

A machine learning approach to distinguish Parkinson's disease (PD) patient's with shuffling gait from older adults based on gait signals using 3D motion analysis

Satyabrata Aich^{1*}, Pyari Mohan Pradhan², Jinse Park³, Hee-Cheol Kim¹

¹ Department of Computer Engineering/IDA, Inje University, Gimhae, South Korea

² Department of Electronics and Communication Engineering, IIT Roorkee, India

³ Department of Neurology, INJE University College of Medicine, Busan, South Korea

*Corresponding author E-mail: satyabrataaich@gmail.com

Abstract

In recent times the adverse impact of Parkinson's disease (PD) getting worse and worse with the increasing rate of old age population through out the world. This disease is the second common neurological disorder and has a tremendous economical and social impact because the cost associated with the healthcare as well as service is extremely high. The diagnosis process of this disease mostly done by closely observing the patient in the clinic as well as using the rating scale. However, this kind of diagnosis is subjective in nature and usually takes long time and assessment of this disease is complicated and cannot replicated in other patients. This kind of diagnosis method is also not suitable for the early detection of the PD. So, with this shortcoming it is necessary to find a suitable method that can automate the process as well as useful in the initial phase of diagnosis of PD. Recently with the invention of motion capture equipment's and artificial intelligent technique, the feasibility of the objective nature-based diagnosis is getting lot of attention, especially the objective quantification of gait parameters. Shuffling of gait is one of the important characteristics of PD patients and it is usually defined by shorter stride length and low foot clearance. In this study a novel method is proposed to quantify the gait parameters using 3D motion captures and then various feature selection algorithm have used to select the effective features and finally machine learning based techniques were implemented to automate the classification process of two groups composed of PD patients as well as older adults. We have found maximum accuracy of 98.54 % by using support vector machine (SVM) classifier with radial basis function coupled with minimum redundancy and maximum relevance (MRMR) algorithm-based feature set. Our result showed that the proposed method can help the clinicians to distinguish PD patients from the older adults. This method helps to detect the PD at early stage.

Keywords: Shuffling Gait; Feature Selection; Machine Learning; Parkinson's disease; Wearable Sensor

1. Introduction

Parkinson's disease is the second most common neurological disorder and around 7 to 10 million people have been affected by this disease [1]. It is one of the disease that affects the brain cell and that indirectly affect the movement, speech and other cognitive parts of the brain [2]. The number of undiagnosed people affected with PD is around 20% [3]. At the same time, it is very difficult for the clinician to clinically distinguish the PD patients with the older adults because some of the characteristics are observed among old people those are also associated with PD [4]. Gait analysis is treated as one of the important tool for the assessment of PD [5]. It was observed that the spatiotemporal parameters related to gait dynamics helps in objective assessment of PD because those parameters are one dimensional and the analysis can be observed in different dimension [6]. The objective of the feature selection method in the medical field is to identify the most important factor that is associated with the result and also remove the redundant features that provides better and quicker result at

quick time [7]. Many researchers used different classifiers for prediction of disease to get good outputs while prediction coupled with good accuracy [8-10].

In this paper spatiotemporal parameter-based gait analysis is carried out using 3D motion captures for objective measurement of gait parameters and then feature selection algorithms are implemented to know the importance of the features. A performance comparison analysis is carried out with different classification techniques coupled with different feature sets obtained from various feature selection algorithm to clearly distinguish the PD patients with shuffling gait and older adults.

The structure of the paper is organized as follows: Section 2 presents the background. Section 3 describes about the methodology. Section 4 describes about the result and discussions. Section 5 describes about conclusion.

