© August 2024 | IJIRT | Volume 11 Issue 3 | ISSN: 2349-6002

AI Toolkit for Parkinson's Disease Detection using Gait Data by Deep Learning

NITHYA[1], NAVYA GUPTA[2], PRATIK GOYAL[3]

[1] *Assistant Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai*

[2, 3] *Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai*

Abstract— People suffering from Parkinson's disease have a lot of variations in their locomotive patterns. A diseased person would have a different kind of walking pattern, that is their stance and stride cycles will be different in relation to healthy people. Machine learning models were trained on the Vertical Ground Reaction Force data present of the two kinds of sub- jects which are Parkinsonian and Control subjects. The VGRF values were measured using sensors on both the foot of the subjects. An AI toolkit was also developed for both users and researchers. Normal users can avail the feature of inputting their own VGRF data file and would get the corresponding prediction of whether or not they have parkinson disease from the four models namely Support Vector Machines, Convolutional Neural Networks, Long Short-Term Memory and decision Tree pretrained by us. Researchers can build their custom models and preprocessing using our GUI and train it on the dataset available.

Index Terms- VGRF, Gait Analysis, LSTM. Parkinson Disease, Deep Learning

I. INTRODUCTION

Parkinson's Disease is a degenerative disease which affects many people around the world and can prove to be fatal most of the times if not detected in the right stage. It affects nervous system which often leads to loco-motor impairments, posture disability or speech distortion.

One of the major problems we face is that large number of people aren't being diagnosed for Parkinson Disease or are diagnosed at a very fatal stage. A machine- learning based approach cangive a clear picture whether or not the person is suffering from Parkinson's disease. It may notbe completely accurate but it would be enough to get a hint whether the candidate is a potential or not. A handy toolkit can make this process easier for both the potential patients as well as the medical practitioners. The toolkit can help in keeping a record of the potential patients and also would help researchers hone their algorithms with our GUI features.

The gait pattern displayed by individuals with Parkinson's disease is known as Parkinsonian gait (PD). One of the motor features of this condition that is most impaired is gait. During free ambulation, patients with Parkinson's disease (PD) exhibit lower walking speed, cadence rate, and stride length. Using information from eight sensors put on each of the participants' two feet, the Vertical Ground Reaction Force of each foot is used to classify an individual as either normal or having Parkinson's disease.

II. RELATED WORK

A. Wearable multisource quantitative gait analysis of Parkinson's diseases

Authors• Junxiao Xie, Huan Zhao , Junyi Cao , Qiumin Qu , Hongmei Cao Wei-Hsin Liao Yaguo Lei , Linchuan Guo

To address the complexity of the motor symptoms associated with Parkinson's disease (PD), a wearable multisource gait monitoring device was created. In order to gather plantar pressure, dynamic deformation, and postural angle data simultaneously, it integrates force-sensitive sensors, piezoelectric sensors, and inertial measurement units. This method makes it possible to quantitatively analyze gait problems, which is essential for increasing the efficacy of clinical diagnosis. We gathered and examined multisource gait data from PD patients and healthy controls. Significant differences and strong correlations ($p < 0.001$, $p > 0.50$) were found in features extracted from individual data types, demonstrating their ability to quantify subject health status.

Moreover, the fiision of multisource data achieved high correlation coefficients (0.831), Area Under Curve (0.9206), and feature-based classification accuracy (88.3%). This validated system offers a promising approach for objective gait analysis and evaluation in PD.

