RESEARCH ARTICLE

SPECIAL ISSUE

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

Check for updates

http://dx.doi.org/10.47392/irjash.2023.S057

RSP Science Hub

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

Parkinson Disease Diagnosis and Severity Rating Prediction Based on Gait analysis using Deep Learning

Konkimalla Venkata Satya Sai Kumar [1], Indeti Sirisha [1], Kalakoti Vathsalya [1], Kotha Krishna Veera Venkata Vamsi [1]

[1]Department of Electronics & Communication Engineering, R.V.R. & J.C. College of Engineering , Guntur, Andhra Pradesh, India

Email: sai12345satya@gmail.com

Article History

Received: 28 February 2023
Accepted: 19 March 2023

***Keywords*:**

VGRF;
CNN;
Gait analysis;
CNN-LSTM;
Parkinson's disease.

Abstract

Parkinson's disease can present with both physical and emotional symptoms, which can vary widely between individuals. These symptoms may include problems with walking, shaking or tremors, difficulty maintaining posture, slowed movement, instability while walking, feeling tired or fatigued, experiencing feelings of sadness or hopelessness and have trouble in sleeping. It is important to diagnose Parkinson's disease as early as possible.This can greatly aid in managing the disease effectively, particularly in the healthcare industry.In this study, we used Convolutional Neural Networks (CNN) to diagnose PD and CNN-Long Short-Term Memory (LSTM) network using Hoehn & Yahr rating scale leads to severity rating prediction. Our dataset was obtained from Physionet which consists gait patterns of 93 PD patients and 73 healthy controls. We captured gait patterns using eight Force-Sensitive Resistors (FSR) located under the foot, which measured the Vertical Ground Reaction Forces (VGRF). To extract spatiotemporal features, including the swing phase, stance phase, and gait phase (a combination of swing and stance phase), we have used proposed classifier. We used spatial features which are measured across spatio-temporal features to predict gait abnormality. We had implemented the feature extractor using deep learning during the training process, which is a more efficient approach than manual implementation. We used the CNN for PD classification and the CNN-Long Short Term Memory for severity scale prediction based on the widely used Hoehn and Yahr scale.

1. Introduction

Parkinson's disease is a devastating condition that affects both motor and non motor functions of the body. (Vidya, Sasikumar, et al.) PD is a condition that results from a shortage of nerve cells in the brain that produce a chemical called dopamine. Dopamine is responsible for transmitting messages that control movement and coordination. (Camps et al.) As a result of this shortage, people with Parkinson's experience difficulties with movement and coordination, such as slowness of movement, muscle rigidity, problems with balance, and tremors (uncontrolled shaking). These motor symptoms are often the most noticeable and dis- ruptive aspects of the disease. PD can also lead to various non- motor symptoms. Analysis of gait is an essential component of the diagnostic process for Parkinson's disease, as it is often one of the earliest and most prominent symptoms. (Wahid et al.) Previous research has explored the use of machine learning methods for

PD classification. However, these algorithms typically require handcrafted feature extraction, which can be time consuming and labor intensive. We suggest a new way of classifying using Convolutional Neural Network (CNN) of Parkinson's disease patients and healthy subjects based on gait data.CNN can automatically extract features from raw data, making them well suited to gait analysis. Our proposed model aims to improve the accuracy and efficiency of Parkinson's disease classification. (·i Canturk)

The CNN-LSTM model by combining the strength of CNN and Long Short-Term Memory (LSTM) neural networks, spatiotemporal features are extracted from gait data. This approach has shown promise in previous research for other types of time-series data analysis. Overall, our proposed models have the potential to improve the accuracy and efficiency of Parkinson's disease diagnosis and severity rating prediction. (E, Brindha, Balakrishnan, et al.)