2. Related work

In the past different researchers have used wearable devices as well machine learning techniques for the prediction of the chronic diseases. Some of the works are as follows. An inertial measurement

units (IMU) which consists of gyroscope and accelerometer were used to do the gait analysis of healthy subjects and PD patients. The proposed algorithm was able to estimate the stride length from the data collected from the IMU. The results were validated by measuring the stride length using GAITrite walkway system data, which is mentioned as gold standard. The comparison has done with the algorithm and GAITrite system and it was found that the mean error over all the stride lengths were about 6 % for healthy subjects and around 10.3% for PD patients [11]. A novel method is proposed to measure the spatiotemporal gait parameters of children's those are suffered with cerebral palsy using wearable sensors. The signal processing algorithm is used to quantify the gait parameters from the data collected from the wearable devices and validated the result with optoelectronic system which is served as a gold standard. It was found that the wearable device-based method able to provide good result and having a good level of agreement between the two methods has given an assurance to the clinicians to recommend it in practical life [12]. A method is proposed that used ANN and SVM for distinguishing the walk pattern during the speed walking of PD patients. Three different parameters namely spatiotemporal, kinematic and kinetic were used to distinguish the pattern among PD patients. It was observed that spatiotemporal parameter used as feature vector contributed perfect classification using ANN and SVM classifier [13]. A method is proposed for automatic recognition of health problem using gait data with machine learning techniques. The method used neural network and k nearest neighbor classifier based on the gait data of people related to five different condition such as normal, with hemiplegic, with Parkinson's disease, with pain in the back and with pain in the leg. semantic features were used in their approach. It was found the classification accuracies ranging from 99% to 100 % by using the neural networks and k nearest neighbor classifier [14]. A method is proposed for the diagnosis of PD based on the characteristics features of a person's voice. The approach used decision tree-based classification approach using the threshold value. It also used mixed classification approach for the prediction of the Parkinson's disease using nonlinear classifiers and found classification accuracy of 90% [15]. The above past work motivated us to develop the current approach described in the later part of the paper.

3. Methodology

The study comprised of 40 individuals from both the groups. 20 patients from PD group with shuffling gait and 20 older adults mentioned as control group. 3D motion Analysis system is used to quantify the spatiotemporal gait parameters at Inje Paik hospital, Busan, South Korea. We have collected 12 features by measuring the spatiotemporal parameters. We have used four feature selection methods in this study such as minimum redundancy maximum relevance (MRMR), Fisher's score(FS) method, sequential forward selection(SFS) method and Principal Component Analysis (PCA) method. After feature selection methods each of feature sets is implemented in different classification technique such as support vector machine (SVM), random forest (RF), and Naïve Bayes (NB) for performance comparison. The performance measures those are included in this study are accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV). Fig 1 shows the flow chart of the proposed methodology.

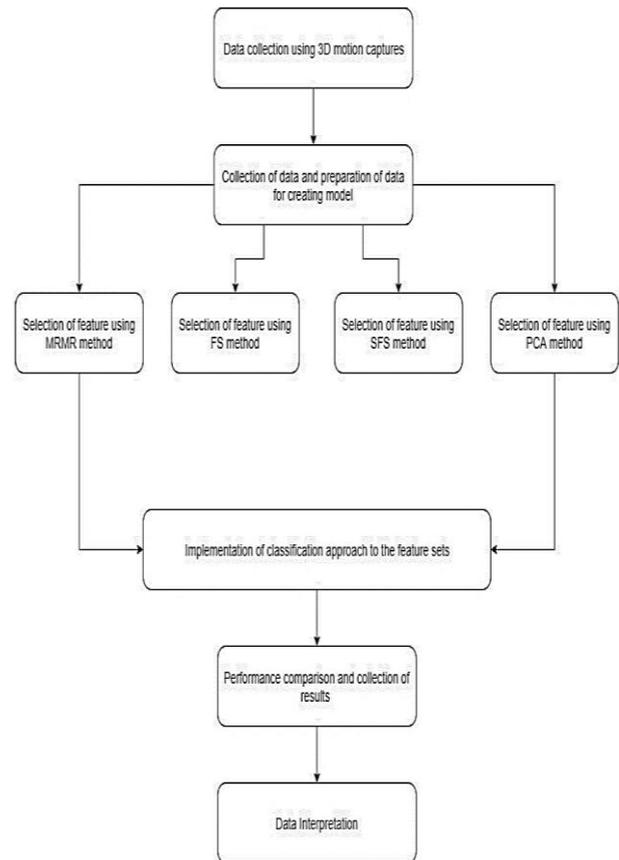


Fig. 1: Flow Chart of Proposed Methodology.