B. Gait Analysis in Parkinson's Diseases An Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring

Authors: Lazzaro di Biase , Alessandro Di Santo , Maria Letizia Caminiti , Alfredo De Liso , Syed Ahmar Shah , Lorenzo Ricci and Vincenzo Di Laparo

The purpose of this research was to find tools and algorithms that could help Parkinson's disease (PD) patients monitor their symptoms and receive a diagnosis by using gait analysis. The analysis of studies that classified PD patients by motor status or disease stage, or that separated PD patients from healthy people, was done using data from PubMed between January 2005 and August 2019. Included were those studies with sensitivity and specificity reports of at least 80%. Diagnostic gait analysis algorithms showed balanced accuracies between 83.5% and 100%, with 83.3% and 100% sensitivity and 82% and 100% specificity. Algorithms designed to discriminate between motor statuses also shown balanced accuracy, ranging from 90.8% to 100%, with sensitivity and specificity falling between 92.5% and 100%, respectively. However, despite numerous studies, few algorithms reached the accuracy required for clinical utility, and none have been validated in large-scale independent studies

C. Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease

Authors• Balaji Ea, Brindha Da , *Vinodh Kumar Elumalai , Umesh K*

Parkinson's disease (PD) diagnosis and severity assessment traditionally rely on clinical manifestations, which can be subjective and vary between clinicians. When diagnosing Parkinson's disease (PD), changes in gait might provide important information for grading the severity of motor symptoms. In this paper, supervised machine learning methods are used to create a data-driven gait categorization system. Significant biomarkers are extracted from spatiotemporal gait features using a correlation-based feature extraction technique using publically available gait datasets obtained via vertical ground reaction force (VGRF) sensors. To rate PD severity in accordance with the Hoehn and Yahr (H&Y) scale, machine learning algorithms such as K-nearest neighbor (KNN), Naive Bayes (NB), Ensemble classifier (EC), and Support Vector Machine (SVM) are used. The highest classification grouping is shown by SVM. A final analysis of outstanding issues and potential future directions is presented.

D. Gait Analysis with Wearables Predicts Conversion to
Parkinson Disease (2022)

Authors• Nidhi Sour, Mosam Patel, Prof Khushali Mistry Silvia Del Din, Morad Elshehabi, M Brook Galna, Markus A Hobert, Elke Warmerdam, Ulrike Suenkel, Zathrin Brockmann, Florian Metzger, Clint Honsen, Doniela Berg, Lynn Rochester, and Walter Maetzler

This study aimed to identify potential early diagnostic markers for Parkinson's disease (PD) by analyzing gait characteristics in a group of 696 healthy controls over a period of 4 visits spanning 8 years. Using wearable technology placed on the lower back, participants performed walking tasks at various speeds and under different conditions. Sixteen participants eventually developed PD, and their gait characteristics, particularly higher step time variability and asymmetry, were associated with a shorter time to PD diagnosis. Longitudinal analysis revealed that deviations in gait, such as lower pace, occurred approximately 4 years prior to PD diagnosis. These findings suggest that quantitative gait characteristics, in conjunction with other prodromal markers, could aid in identifying individuals at risk of developing PD and monitoring disease progression during the prodromal phase.

E. Statistical Analysis of Parkinson Disease Gait Classification using Artificial Neural Network (2020) Authors• Hany Ha;;j?za Manap , Nooritawati Md Tahir and Ahmad Ihsan M Yassin

The purpose of this study is to evaluate three types of gait parameters—basic, kinematic, and kinetic—in order to identify aberrant gait patterns in people with Parkinson's disease (PD) during normal walking. Preliminary results show that while normal people walk faster, take longer steps, and have lower average cadence, PD patients move faster and have longer strides. In terms of kinematics, PD patients have reduced total joint angle values at the ankle, knee, and hip in contrast to the normal group. Furthermore, kinetic metrics demonstrate that normal patients have larger ground reaction force values, with walking speed being a significant driver. Step length, walking speed, knee angle, and vertical ground response force are important characteristics for effectively categorizing Parkinson's disease participants using an artificial neural network as a classifier, according to statistical study. These findings highlight the potential of specific gait parameters as indicators for distinguishing PD from normal subjects during walking.