2. Related work

PD is a progressive diseases of the nervous system that progressively affects motor function, causing symptoms such as tremors, rigidity, and slowness of movement. Gait analysis is analyzing a person's walking pattern. By evaluating the severity of these abnormalities, gait analysis can be helpful in assessing the progression of Parkinson's disease over time. (Aite et al.) This approach can provide important insights into a person's condition and help healthcare providers develop more effective treatment plans. Recently, the application of Deep Learning(DL) techniques in the analysis of gait data has demonstrated significant potential for diagnosing and predicting the severity of Parkinson's disease. (Zhao et al.) Deep learning algorithms are increasingly being used in gait analysis to automatically extract features from data and classify patients based on their gait patterns. (Wu, Krishnan, et al.) By leveraging neural networks with multiple layers, enabling more accurate and objective assessments of gait abnormalities in individuals with conditions such as Parkinson's disease. Various studies have reported successful use of Deep Learning(DL) in PD evaluation and severity rating gait analysis based forecasting. (Hausdorff et al.) Machine learning (ML) approaches for PD diagnosis require handcrafted feature extraction, which is a major limitation. As a result, data-driven deep learning (DL) models have gained considerable attention for PD diagnosis. Deep learning models generated by data are used for Parkinson's disease (PD) diagnosis. These models can automate feature vector selection and considerably reduce the amount of human assistance needed to solve the categorization problem. (Kumar et al.) Two recent studies, by Zhao et al. in 2019 and Zhang et al. in 2021, have successfully applied deep learning models to PD diagnosis. (Abdulhay et al.)

3. Dataset

We utilized a time series dataset of gait obtained from the Physionet database for our study. The dataset was collected using eight Force Sensitive Resistors to measure the Vertical Ground Reaction Forces (VGRF) during three gait examinations: normal walking, treadmill walking, rhythmic auditory walking. We recorded the VGRF of both Patients with Parkinson's disease and healthy people. (W Langston) The dataset contains information from 166 participants, comprising 93 individuals diagnosed with Parkinson's Disease and 73 healthy individuals. This dataset provides a valuable resource for analyzing spatiotemporal gait features in Parkinson's Disease patients, which can contribute to the development of accurate diagnostic tools and treatment strategies. (Das)

3.1. Spatio-temporal features

Spatiotemporal features are used in data analysis when data is collected across both space and time. Spatial features include Step Length, Step Width, Stride length and the temporal features include Step Duration, Stance Phase, and Swing Phase. With the proposed methodology, features are automatically extracted using a CNN, removing the necessity for feature extraction that is manually crafted. (Benmalek, Elmhamdi, Jilbab, et al.)

TABLE 1. Number of individuals in each subject group for gait analysis in Parkinson's disease

Subject	Count
Si	64
Ju	54
Ga	47

However, in the proposed technique, features are automatically extracted using CNN, which eliminates the need for handcrafted feature extraction. This approach can lead to more accurate and efficient classification and prediction of spatiotemporal data. (Balaji, D. Brindha, Balakrishnan, et al.)

4. Methodology

In proposed methodology, a Convolutional Neural Network (CNN) is used to categorize individuals as either healthy or diagnosed with Parkinson's Disease. This classification is binary, with '0' representing healthy individuals and '1' representing those with the disease. In addition to this binary classification, a CNN-Long Short Term Memory (CNN-LSTM) model is used to estimate the severity rating of Parkinson's Disease. This is a multi class classification problem, and the Hoehn&Yahr scale is a widely used tool to evaluate the Severity of movement problems in individuals with Parkinson's. This scale involves rating a patient's symptoms based on the presence and extent of features such as tremors, rigidity, and bradykinesia. The scale is often used to assess disease progression and response to treatment, and can help healthcare providers tailor care plans to the needs of individual patients. The use of these deep learning techniques allows for automated feature extraction, which can lead to more accurate and efficient classification and prediction of Parkinson's Disease. This can ultimately change to better diagnosis and treatment for those affected by the disease.

4.1. Programming languages and libraries

We implemented our proposed models using Python programming language, specifically Python 3.7. We used the following libraries for data preprocessing, model building, and evaluation:

- NumPy for numerical computations
- Pandas for data manipulation
- Keras with TensorFlow backend for deep learning model building
- Evaluation metrics for classification tasks include accuracy, precision, recall, F1-score

4.2. Pre-processing the Gait dataset

Pre-processing refers to a below series of steps on Gait dataset.

