Data Mining
Parkinson’s Disease
Freezing of Gait

University of Delhi

Department of Computer Science

Batch(2023 - 25)

Guidance of:
Dr.Bharti Rana

Team - Members
Prachi Goel
Deepak Kumar
Harsh Yadav
Agam Gupta

Introduction

Parkinson's disease is a degenerative, long-term neurological condition that greatly dopaminergic neurons in the brain which impairs coordination and mobility. This disorder affects the neurological system, causing tremors, stiffness, and a variety of mobility issues.

Parkinson's disease patients can have tremors in their hands, arms, legs, jaw, or brain, among other symptoms. Common symptoms of the illness include muscle stiffness, slowness of movement, imbalance, and coordination problems. Parkinson's disease can affect a person's mental health in addition to their physical health, which can lead to symptoms like depression and other emotional disturbances. Furthermore, swallowing, chewing, and speaking difficulties are among the everyday activities that people with Parkinson's disease may experience. A wide range of symptoms, including those connected to the skin and urinary system, are linked to this intricate neurological condition.

Parkinson's Disease Gait Analysis is crucial for understanding ***Freezing of Gait*** **(FOG)** episodes. FOG is a common and debilitating symptom in Parkinson's patients, characterized by brief, episodic inability to initiate or continue movement. This report aims to unravel the enigma of FOG and explore the role of gait analysis in its understanding.

1.1. Gait Features

Human gait is a sequence of involuntary movements, cyclically repeated and triggered by voluntary movement. Several components could be used to objectively measure and analyze gait cycle. These components are typically categorized into spatiotemporal, kinematics and kinetics features.

1.2. Spatiotemporal Features

Several spatiotemporal features can be used for gait analysis (Table 1). These features are the more commonly used types of features to objectively describe the gait pattern in healthy subjects and patients with several diseases. Spatiotemporal features could refer to the global gait cycle or to the stride cycle.

Related-Work

[1] *Gait Analysis in Parkinson's Disease: An Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring*

- ## Introduction:

Parkinson's disease (PD) is a dynamic disease that requires continuous adjustment of therapy, with a diagnostic error rate of around 20%. The gold standard for diagnosis and symptoms monitoring in PD is based on clinical evaluation, which includes subjective components. Gait impairment is an evolving condition, with different patterns of gait disturbances detected throughout the disease course. An objective and quantitative gait analysis system could potentially improve current practice in diagnosing, symptom monitoring, therapy management, rehabilitation, and fall risk assessment and prevention in Parkinson's disease patients.

- ## Technology Used:

The researchers used various technologies, including wireless inertial sensors placed on the foot in PD patients and healthy subjects. They detected physical kinematic features of pitch, roll, and yaw rotations of the foot during walking and used principal component analysis (PCA) to select the best features. The SVM method was used to classify PD patients, with and without gait impairment, and healthy subjects.

The selected features include pitch, roll, and yaw features, which have very high sensitivity, specificity, and positive predict values.

The study also included decision trees, neural networks, linear discriminant analysis, random forest, Bayesian probability, SVM, decision tree, and tensor decomposition. The results showed that the proposed classification has very high sensitivity, specificity, and positive predict values (93.3%, 95.8%, and 97.7% respectively) to distinguish PD patients from healthy subjects.

- **Conclusion:**

The proposed classification system for Parkinson's disease (PD) patients compared to healthy subjects has high sensitivity, specificity, and positive predict values (93.3%, 95.8%, and 97.7% respectively). The classification features include pitch, roll, yaw, decision tree, neural network, LDA, random forest, Bayesian probability, SVM, decision tree, kNN, and tensor decomposition. The discrimination features include pitch, dorsiflexion, roll, yaw, distance between elbow and hip, average angle, knee and knee angles, height, speed, and angle between shoulder and wrist.

