classifier and Voting Classifier techniques were used for this comparative study. The used techniques performances were compared in differentiating between healthy individuals and infected ones with PD. The obtained performance results showed that Stacking Classifier technique is more efficient than Voting Classifier technique in differentiating between healthy people and people with PD. Stacking Classifier technique reaches (92.2%) accuracy, while voting classifier technique was only achieved (82.57%) accuracy for the same dataset.

Therefore, before this technology can be used as a diagnostic decision support instrument, the results of this study will have to be further checked with separate data sets. The applicability of these results will be further confirmed by a review with a cohort of subjects with PD like vocal signs, but without the PD. While expression in other studies was used, in controlled circumstances the selection of lasting vowels decreases intraspeaker uncertainty and linguistic misunderstanding and can contribute to improved outcomes. Nevertheless, the synthesis of the two methods may be explored in future experiments to obtain data from both sustainable vowels and voice. A very comprehensive archive, containing voices from different diseases, can be used to assess the root pathology among a broad variety of potential conditions by using advanced dysphonic

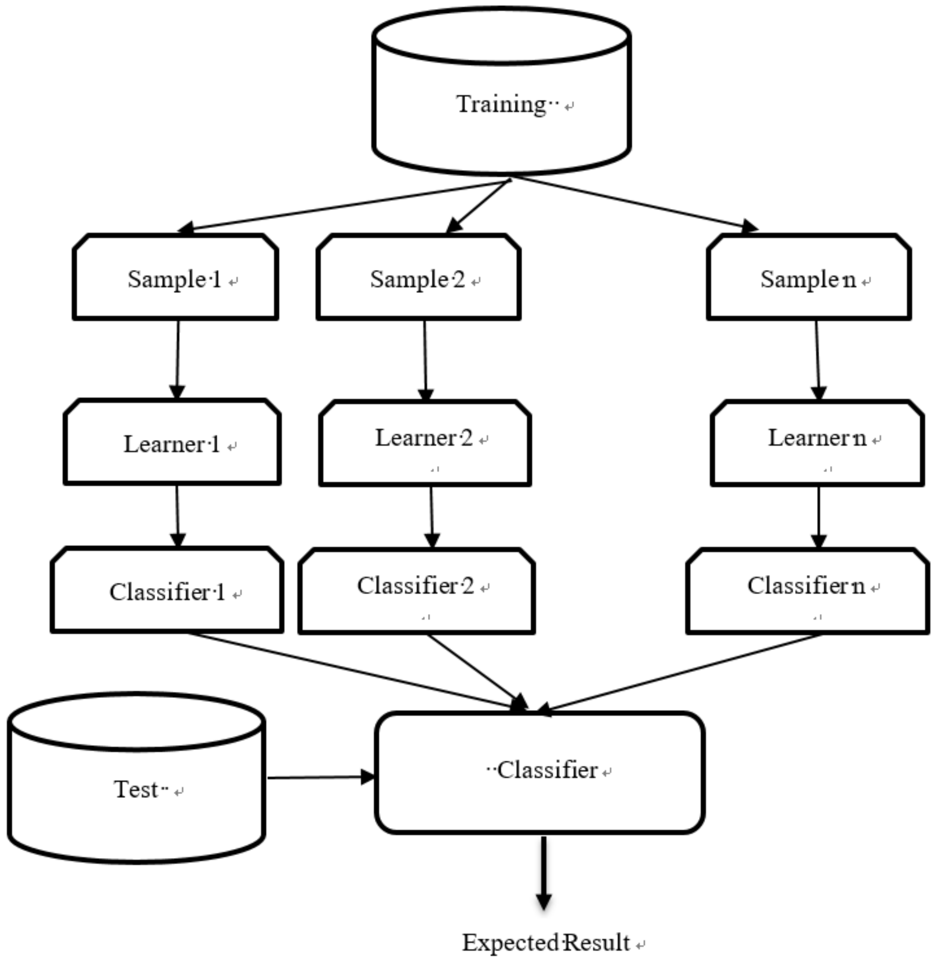

Figure 4: Flowchart of Stacking Technique

acts. It is fascinating in addition to we are actively working on applying these results to more functional sound configurations which would expand the technologies being proposed in more specific environments in the acoustically regulated atmosphere in this report. Finally, future work needs more analysis could provide more knowledge from physical models of processes of voice output to boost the precision of jitter, shimmer and HNR estimates using glottal-source signals from voice capture.

ACKNOWLEDGMENTS

The authors are very grateful to the University of Mosul, College of Engineering for their provided facilities, which helped to improve the quality of this work.

REFERENCES

[1] Pillon, B., Dubois, B., Cusimano, G., Bonnet, A.-M., Lhermitte, F., & Agid, Y. (1989). Does cognitive impairment in Parkinson's disease result from non-dopaminergic lesions? Journal of Neurology, Neurosurgery & Psychiatry, 52(2), 201-206.

[2] Bloem, B.R. and M. Munneke, Revolutionising management of chronic disease: The ParkinsonNet approach. Bmj, 2014. 348: p. g1838.

[3] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," Neurocomputing, vol. 184, pp. 131-144, 2016.

[4] D. Calne, B. Snow, and C. Lee, "Criteria for diagnosing Parkinson's disease," Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, vol. 32, no. S1, pp. S125-S127, 1992.

[5] H.-L. Chen *et al.*, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach," Expert systems with applications, vol. 40, no. 1, pp. 263-271, 2013.