4. Results and discussion

In this study we have implemented different feature selection algorithm and we found MRMR feature selection algorithm performed well with different classifiers. We have used 80% of the data for training set and 20% of the data as test set. The comparison of accuracy with different feature sets are shown in the fig 2.

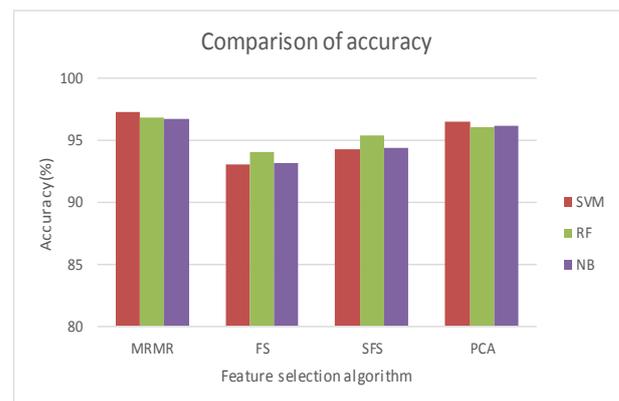


Fig. 2: Accuracy of the MRMR, FS, SFS and PCA Results.

In the above figure it is shown that the accuracy of SVM classifier with rbf kernel function coupled with MRMR feature sets has the highest accuracy of 98.54% followed random forest classifier and Naïve Bayes classifier with the accuracy of 96.89%, and 96.75% respectively. MRMR feature sets perform better compared to all other feature sets mentioned in our approach.

The comparison of sensitivity of different feature sets coupled with classifiers is shown in the figure 3.

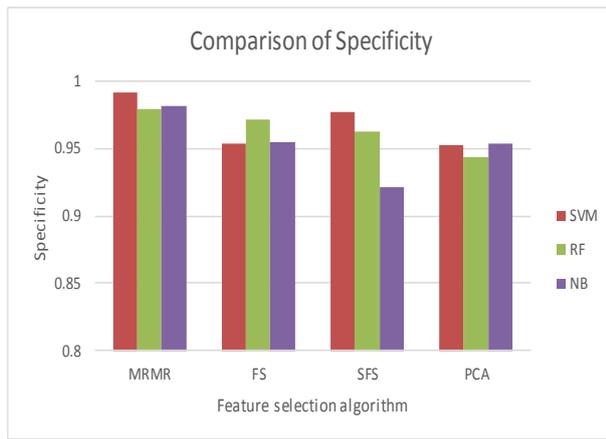


Fig. 3: Sensitivity of the MRMR, FS, SFS and PCA.

Results the figure 3 showed that the sensitivity of SVM classifier with rbf kernel function coupled with MRMR feature sets has the highest sensitivity of 0.9891 followed random forest classifier and Naïve Bayes classifier with the sensitivity of 0.9844 and 0.9722 respectively.

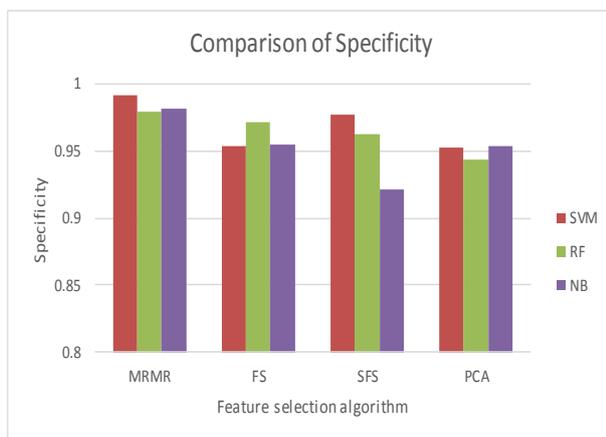


Fig. 4: Specificity of the MRMR, FS, SFS and PCA.