[2] *Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network*

- **Objective:**

The objective of the research article is to design and investigate a gait analysis based classifier model using a hybrid CNN–LSTM network to predict the severity rating of PD. PD is a progressive neurodegenerative disorder that affects the motor system and causes various symptoms such as tremor, rigidity, bradykinesia, and postural instability. One of the most common and early signs of PD is gait impairment, which can be measured by the VGRF signals. The VGRF signals reflect the interaction between the foot and the ground during walking and can capture the spatiotemporal features of gait. The authors propose to use EMD to decompose the VGRF signals into IMFs and select the dominant IMF based on power spectral analysis. The dominant IMF is then fed into the CNN–LSTM network to learn the spatiotemporal features of gait and classify the PD stages according to H&Y scale. The H&Y scale is a widely used clinical rating scale that assesses the severity of PD based on five stages. The authors claim that their proposed model can achieve a high accuracy of 98.32% in multi-class classification of PD severity and outperform several existing methods that have used gait pattern to diagnose PD. The authors also highlight the advantages of their data-driven and adaptive approach, which does not require any prior knowledge or assumptions about the gait signals. The authors hope that their proposed model can assist the physicians to diagnose PD effectively and accurately by exploiting the gait impairment as one of the early and important symptoms of PD.

- **Methods:**

  The proposed model uses EMD to decompose the VGRF signals from three walking tests into IMFs and extracts the dominant IMFs based on power spectral analysis1. The IMFs are then fed into the CNN–LSTM network to learn the spatiotemporal features of gait and classify the PD stages according to H&Y scale.

Adam optimization algorithm: A method to minimize the loss function of the CNN–LSTM classifier model. It uses adaptive learning rates and first and second moments of gradient to update the weights iteratively. It has some hyper-parameters such as learning rate, exponential decay rates, threshold and constant. It requires less memory and minimal tuning1. Algorithm shows the pseudo code of the Adam optimization algorithm.

- **Results:**

Experimental results and discussion: A section that describes the data, methods, metrics, and analysis of the proposed framework. It uses the VGRF time series data from Physionet database for three walking tests3. It divides the data into 80% for training and 20% for testing4. It applies EMD technique to decompose the VGRF signals into IMFs and selects the second IMF as the dominant one5. It uses the CNN–LSTM network with four convolutional and max-pooling layers, two LSTM and dropout layers, and a fully connected layer with softmax activation. It trains the network for 25 epochs and uses the Adam optimizer with the specified hyper-parameter settings. It evaluates the performance using accuracy, sensitivity, specificity, PPV, F-score, and MCC. It compares the results with the baseline CNN model and other existing methods. It shows that the proposed framework achieves the highest accuracy of 98.32% and outperforms the state-of-the-art techniques.

[3] *Freezing of gait in Parkinson's disease: Classification using computational intelligence*

- ## Introduction:

Parkinson's disease is a progressive disorder that affects the nervous system and the parts of the body controlled by the nerves. The most prominent signs and symptoms of Parkinson's disease occur when nerve cells in the basal ganglia, an area of the brain that controls movement, become impaired and/or die. Normally, these nerve cells, or neurons, produce an important brain chemical known as dopamine. When the neurons die or become impaired, they produce less dopamine, which causes the movement problems associated with the disease. Scientists still do not know what causes the neurons to die. This research paper tried to solve two problems:
Problem1 – To distinguish PD patients from Healthy individuals.
Problem2 – Differentiating between patients with PD and FOG in the off and on medication states.
Seven different Classifiers were used : KNN, Decision Tree, Linear SVM, Random Forest, Naïve Bayes, Multilayer Perceptron, Quadratic Discriminant Analysis.

- ## Methodology:

Using the Tsfresh Library, 200 features were extracted. These include, for example, absolute energy, absolute maximum, and absolute sum of changes. All the extracted features were stored in CSV files. At this step, all the Not A Number (NaN) records were removed. In this experiment, 9 individuals from each of these groups were selected. Each individual carried out the same task three times at different time intervals to ensure result validity. This results in 27 healthy subjects and 27 PD subjects for each task. The initial 67% of the data was employed for training, while the remaining 33% was designated for validation.
The training dataset was further divided into five distinct subsets for cross validation, to avoid overfitting, and also to ensure the consistency of the accuracy result.