4.2.1. Removing unwanted rows:

Observe the dataset for any irrelevant or duplicate rows. Duplicate rows contain redundant information and should be removed. Similarly,any rows that are not relevant to the analysis should also be removed.

4.2.2. Normalize the data:

To ensure that all data points in the dataset are on the same scale, it is necessary to normalize the data. Techniques such as min max scaling or standardization can be used to achieve this normalization. This ensures that no one feature dominates the analysis due to its larger scale.

4.2.3. Split the dataset:

To effectively train and validate models using DL, it is often necessary to split the dataset into separate training and testing sets. This involves partitioning the data into two distinct subsets, with the larger portion reserved for training the model and the smaller portion used for testing and validating the model's performance. This helps to ensure That the model is not fitting noise in the data can accurately generalize to new, unseen data.

4.3. Problem Statement

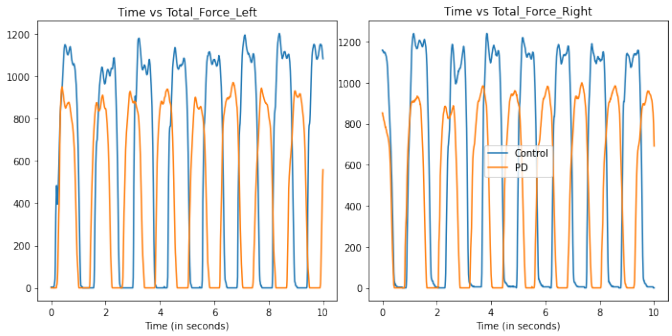
In previous research, machine learning algorithms have been utilized to address issues related to Parkinson's Disease. However, one challenge with these algorithms is that features must be manually extracted. To overcome this challenge, deep learning techniques like CNN can be used to classify Parkinson's Disease patients. Leveraging the capabilities of deep learning techniques, such as CNN and CNN-LSTM networks, can facilitate the extraction of meaningful features from complex gait data. By doing so, these techniques can significantly improve the accuracy and precision of Parkinson's Disease severity classification and prediction. Using deep learning techniques like CNN and CNN-LSTM can help in the automatic extraction of relevant features and enable accurate classification and prediction of Parkinson's Disease severity. These techniques can provide valuable insights for clinicians and researchers in the field, aiding in the development of more effective treatments for Parkinson's Disease.

4.4. Data Visualization

Analyzing features of the gait of PD individuals using Vertical Ground Reaction Forces (VGRF) collected from FSR sensors located under each foot can provide valuable insights into the effects of the disease on gait. One important comparison is between the gait characteristics of Parkinson's disease affected persons and people in good health. By comparing time and total force of VGRF between these two groups, it is possible to identify significant differences in gait characteristics. For example, Parkinson's disease patients may exhibit slower walking speeds, shorter stride lengths. Visualizing this data can be done using a variety of methods, such as scatter plots, line graphs, or box plots.

4.5. Splitting dataset into Training and Validation

To prepare dataset for training and testing, we need to load the files and convert them into NumPy arrays. Once we have the numpy arrays, we can reshape them into a 3D array of samples, timesteps, and features. The shapes of the resulting train



Graph 1: The graph describes the relation between Time(in seconds) vs Total force left and right.

and test files will indicate the number of samples, timesteps, and features in each file. In a time series dataset, the samples refer to the number of rows in the dataset. Each row represents a single observation at a particular point in time. A timestep is a number of sequential steps that make up a single input to the neural network. The number of features in a time series dataset refers to the number of columns in the dataset. By splitting the dataset into training and testing files, We can validate the model's performance on previously unseen data. This helps to prevent overfitting and ensures that the model generalizes well to new data.