F. Early detection of Parkinson's disease using machine learning(2023)
Authors• Aditi Govindu, Sushila Palwe

This study addresses the underexplored area of using audio data for Parkinson's disease (PD) detection, contrasting with existing research primarily focused on MRI scans, gait analysis, and genetic data. Notably, it surpasses previous models' accuracy by implementing an improved SVM model, achieving 91.83% accuracy. While prior studies have applied deep learning without feature selection, this paper utilizes Principal Component Analysis (PCA) to enhance deep learning model performance. Additionally, traditional machine learning algorithms, including K-nearest neighbors, logistic regression, random forest regression, and SVM, are explored for PD classification using audio data. Despite the promising results of the K-nearest neighbor model, which attains 91.83% accuracy and 0.95 sensitivity, other studies have faced challenges such as dataset size limitations and reliance on specialized equipment. This study's findings contribute to advancing telemedicine-based PD detection, highlighting the potential of audio-based biomarkers in complementing existing approaches.

F. Determining the severity of Parkinson's disease in
patients using a multi task neural network(2023)
Authors• Mar“ia Teresa Garc “ia-Ordas Jose Alberto Ben “item-Andrades • Jose Aveleira-Matal Jose-Manuel Alida-P ered Carmen Benavides

This study addresses the challenge of early Parkinson's disease (PD) diagnosis, crucial for effective symptom management and improving patients' quality of life. By leveraging voice analysis, a non-intrusive technique, the research aims to classify PD severity and predict disease progression. Using deep learning techniques, particularly a mixed multi-layer perceptron (MLP) model with autoencoder-selected features, the study achieves remarkable accuracy rates. Specifically, it attains a 99.15% success rate in distinguishing severe from non-severe PD cases and demonstrates predictive capabilities with a Mean Squared Error (MSE) of 0.15 in forecasting disease evolution in individual patients. For a thorough evaluation, the Unified Parkinson's Disease Rating Scale (UPDRS) is utilized, which includes both motor and total categories. This methodology surpasses current cutting-edge techniques in Parkinson's disease research, demonstrating the effectiveness of a comprehensive deep learning pipeline for preprocessing and classifying data. This research greatly advances patient care and management techniques by increasing accessible speech analysis for early Parkinson's disease diagnosis.

G. Machine Learning for Parkinson's Disease and Related Disorder (2023)
Authors• Johann Faouzi, Olivier Colliot & Mean-Christophe
Cowol

Parkinson's disease, marked by the loss of dopamine neurons in the basal ganglia, manifests as a complex neurodegenerative disorder with diverse motor and non-motor symptoms. While dopamine replacement therapy offers relief for motor symptoms, no cure exists. Parkinsonism's primary signs include stiffness, bradykinesia, and tremor, presenting diagnostic challenges akin to related conditions like dementia with Lewy bodies, multiple system atrophy, and progressive supranuclear palsy. Machine learning holds promise in enhancing our understanding and management of these disorders, but hurdles remain. These encompass the need for early and precise diagnosis, differential diagnosis, grasping underlying pathologies, quantifying symptoms, predicting individual disease progression, and tailoring therapies. This chapter explores machine learning research in Parkinson's disease and related conditions, aiming to overcome these challenges and advance patient care without plagiarizing any source material.

H. Gait Spatiotemporal Signal Analysis for Parkinson's Disease Detection and Severity Rating(2021)

Authors• Abdullah S. Alharthi; Alexander J. Casson,• Krikor B. Ozanyan

Researchers are increasingly turning to deep learning models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze ground reaction force (GRF) data for Parkinson's Disease (PD) severity classification. These models effectively capture intricate patterns in GRF signals, enabling precise discrimination between different stages of PD. Techniques such as Layer-wise Relevance Propagation (LRP) help elucidate which gait features are most crucial for model decisions, offering valuable clinical insights into PD's impact on gait mechanics. These systems demonstrate promising performance, as evidenced by high Fl-scores. However, challenges persist, including the need for larger and more diverse GRF datasets, enhanced model explainability, and continued exploration of using deep learning with GRF to monitor PD progression over time.