Results. the figure 4 showed that the specificity of SVM classifier with rbf kernel function coupled with MRMR feature sets has the highest specificity of 0.9912 followed Naïve Bayes classifier and random forest classifier with the specificity of 0.9811 and 0.9789 respectively.

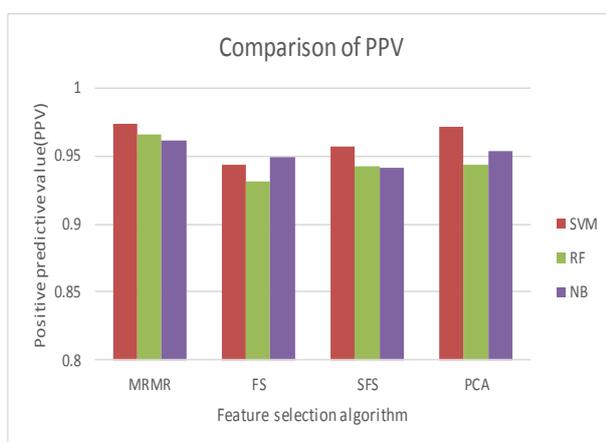


Fig. 5: PPV of the MRMR, FS, SFS and PCA.

Results. The figure 5 showed that the positive predictive value (PPV) of SVM classifier with rbf kernel function coupled with MRMR feature sets has the highest PPV of 0.9734 followed random forest classifier and Naïve Bayes classifier with the PPV of 0.9654 and 0.9611 respectively.

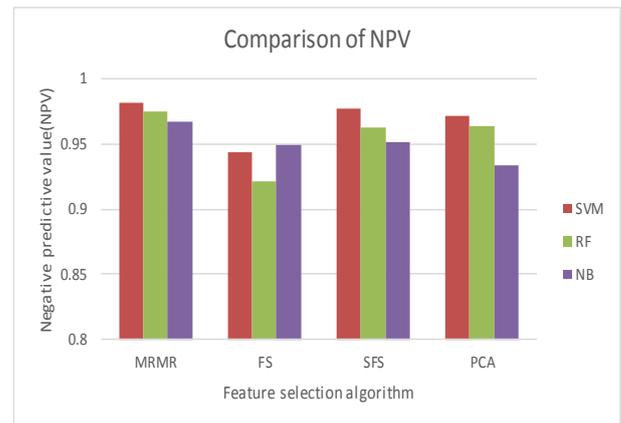


Fig. 6: NPV of the MRMR, FS, SFS and PCA.

Results. The figure 6 showed that the negative predictive value (NPV) of SVM classifier with rbf kernel function coupled with MRMR feature sets has the highest NPV of 0.9815 followed random forest classifier and Naïve Bayes classifier with the NPV of 0.9754 and 0.9667 respectively.

From the above observation it was found that feature selection method has an important role to play that ultimately provide better result at quick time. It also removed the redundant features that will help clinicians to focus on important features during diagnosis of PD. Our proposed method has carried out a performance comparison-based study by using different classification technique based on different feature sets. The performance measures help to identify the potential of each classifier as well as automatize the classification process that will help the clinicians to clearly distinguish between two groups. Our proposed study provides an accuracy of 98.54%, compared to the previous studies. One of the previous study used fuzzy c means clustering based feature weighting and KNN and found accuracy of 97.93% [16] and other study used particle swarm optimization and KNN and found accuracy of 97.47% [17]. In summary our proposed method has better accuracy and can be recommended for clinical trials.