[6] C. Sommer and D. W. Gerlich, "Machine learning in cell biology–teaching computers to recognize phenotypes," J Cell Sci, vol. 126, no. 24, pp. 5529-5539, 2013.

[7] M. Nilashi, O. Bin Ibrahim, A. Mardani, A. Ahani, and A. Jusoh, "A soft computing approach for diabetes disease classification," Health Informatics Journal, vol. 24, no. 4, pp. 379-393, 2018.

[8] S. Ozer *et al.*, "Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI," Medical physics, vol. 37, no. 4, pp. 1873-1883, 2010.

[9] Nilashi, M., *et al.*, A soft computing approach for diabetes disease classification. Health Informatics Journal, 2018. 24(4): p. 379-393.

[10] Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and systems magazine, 6(3), 21-45.

[11] Ö. Eskidere, F. Ertaş, and C. Hanilçi, "A comparison of regression methods for remote tracking of Parkinson's disease progression," Expert Systems with Applications, vol. 39, no. 5, pp. 5523-5528, 2012.

[12] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," Computer methods and programs in biomedicine, vol. 113, no. 3, pp. 904-913, 2014.

[13] T. Peterek, P. Dohnálek, P. Gajdoš, and M. Šmondrk, "Performance evaluation of Random Forest regression model in tracking Parkinson's disease progress," in

13th International Conference on Hybrid Intelligent Systems (HIS 2013), 2013: IEEE, pp. 83-87.
[14] W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen, "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," Biomedical Signal Processing and Control, vol. 8, no. 4, pp. 364-373, 2013.
[15] M. S. Islam, I. Parvez, H. Deng, and P. Goswami, "Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder (dysphonia)," in 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014: IEEE, pp. 1-7.
[16] V. Despotovic, T. Skovranek, and C. Schommer, "Speech Based Estimation of Parkinson's Disease Using Gaussian Processes and Automatic Relevance Determination," Neurocomputing, 2020.
[17] K. Polat, "A Hybrid Approach to Parkinson Disease Classification Using Speech Signal: The Combination of SMOTE and Random Forests," in 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 2019: IEEE, pp. 1-3.
[18] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Apaydin, H. (2019). A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 74, 255-263.
[19] Froelich, W., K. Wrobel, and P. Porwik, Diagnosis of Parkinson's disease using speech samples and threshold-based classification. Journal of Medical Imaging and Health Informatics, 2015. 5(6): p. 1358-1363.
[20] Logemann, J.A., *et al.*, Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and hearing Disorders, 1978. 43(1): p. 47-57.
[21] Parisi, L., N. RaviChandran, and M.L. Manaog, Feature-driven machine learning to improve early diagnosis of Parkinson's disease. Expert Systems with Applications, 2018. 110: p. 182-190.
[22] Xie, Y., Y. Liu, and Q. Fu, Imbalanced data sets classification based on SVM for sand-dust storm warning. Discrete Dynamics in Nature and Society, 2015. 2015.
[23] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
[24] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2004). Effective voting of heterogeneous classifiers. Paper presented at the European Conference on Machine Learning.
[25] Sikora, R. (2015). A modified stacking ensemble machine learning algorithm using genetic algorithms Handbook of Research on Organizational Transformations through Big Data Analytics (pp. 43-53): IGi Global.
[26] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264-1271, May 2012, doi: 10.1109/TBME.2012.2183367.
[27] B. Pittman, R. H. Ghomi and D. Si, "Parkinson's Disease Classification of mPower Walking Activity Participants," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 4253-4256, doi: 10.1109/EMBC.2018.8513409.
[28] S. Magotra and K. Kumar, "Detection of HELLO flood attack on LEACH protocol," in Proceedings of the IEEE International Advance Computing Conference (IACC '14), pp. 193–198, Gurgaon, India, 2014.
[29] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Trans. Biomedical Engineering 56(4), 1015-1022, 2009.
[30] S. Sapir, L. Ramig, J. Spielman, C. Fox, Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. Journal of Speech Language and Hearing Research, 53, 114-25, 2010.
[31] A. Tsanas, M.A. Little, P.E. McSharry, L.O. 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, 842-855, 2011.
[32] I.R. Titze. Principles of Voice Production. National Center for Voice and Speech, Iowa City, US, 2nd ed., 2000.
[33] Aida-zade K, Xocayev A,Rustamov S. Speech recognition usingsupport vector machines. In: AICT'16.10th IEEE International Conference on Application of Information and Communication Technologies; 2016
[34] Tirumala SS, Shahamiri SR,Garhwal AS, Wang R. Speaker identification features extraction methods: A systematic review. Expert Systems with Applications. 2017; 90:250-271
[35] V. Nigam, S. Jain, and K. Burse, "Profile based scheme against DDoS attack in WSN," in Proceedings of the 4th International Conference on Communication Systems and Network Technologies (CSNT '14), pp. 112–116, Bhopal, India, 2014.
[36] Krassovitskiy A, Mussabayev R. Energy-based centroid identification and cluster propagation with noise detection. In: Nguyen N, Pimenidis E, Khan Z, Trawiński B, editors. Computational Collective Intelligence. Lecture Notes in Computer Science. Vol. 11055. Cham: Springer; 2018. pp. 523-533. DOI: 10.1007/978-3- 319-98443-8_48. ICCCI 2018
[37] P. Yi, T. Zhu, Q. Zhang, Y. Wu, and J. H. Li, "A denial of service attack in advanced metering infrastructure network," in Proceedings of the IEEE International Conference on Communications (ICC '14), Sydney, Australia, June 2014.