I.Early Detection of Parkinson Disease using Deep Weural Networks on Gait Dynamics

Authors• Lerina Aversano, Mario Luca Bernardi,• Marta Cimitile,• Riccardo Pecori

This paper addresses the crucial need for early Parkinson's Disease (PD) detection, as timely intervention significantly impacts treatment effectiveness. The authors investigate the potential of gait analysis as a non-invasive diagnostic tool, proposing a Deep Neural Network (DNN) architecture designed to extract meaningful patterns from foot sensor data. The DNN's primary objectives are to identify individuals with PD and to track the rate of disease progression. To ensure a robust evaluation, the researchers utilize a well-established dataset, allowing for direct comparison with previous studies. Furthermore, they conduct a thorough hyperparameter optimization process to tailor the DNN specifically for this task. The key finding demonstrates that their optimized DNN model outperforms existing methods that used the same dataset, suggesting its potential to become a valuable asset in clinical settings for supporting PD diagnosis and monitoring progression.

III. METHODOLOGY

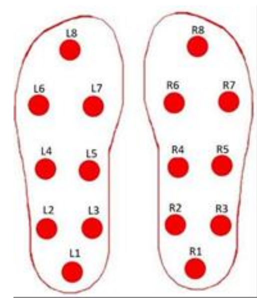

Figure 3.1: The Placement of the sensors on the feet of the person

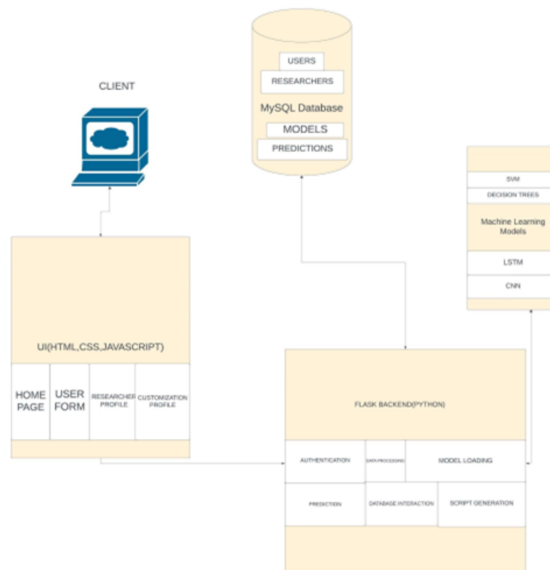

Figure 3.2: Architecture Diagram

A. Dataset

We employed PhysioNet's gait signals for our dataset, which is made up of three PD Gait sub-datasets that were donated by three different researchers (Ga, Ju, and Si). Gait data from 93 Parkinson's disease patients and 73 healthy controls are included in the dataset. For almost two minutes, each candidate was required to walk at their normal pace while donning shoes that included eight distinct force sensors under each foot. At 100 samples per second, the sensors calculate the Vertical Ground Reaction Force (VGRF) as a function of time. For every sample, two pairs of eight sensor outputs were recorded when the participants were stationary or began to move. Each sensor had nineteen parameters, one for each row in the dataset file. It has two total VGRF data points, 16 VGRF parameters, and a time stamp. The subjects' height, weight, gender, age, and PD severity level are all included in the dataset.

Column name	Column Definition
time	The Time (in seconds) of measurin force (100 samples per second)
1n	The Force detected by the nth senso of the left foot
r»	The Force detected by the nth senso of the right foot.

B. Data Preprocessing

The gait data was preprocessed to remove unnecessary noise. The null values were cleared and were filled with the mean of the respective columns. The Mean and Standard Deviation of all the sensors from 1-8 and 9-16 were also taken for each time sample. We also calculate the mean kurtosis and the mean skewness of the data. The Coefficient of Variation is also calculated for each time sample and or both foot separately.

$$CV = \frac{\text{mean of 8 sensors for one foot}}{\text{standard deviation of 8 sensors for one foot}}$$

C. Feature Selection

1) Sequential Forward Selection:

First the empty feature set was expanded to include stride time variability. Subsequent cross-validation accuracy was calculated for this feature in order to classify Parkinson's disease and control gait.