- **Technology:**

For data collection, inertial sensors were used to collect gait and episode information of FOG were coupled to three smartwatches, with a Movement Disorders Monitoring System (NetMD), in order to analyze and remotely and continuously monitor movement disturbances through inertial signals. This NetMD used the joint action of an Android mobile with smartwatch devices with communication being established via Bluetooth. The system used to generate a text file with 10 columns containing the values of the inertial signals for each smartwatch (time in milliseconds, sensor name, battery status, and accelerometer and gyroscope on the x, y and z axes).

- **Results:**

While Classifying PD patients with healthy ones, the majority of the classifiers displayed satisfactory results, but the MLP outperformed others, achieving a 96% F1-score, 96% accuracy, and 99% recall.
While checking for effectiveness of medication, because it had a score of 99% in all four categories of accuracy, F1-score, precision, and recall, the Decision Tree was considered the most effective method of categorization for this issue.

- **Conclusion:**

A patient-wise analysis may give an opportunity to further understand the patient's symptoms, leading to more effective treatment strategies in the future. Furthermore, feature importance can be calculated to better understand gait impact in Parkinson's disease. This can be achieved through methods such as feature selection, dimensionality reduction, and model-based methods. The results can provide valuable insight into the most effective features for analysis and the underlying biological mechanisms of the disease.

Methodology

Our Dataset comprises lower-back 3D accelerometer data from subjects exhibiting **freezing of gait** episodes, a disabling symptom that is common among people with Parkinson's disease. Freezing of gait (FOG) negatively impacts walking abilities and impinges locomotion and independence.

Your objective is to detect the start and stop of each freezing episode and the occurrence in these series of three types of freezing of gait events: **Start Hesitation**, **Turn**, and **Walking**.

The Datasets:

The data series include three datasets, collected under distinct circumstances:

- The **tDCS FOG** (tdcsfog) dataset, comprising data series collected in the lab, as subjects completed a FOG-provoking protocol.
- The **DeFOG** (defog) dataset, comprising data series collected in the subject's home, as subjects completed a FOG-provoking protocol.

Attribute-Information:

- **Time** An integer timestep. Series from the **tdcsfog** dataset are recorded at 128Hz (128 timesteps per second), while series from the **defog** and **daily** series are recorded at 100Hz (100 timesteps per second).
- **AccV**, **AccML**, and **AccAP** Acceleration from a lower-back sensor on three axes: **V - vertical**, **ML - mediolateral**, **AP - anteroposterior**. Data is in units of *m/s^2* for **tdcsfog/** and *g* for **defog/** and **notype/**.
- **StartHesitation**, **Turn**, **Walking** Indicator variables for the occurrence of each of the event types.
- **Event** Indicator variable for the occurrence of *any* FOG-type event.

- **Valid** There were cases during the video annotation that were hard for the annotator to decide if there was an Akinetic (i.e., essentially no movement) FoG or the subject stopped voluntarily. Only event annotations where the series is marked true should be considered as unambiguous.
- **Task** Series were only annotated where this value is true. Portions marked false should be considered unannotated.

To treat the **Imbalance Data**, we had two choices:

1. Either to do **down-sampling**, and take samples of all classes equal to of minority class(i.e 500).
2. The second option we had, of generating synthetic data and leveraging the **SMOTE** technique, in result up-sampling the data.

We tried both possibility, and we found out the *combination of Smoting and downsampling*, as better approach.

Experimental Results

In the experimental phase of the project, we trained and tested the results on 6 classifiers.