4.6. Convolutional Neural Network (CNN)

Parkinson's disease can be classified using gait analysis, a technique that evaluates how people walk and move. This method can be enhanced by using Convolutional Neural Network (CNN), which are commonly used in image recognition. In this approach, a 1D CNN is applied to gait data collected from force-sensitive resistors placed under

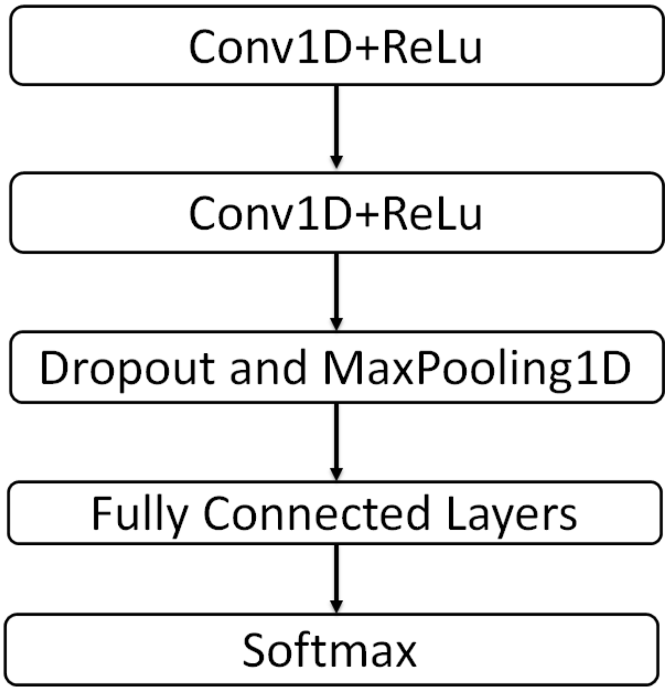


FIGURE 1. CNN Architecture

each foot. The time-series data from the resistors is fed as input to the CNN, and the output is processed by a 1D CNN to extract features. These features are then fed into a layer and a Max-Pooling layer, which flatten the predicted data and prepare it for the fully connected network. The flattened data is then supplied to dense layers that form the output layer for Parkinson's disease classification. The final step of the classification process involves the dense layer, which calculates the probability of each possible class based on the extracted features. The class with the highest probability is then chosen as the predicted class for the given input data. This method provides a reliable way to detect and watch Parkinson's disease using gain analysis and CNN.

4.7. CNN-Long Short Term Memory

The model architecture comprises several layers of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, which are a type of recurrent neural network designed to process sequential data.

The model architecture comprises several layers of Convolutional Neural Network(CNN) and Long Short-Term Memory (LSTM) networks, which are a type of recurrent neural network designed to process sequential data. The initial layers of the model consist of Conv1D layers with 64 filters and a kernel size of 5. These layers apply filters to the

