**Fall Risk Prediction in Parkinson’s Disease Using Real-World Inertial Sensor Gait Data**

Link for this journal: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9924545>

**Introduction:**

This paper is all about creating a machine learning model which analyzes the behavior of the patients suffering from Parkinson’s disease and tries to predict the fall risk possibility for each patient. The authors are going to perform an experimentation to study these patients.

As mentioned by the authors, there were many papers which already tried to solve this problem by different approaches, but apart from those this paper differentiates on the basis of several factors like the type of data being used and also the methodologies which are being used.

This study helped the authors to achieve interesting findings as they classified the test results on the basis of three different categories. And they predicted the findings based on these categories to decide which way of selecting the features impacted for better accuracy, sensitivity and specificity.

The results obtained from the prediction are a combination of different machine learning algorithms calculated on a regular interval of time. In terms of the values of the key metrics like sensitivity, specificity, accuracy etc., the results from the previous studies have strong values. But all those values are based on either unsupervised gait data or unconstrained real-world data calculated for several days. But for this experimentation the authors used both the data combined for the first time

**Technical Contributions:**

For this experiment, the authors tried to setup the process by involving 40 Parkinson’s disease affected patients and they tried to attach their required equipment with the patients before starting the experiment process. These sensors are foot worn IMU’s (Inertial Measurement Units). Out of them 5 patients were taken off from the experimentation as they did not contribute well.

The experiment was conducted for a time period of two weeks for gathering the data from the patients continuously. This process is followed up for three months. So, this experimentation tries to gather data from two methodologies 1) By making the patients follow 4x10 Meter Walking Tests daily morning, noon and evening collecting the gait data. 2) By unsupervised monitoring of the patients daily in the real-world scenario.

Both the data are used for extracting features and selecting appropriate machine learning models for the test. In this experimentation the authors tried to capture the feature parameters involved in this experiment by the method of data aggregation and divided the obtained datasets as three types 1) Bout wise 2) Daily wise 3) Participant wise.

They considered the process of feature selection with the help of Select-K-Best based and Recursive Feature Elimination (RFE). The filter of selecting the features was based on Fibonacci data collection and the models like SVM-lin, SVM-rbf, RF (Random Forest classifier) and GB (Gradient Boosting algorithm) are used for training and testing purposes for getting the metric values.

The Cross-validation technique used in the testing process was LOPO-CV (Leave one person out cross validation). They observed that the Participant wise data aggregation method gave the best results as the features had best correlation for analysis compared to other two approaches.

With all these methods of approach they were able to get the astounding result of 74% accuracy with 60% sensitivity and 88% specificity.

**Improvisations:**

This method is very practical compared to the previous papers as mentioned by the author as no one worked with both real time data and gait data both at a time. But this method has several disadvantages as mentioned by the author himself in a statement.

The data collection period needs to be improved in a more longitudinal kind of approach so that the models can be using deep learning technology. The accuracy given by the test result is lower compared to the one mentioned by the other authors.

Coming to one of the improvisations according to me is to use the approach longitudinal feature selection instead of Fibonacci approach as the data is mostly based on time series analysis. This may help in understanding the features in depth resulting in better correlation among the parameters.

When dealing with time series data it is always important to consider the cross validation used must satisfy time series approach, coming to this experiment according to my analysis the LOPO-CV may not be perfect for analysis instead the author may have approached with some of the methods like Time-Series split Cross-validation which may be best for temporal data.

Now coming to the approach of feature selection, it may be best to also consider the methodology of Frequency-Domain Feature Extraction technique which is basically based on the approach of identifying the frequency of the classes for better correlation.