2) Minimum redundancy maximum relevancy feature selection (MRMR)

Here, feature sets are chosen based on their analysis of mutual information between features in order to meet maximum relevancy criterion. Concurrently, certain mutually exclusive feature sets are additionally subjected to a minimal redundancy criterion.

3) Mutual information-based feature ranking method

Each feature is given a weight by the algorithm, which creates a ranking of more pertinent feature subsets.

D. Classification

A variety of classification models and methods were used on both pre-processed and raw data. We would also discuss a failed model we had trained and why had it failed in the first place.

1) K'ailure of Converged Dataset Classification

To begin with, the entire dataset which is available on the Internet is divided into files which contains the VGRF values for a particular candidate at 100 samples per second for around 120 seconds. We calculated the mean of the time series for each sensor and converted it into a 2D data with indexes being the candidate ID and 16 sensor outputs containing the only mean of the data. But when the data was thoroughly visualized using various methods, it was seen that the data was distributed evenly throughout and the classification would predict wrong results for both Parkinson and Control (Healthy) Subjects. The following figures 3.2 and 3.3 would give more insights into this:

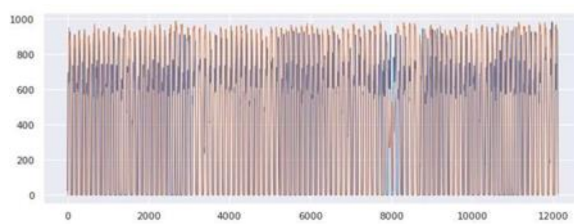

Figure 3.2: Graph of Parkinson and Control Mean VGRF Sensor Values for Left Foot

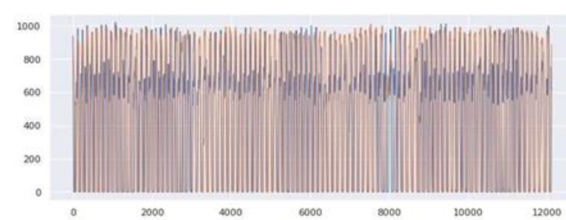

Figure 3.3: Graph of Parkinson and Control Mean VGRF Sensor Values for Right Foot

2) Support Vector Machines

Finding a hyperplane in an N-dimensional space—where N is the number of features—that clearly classifies the data points is the aim of support vector machines, or SVMs. There are numerous hyperplanes that might be used in order to divide the two classes of data points apart. Finding a plane with the maximum margin—that is, the maximum separation between data points for both classes—is our goal. In order to classify subsequent data points more confidently, it is helpful to maximize the margin distance. Decision boundaries known as hyperplanes aid in the classification of the data points. The classes of the data points that lie on either side of the hyperplane are distinct. By employing these support vectors, we are able to increase the classifier's margin.

3) Convolutional Neural Networks

Since the data set comprised of time series data for each candidate, so we reshaped it into a 3D data with a frequency of taking 100 samples with a crossover of 50 %. A 50 % crossover basically means that on taking each 100 groups we can have an overlap of 50 %. So, the final shape becomes N • 100 18 with also removing the first column of each file which was" time". After reshaping the data, we created a CNN model. So, the Input tensor was passed into the above Conv-net and was undergone a 10-fold classification. The weights were saved after training it with a 10-fold classification.