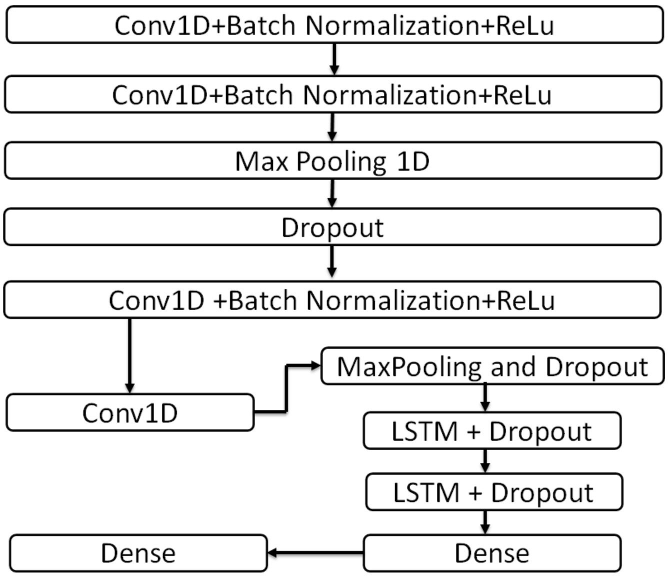


FIGURE 2. CNN-LSTM Architecture

input data to extract relevant features. To standardize and improve the performance and stability of the model, batch normalization layers follow each Conv1D layer. MaxPooling1D layers with a pool size of 2 are then incorporated to downsample the data and remove the parameters in the model. In order to prevent the model from memorizing the training data too closely, which could lead to poor generalization of new data, Dropout layers are included after the MaxPooling1D and LSTM layers. These Dropout layers randomly drop out a fraction of the nodes in the layer during training, forcing the remaining nodes to learn more robust features that are less dependent on the specific input data. This helps improve the model's ability to generalize to new data and improve its overall performance. LSTM layers are a type of recurrent neural network (RNN) layer that can process sequential data, allowing the model to capture the temporal dependencies or patterns in the gait data. The first LSTM layer has 128 units, while the second layer has 64 units. These units are like individual processing nodes that work together to analyze and learn from the gait data. Overall, the use of LSTM layers in the subsequent layers of the model enables it to better process and understand the sequential nature of the gait data, leading to improved performance when it is applied to new data. By capturing temporal dependencies in the data, the model can generalize and make better predictions, even when faced with new or previously unseen gait data. Ultimately, two dense layers are appended with 32 and 4 units, respectively. The last dense layer features a softmax activation function that generates a probability distribution over the severity ratings. The model is constructed using the Adam optimizer and categorical cross-entropy cost function. During training, accuracy is used as a metric to evaluate the model's performance.

5. Block Diagram

• Collecting Gait Patterns using FSRs: In this block, gait patterns are collected from 93 PD patients and 73 healthy subjects using 8-Force Sensitive Resistors (FSRs) located under their feet.

• Preprocessing the Data: The collected data is cleaned and processed to remove noise or errors that could affect the accuracy of classification.

• Visualizing the Data: The preprocessed data is visualized to gain insights into its distribution and characteristics.

• Data Preparation: The preprocessed data from 303 files are concatenated to create 19 separate files, each containing the combined data of a single column.

• Creating the CNN Model: A Convolutional Neural Network (CNN) model is designed to verify Parkinson's disease based on the gait patterns of the subjects. The CNN model uses spatiotemporal features, such as swing phase and stance phase, to predict gait abnormality.

• Classifying Parkinson's Disease: The gait patterns are classified into two categories PD patients and healthy subjects using the CNN.

• Creating the CNN-LSTM Model: A Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) network is used to predict the severity rating of Parkinson's disease. The CNN-LSTM model takes the gait patterns of the PD patients as input and predicts their severity rating based on temporal features.

• Predicting Severity Rating: The severity rating of Parkinson's disease is predicted using the CNN-LSTM model severity rating is commonly assessed using rating scales such as Hoehn and Yahr.

6. Experimental results and discussions

The proposed CNN-based Parkinson's disease classifier and CNN-LSTM based severity rating prediction models were implemented using Tensorflow with the Keras library in the conducted experiments. The time series dataset is split into 75% used for training and 25% used for testing purposes. The

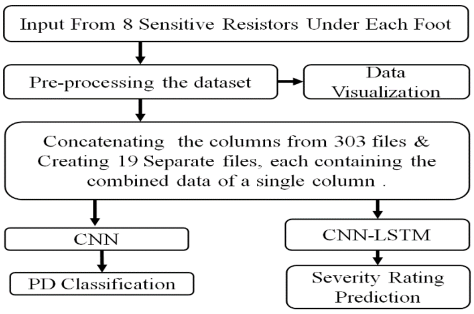


FIGURE 3. Block Diagram for PD Classification & Severity Rating Prediction

number of samples used for training was 34,575, while for testing, it was 8,968 samples.