Classifiers:

1. Decision Tree
2. Random Forest
3. Naive Bayes
4. K-NN
5. XGBoost
6. LGBM

The Decision Tree classification model was chosen for DeFOG dataset, achieving an impressive accuracy of 95.15%. The model's performance was evaluated using a confusion matrix and further detailed through a classification report, highlighting precision, recall, and F1-score metrics for each class. The precision, recall, and F1-score for each class were as follows:

- Class 0: Precision of 99%, Recall of 95%, and F1-score of 97%
- Class 1: Precision of 92%, Recall of 97%, and F1-score of 94%
- Class 2: Precision of 79%, Recall of 94%, and F1-score of 86%
- Class 3: Precision of 87%, Recall of 95%, and F1-score of 91%

Here **0** Signifies, the person is normal and doesn't seem to have FoG, whereas, **1** signifies, FoG occured with hesitation while patient started to walk, **2** signifies the occurence of FoG while the patient took a Turn, and **3** signifies that FoG triggered while the patient was normally walking.

Keep in mind, only a single sensor data does not prove that the specific person is FoG+ or not, but the collection of predictions made by the model on the series of data does.

We found out that in **defog** {home dataset}, the **Decision tree** was giving us the best results, *that too by applying smote only*.The overall weighted average precision, recall, and F1-score were reported as **96%, 95%,** and **95%**, respectively, across all classes. The macro average precision, recall, and F1-score were calculated as **89%, 95%,** and **92%**, providing a comprehensive overview of the model's performance across multiple classes.

Same process we repeated for **tdcsFOG** (lab dataset), but surprisingly the only smote did not work for this dataset, *The combination of two gave the best results*, that too with **Random Forest Classifier**. The overall weighted average precision, recall, and F1-score were reported as **86%, 86%,** and **86%**, respectively, across all classes. The macro average precision, recall, and F1-score were calculated as **86%.**

Conclusion and Future Work

In summary, the classification model demonstrated an impressive accuracy of 95.15% for deFOG dataset and 86.03% for tdcsFoG dataset, showcasing its effectiveness in accurately predicting various classes. The detailed classification report further illustrates the model's precision, recall, and F1-score metrics, providing valuable insights into its performance across different categories.

Although the model demonstrated a high level of skill in detecting FOG events, continuous improvement is necessary to handle false positives and negatives. For a more nuanced understanding and successful management of FOG, future research should concentrate on advanced sensor technologies, multimodal data fusion, real-time interventions, and longitudinal studies. These efforts will improve patient-centric care and contribute to the changing field of Parkinson's disease monitoring and intervention.

The development of advanced sensor technologies, such as wearable devices and machine learning models, is being explored for detecting and managing Parkinson's disease (FOG). These technologies can capture detailed data on gait abnormalities, such as stride length, cadence, and joint angles. Additionally, multimodal data fusion can be used to improve the accuracy of FOG detection. Real-time monitoring systems and assistive technologies can provide immediate feedback to patients. Longitudinal studies can identify patterns of FOG progression and enable early intervention. Patient-centric approaches, such as user-adaptive systems, telemedicine solutions, and interactive interfaces, can also be developed. Collaboration with healthcare professionals and patient education programs can ensure successful implementation of these technologies.

References/ Bibliography

[1] *Gait Analysis in Parkinson's Disease: An Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring*
Lazzaro di Biase 1, Alessandro Di Santo 1, Maria Letizia Caminiti 1, Alfredo De Liso 1, Syed Ahmar Shah 2, Lorenzo Ricci 1, Vincenzo Di Lazzaro

[2] *Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network*
B. Vidya, Sasikumar P.

[3] *Freezing of gait in Parkinson's disease: Classification using computational intelligence*
Omid Mohamad Beigi 1, Lígia Reis Nóbrega 2, Sheridan Houghten 3, Adriano Alves Pereira 2, Adriano de Oliveira Andrade 2

[4] @misc{tlvmc-parkinsons-freezing-gait-prediction,
author = {Addison Howard, amit salomon, eran gazit, HCL-Jevster, Jeff Hausdorff, Leslie Kirsch, Maggie, Pieter Ginis, Ryan Holbrook, Yasir F Karim},
title = {Parkinson's Freezing of Gait Prediction},
publisher = {Kaggle},
year = {2023},
url = {https://kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction}
}