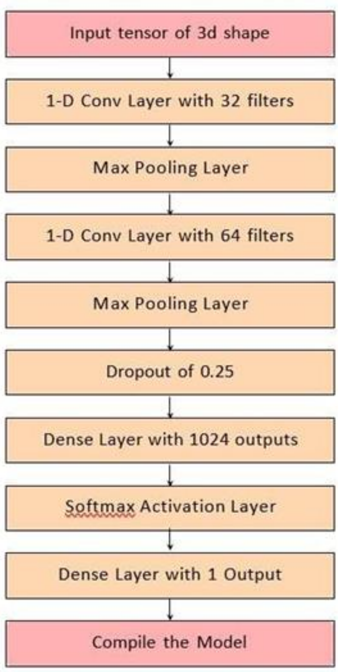

Figure 3.4: Flowchart for CNN

4) Long Short-Term Memory

Taking the same 3D data with the frequency of 100 samples per second with 50 % crossover, a LSTM network was trained on the formatted dataset. Since the sampling was done at a sampling frequency of 100, the units of the LSTM network were taken to be 100 with sigmoid activation. A similar 10-fold classification was also done for this network

5) Decision Trees

Decision Trees (DTs) are a non-parametric supervised learning method used for classification. The model learns basic decision rules derived from the data attributes in order to forecast the value of a target variable. It is comparable to an approximation of a piece-wise constant. Both raw and preprocessed data were used to train the decision tree, with the data being split using a random strategy and the leave one-out method.

6) Precision Model

We had trained another Logical Regression model which predicts if the given data is Parkinsonian or not from the predictions of the results from the above four models. Basically, the models discussed above would predict a value among 0 or 1 for a given data file and to assure which of them gave the most correct answer we trained another model using the data of 0's and 1's for each file which was available in the data. In this way, the sanctity of the prediction was maintained.

IV. SOFTWARE TOOLKIT

We have also developed a software that would facilitate future model training and would also cater to the normal people in getting an early-on prediction whether they have Parkinson's disease or not. Agile software development life cycle was employed for the development of the software. Agile software cycle consists of Concept, Iteration, Iteration, Release, Production and Retirement.

The software is developed on Flask, a Python framework used for designing web-based applications. MySQL is used for database management, HTML/CSS and Javascript are used for the frontend development while Python is the backend which is used by Flask. To build such a web system, we need three major parts for each component: database, user interface and the functions to interact between them.

1) Database

We created two models namely Users and Researchers and defined the different columns in the model itself using Flask Syntax. Every model is basically a class in terms of object-oriented programming. We have also implemented a Dashboard feature for the researchers by creating Profiles for them using the Flask-Login feature.

2) User Interface

The User Interface (UI) was developed by Flask templates. Templates in Flask allows us to create dynamic web pages and using Jinja templates we could extend the layout of some of the major components of the website to different web pages as well. A template, in layman terms, is just a simple HTML file with some Flask syntax blended in. Every web page has a corresponding template which allows the users to interact with it.

3) Modules

The entities of the project namely the Users, the Researchers, the Home Page, the Authentication functions were all divided into modules for ease in programmers' perspective and in case of Flask, these modules are named as Blueprints. So we have created four blueprints namely the user blueprint, the researcher blueprint, the Home blueprint and the Authentication blueprint. A blueprint simply can be understood as a separate Flask entity wherein it has its own templates and static folder, its own functions and its own interface.

Blueprints help in proper division of the program and also future changes in the code can be made more feasible with this technique. A proper structure of the project is thus maintained if blueprints are employed. These blueprints consists of a main file where all the required routes and methods are written which connect the user interface to the back-end of the toolkit.

V. RESULTS AND DISCUSSIONS

The models were trained on two different set of experiments, one in which the whole dataset was flattened to a 2-D space and the other one in which the whole dataset was re-visualized in a 3-D space with 100 samples taken at a time with a 50 % crossover.

The first experiment was done without any pre-processing and all the models were trained on the raw data. Both the datasets were normalized by their corresponding body weights since the forces are dependent of the body weights. After all these initial preprocessing steps, the additional advanced features were trained on the model and the corresponding accuracies were measured.