TABLE 2. CNN results of PD classification of gait data

PD Classification	CNN		
	Precision	**Recall**	**F1-Score**
0	0.96	0.98	0.97
1	0.96	0.92	0.94
Accuracy		0.95	-

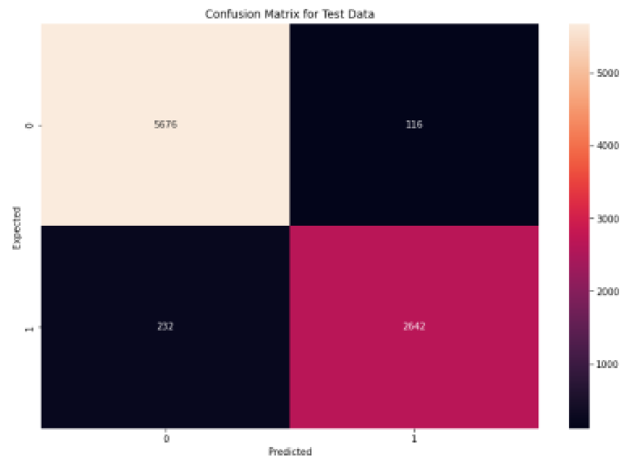


FIGURE 4. Confusion matrix for PD classification using CNN.

The time steps used were 100, and the dataset had 19 features. The accuracy of Parkinson's disease classification achieved 95% using the CNN model, while the severity rating prediction achieved 88% accuracy using the CNN-LSTM network. Additionally, a CNN model was employed for estimating the seriousness rating of this disease, resulting in an accuracy rate of 77%. Precision and recall are two common performance metrics used to evaluate the accuracy of machine learning models, particularly for binary classification problems. Precision represents the fraction of true positive predictions among all positive predictions, while recall represents the fraction of true positive predictions among all actual positive samples. A high precision value indicates that the model has a low false positive rate, while a high recall value indicates that the model has a low false negative rate. These metrics are important to consider in different scenarios, depending on the desired trade-off between minimizing false positives or false negatives. The accuracy is the proportion of correct predictions out of all predictions made (correct predictions / total predictions).

TABLE 3. CNN-LSTM results of severity rating prediction on gait data

Severity rating of prediction	CNN-LSTM		
	Precision	**Recall**	**F1-Score**
0	0.84	0.93	0.88
1	0.91	0.90	0.90
2	0.91	0.78	0.84
3	0.86	0.94	0.90
Accuracy		0.88	

TABLE 4. CNN results of severity rating prediction on gait data

Severity rating of prediction	CNN		
	Precision	**Recall**	**F1-Score**
0	0.78	0.81	0.79
1	0.85	0.83	0.84
2	0.68	0.76	0.72
3	0.93	0.58	0.71
Accuracy		0.78	

7. Limitations of proposed approach

This model was developed using a publicly available dataset, it is reasonable to expect that it would perform effectively on larger datasets as well. However, it is important to note that further validation on larger datasets is still necessary to confirm the generalizability of the approach and ensure that the model can be reliably applied to a wider population.

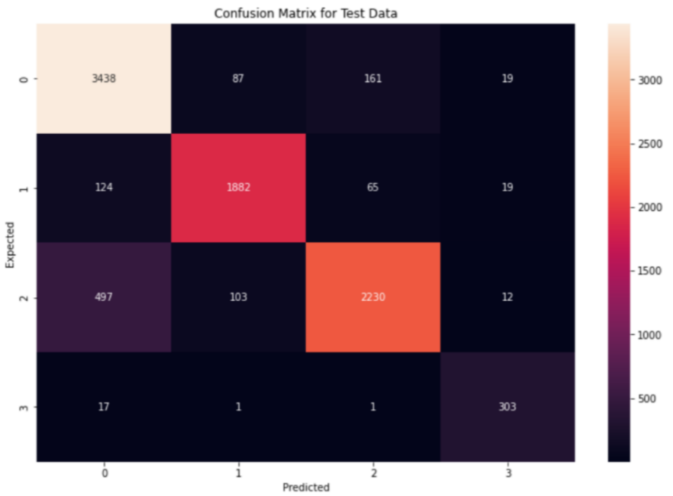


FIGURE 5. **Confusion matrix for severity rating using CNN-LSTM**

This would enable the development of more comprehensive approaches to PD care that consider the diverse range of symptoms associated with the disease. By incorporating data on other symptoms such as tremors, speech, and cognitive function, it may be possible to develop more effective and personalized treatments for PD. This highlights the importance of continuing to explore different sources of data and developing more holistic approaches to PD diagnosis and management.