5. Conclusion and future work

This paper discussed about different feature selection algorithm as well as different classification techniques as well as their performance for better accuracy as well as for getting quicker results. Among the implemented feature selection algorithm and classification techniques MRMR algorithm coupled with SVM with rbf produced an accuracy of 98.54%. This development will help the clinicians and doctors to automate the whole process with short time and will help them to detect the Parkinson's disease at early stage and assess the patient's condition at different stages.

In the future we will develop the algorithm to quantify the spatio-temporal gait parameters from the data collected from the wearable sensors and then implement the same feature selection algorithm as well as classification technique to compare the difference between the two methods as well as to check the feasibility of the wearable sensors for objective quantification of gait parameters.

References

- [1] M.Hermanns, "Weathering the storm: living with Parkinson's disease," *Journal of Christian Nursing*, vol. 28, no. 2, pp.76-82,2011.
- [2] J.M. Beitz, "Parkinson disease: a review" *Frontiers in Bioscience*, S6, pp.65-74, 2014.
- [3] A. Schrag, Y. Ben-Shlomo, N. Quinn, "How valid is the clinical diagnosis of Parkinson's disease in the community?," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 5, pp. 529-534, 2002.
- [4] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp.1568-1572, 2010.

- [5] P.H. Chen, R.L. Wang, D.J. Liou, J.S. Shaw, "Gait disorders in Parkinson's disease: assessment and management," *International Journal of Gerontology*, vol. 57, no. 4, pp.189-193, 2013.
- [6] S. Hou, L. Li, R. Bo, W. Wang, T. Wang, "Feature generation using recurrence quantification analysis with application to fault classification," In *System Science, Engineering Design and Manufacturing Informatization (ICSEM)*, 2011 International Conference on, Vol. 2, pp. 43-46, IEEE, October 2011.
- [7] D.Jain, V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informatics Journal*, 2018.
- [8] J. Han, J. Pei, M. Kamber, "Data mining: concepts and techniques," Elsevier, 2011.
- [9] P.N. Tan, "Introduction to data mining," Pearson Education India, 2006.
- [10] M.H. Dunham, "Data mining: introductory and advanced topics," Pearson Education India, 2006.
- [11] B. Sijobert, M. Benoussaad, J. Denys, R. Pissard-Gibollet, C. Geny, and C.A. Coste, "Implementation and Validation of a Stride Length Estimation Algorithm, Using a Single Basic Inertial Sensor on Healthy Subjects and Patients Suffering from Parkinson's Disease," *Health*, 7, pp.704-714, 2015.
- [12] L.Carcreff, C.N. Gerber, A. Paraschiv-Ionescu, G. De Coulon, C.J. Newman, S. Armand, and K. Aminian, "What is the Best Configuration of Wearable Sensors to Measure Spatiotemporal Gait Parameters in Children with Cerebral Palsy?," *Sensors* vol.18, no. 2, p.394, 2018.
- [13] N.M. Tahir, H.H. Manap, "Parkinson Disease Gait classification based on Machine Learning Approach," *Journal of Applied Sciences*, vol. 12, no. 2, pp. 180-185, 2012.
- [14] B. Pogorelc, Z. Bosnić, M. Gams, "Automatic recognition of gait-related health problems in the elderly using machine learning," *Multimedia Tools and Applications*, vol. 58, no.2, pp.333-354, 2012.
- [15] W. Froelich, K. Wrobel, and P. Porwik, "Diagnosis of Parkinson's disease using speech samples and threshold-based classification," *Journal of Medical Imaging and Health Informatics*, vol. 5, no.6, pp. 1358-1363, 2015.
- [16] K. Polat, "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering," *International Journal of Systems Science*, vol. 43, no. 4, pp.597-609, 2012.
- [17] W.L. Zuo, Z.Y. Wang, T. Liu, 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.