TABLE II. COMPARISON OF THE EXPERIMENTED MODELS

Model	Data used with or without augmentation	*Accuracy*
SVM	50 % Raw data without any feature normalization or selection	51.4%
SVM	50 % Data only with weight normalization	64.2%
SVM	Data with more advanced features aforementioned	88.2%
Decision Tree	Raw data without any feature normalization or selection	8z%
Decision Tree	Data with more advanced features aforementioned	72%
CNN	Data only with weight normalization	96%
LSTM	Data with weight Normalization	92%
Custom Model	Predictions from the above 4 models	94%

VI. CONCLUSION AND FUTURE WORK

The software toolkit developed consists of two sections for users and researchers. User section contains the functionality where user can upload his previously recorded GAIT data file and obtain results regarding being a parkinsonian from the four models and the accurate result would be displayed by the custom logical regression model we had trained. In researcher's part we developed a section where he develop his own model along with the ability to choose algorithm, sensor data and features to be included. The toolkit then automates the process by generating a Python script which can be downloaded and later run locally. The researcher's profile would be also be populated by the details of the models he had previously created using the GUI. This would allow him to compare the various models and develop a more comprehensive study.

In terms of the models used, we employed primarily a Logistic regression model which took input from the output of the four other Machine Learning models namely LSTM, CNN, SVM and Decision Tree. We had introduced many new features derived from the VGRF forces and had used two strategies were Leave-one-out and Random 2/3 method for splitting the dataset. Using 100 LSTM cells instead of a random number increased the accuracy considerably. This decision was based on the fact the data collection frequency matched with the number of cells. CNN stood out as the best model for the dataset. The logical regression model we had trained also substantially improve the sanctity of our accuracy for a particular data file. Decision of choosing outputs of four machine learning models for our custom model was based on combining both shallow and deep machine learning techniques for optimal output.

In the future versions, we would also include the option to add new models, new preprocessing methods and also allowing the researchers to add custom models and preprocessing methods using some API calls or just adding the names of the methods they need and our software would scrap the web for the snippets of the same. Also, the downloading of the script would be replaced by direct training on the cloud server with high computation. A curated medical report would also be generated for the general users. Furthermore, a setup can be built for the sensors of the users to record the user's VGRF data-file. Also, since all the prediction data for each user is saved, it can be used as a training data for future models, improvement in accuracy and just have a general overview of the trend-lines.

REFERENCES

[1] Wearable multisource quantitative gait analysis of Parkinson's diseases Authors: Junxiao Xie, Huan Zhao Junyi Cao , Qiumin Qu , Hongmei Cao , Wei-Hsin Liao Yaguo Lei , Linchuan Guo

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

[3] Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease Balaji Ea, Brindha Da , Vinodh Kumar Elumalai , Umesh K

[4] Gait Analysis with Wearables Predicts Conversion to Parkinson Disease (2022) Nidhi Soni, Mosam Patel, Prof Khushali Mistry Silvia Del Din, Morad Elshehabi, M Brook Galna, Markus A Hobert, Elke Warmerdam, Ulrike Suenkel, Kathrin Brockmann, Florian Metzger, Clint Hansen, Daniela Berg, Lynn Rochester, and Walter Maetzler

[5] .Statistical Analysis of Parkinson Disease Gait Classification using Artificial Neural Network: Hany Hazfiza Manap , Nooritawati Md Tahir and Ahmad Ihsan M Yassin

[6] Early detection of Parkinson's disease using machine learning Aditi Govindu, Sushila Palwe

[7] Determining the severity of Parkinson's disease in patients using a multi task neural network Mar'ia Teresa Garc'ia- Ordas Jose Alberto Ben ’itez-Andrades Jose Aveleira- Matal Jose-Manuel Alija-P’ erez " Carmen Benavides

[8] Machine Learning for Parkinson's Disease and Related Disorder : Johann Faouzi, Olivier Colliot & Jean- Christophe Corvol

[9] Gait Spatiotemporal Signal Analysis for Parkinson's Disease Detection and Severity Rating : Abdullah S. Alharthi; Alexander J. Casson; Krikor B. Ozanyan

[10] Early Detection of Parkinson Disease using Deep Neural Networks on Gait Dynamics: Lerina Aversano, Mario Luca Bernardi; Marta Cimitile; Riccardo Pecori