8. Conclusions

In this paper, we propose a methodology that utilizes Convolutional Neural Networks (CNNs) to classify Parkinson's Disease (PD) determined by gain analysis, and a CNN-LSTM to estimate the severity rating using a network using the Hoehn Yahr rating scale. The input data consists of time series data collected from force-sensitive resistors placed on each foot. The model architecture consists of multiple layers of CNNs and LSTMs. We train the model using the Physionet dataset, which includes gait patterns of 93 PD patients Our findings demonstrate that the suggested methodology produces great accuracy in both PD classification and severity rating prediction. This paper suggests that deep learning techniques, such as CNNs and LSTMs, can effectively diagnose and monitor PD. Although CNNs are a powerful tool for classification tasks, they may not be optimal for time series data such as gait patterns. Adding LSTM layers to a CNN architecture can improve performance on time series data by capturing temporal dependencies. Therefore, using a CNN- LSTM network for severity rating prediction based on gait analysis is a more effective approach than using only a CNN.

References

Abdulhay, Enas E, et al. "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease". *Future Generation Computer Systems* 83 (2018): 366–373.

Aite, Zhao, et al. "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data". *Neurocomputing* 315 (2018): 1–8.

Balaji, E, D Brindha, R Balakrishnan, et al. "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease". *Appl Soft Comput* 94 (2020).

Benmalek, E, J Elmhamdi, A Jilbab, et al. "UPDRS tracking using linear regression and neural network for Parkinson's disease prediction". *International Journal Of Emerging Trends Technology In Computer Science (IJETTCS* 4 (2015): 189–193.

Camps, Julià, et al. "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit". *Knowledge-Based Systems* 139 (2018): 119–131.

Das, R. "A comparison of multiple classification methods for diagnosis of Parkinson disease". *Expert Systems with Applications* 37.2 (2010): 1568–1572.

E, Balaji, Brindha, R Balakrishnan, et al. "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease". *Appl Soft Comput* 94 (2020): 10649–10649.

Hausdorff, Jeffrey M, et al. "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease". *European Journal of Neuroscience* 26.8 (2007): 2369–2375.

·i Canturk. "A computerized method to assess Parkinson's disease severity from gait variability based on gender". *Biomedical Signal Processing and Control* 66 (2021): 102497–102497.

Kumar, Anil, et al. "A new Diagnosis using a Parkinson's Disease XGBoost and CNN-based classification model Using ML Techniques". *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA* (2022): 1–6.

Vidya, B, P Sasikumar, et al. "Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN-LSTM network". (2022).

W Langston, J. "Parkinson's Disease: Current and Future Challenges". *NeuroToxicology* 23.4-5 (2002): 443–450.

Wahid, Ferdous F, et al. "Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features". *IEEE Journal of Biomedical and Health Informatics* 19.6 (2015): 1794–1802.

Wu, Yunfeng, Sridhar Krishnan, et al. "Statistical Analysis of Gait Rhythm in Patients With Parkinson's Disease". *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.2 (2010): 150–158.

Zhao, Aite, et al. "Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases". *Knowledge-Based Systems* 145 (2018): 91–97.

To cite this Article: , Konkimalla Venkata Satya Sai Kumar, Indeti Sirisha , Kalakoti Vathsalya , and Kotha Krishna Veera Venkata Vamsi . "**Parkinson Disease Diagnosis and Severity Rating Prediction Based on Gait analysis using Deep Learning**." International Research Journal on Advanced Science Hub 05.05S May (2023): 418–425. http://dx.doi.org/10.47392/irjash.2023.S057