A Layer-Wise Relevance Propagation Based Feature Selection and Hybid Classification Model for Automatic Detection of Parkinson's Disease Using Gait Signals

Swathika R[a], Radha N[b]

Abstract
Gait irregularities frequently result from orthopaedic or neurological diseases, and they can have serious repercussions including limiting movement and falling. Gait analysis is essential for tracking gait irregularities, and identifying underlying impairments can aid in creating treatment plans. Spatio-temporal, kinematic, and muscle activation gait features must all be examined in today's multi-modal gait analysis. The second most prevalent neurodegenerative ailment, PD (Parkinson's disease), is brought on by the midbrain's premature neuronal loss. The absence of valid neuropathological criteria prevents a firm pathological diagnosis of PD. To classify PD severity, DLTs (deep learning techniques) are utilised to analyse and combine raw data of gait-induced GRFs (ground reaction force) for PD diseased and healthy patients. The current LSTM model also requires a lot of training time and is very sensitive to random weight initialization. For the effective operation and reliable prediction of PD data, this study introduces an efficient feature selection and hybrid CNNs (Convolution Neural Networks) with LSTM (Long Short Term Memory). In the beginning, processing is done to remove noisy data. Then, to evaluate the results of the models and reveal which elements in the spatiotemporal gait GRFs signals are most important for the models' predictions, feature selection is carried out via LRPs (Layer-wise relevance propagations). This enables their assignment to gait events, suggesting that heel strikes and body balances are best suggestive gait aspects for the categorization of healthy gait. Landing of the foot and body balances are the most affected during late stages of PD. Finally, Hybrid CNNs with LSTM for the reliable prediction of PD data and efficient operation. The suggested models can be helpful for identifying changes in postural balance and grading PD severities since they are robust towards noises and process/ classify big datasets efficiently.
Keywords: Sensor applications, Parkinson's Disease, gait characteristics, Layer-wise relevance propagation (LRP), Hybrid Convolutinal Neural Network (CNNs) with Long Short Term Memory (LSTM).

1. Introduction

Humans walk in a way known as gait, which is caused by the centre of gravity of the body moving forward in two phases. Gaits are actions of lower limbs that result in co-ordinated and recurrent foot's contacts with surfaces (Pardoel et al., 2019) Gait analyses are hot research areas for a variety of activities and applications including biometrics, sports, and healthcare. Clinical manifestations of PDs including bradykinesia, rigidities, decreased amplitudes and reduced automaticity of movements impact gait patterns of PD patients.

[a]Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India. Email: swathikar@ssn.edu.in
[b]Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India. Email: radhan@ssn.edu.in
Corresponding author: Swathika R

Along with reductions in gait speeds and step lengths, patients also exhibit increased axial stiffness and rhythmicity impairments (Abou et al., 2021). Gait issues are a key illness burden that significantly reduces independence and quality of life as the disease develops (panel). Dopaminergic medications enhance certain aspects of walking including step lengths and speeds, but have reduced impacts on reactions to temporal traits and episodic symptoms like freezing of gaits, which is abrupt inabilities in walking continuously. The majority of cross-sectional research used to study gait deficits in PD patients do not offer insights into specific gait modifications linked to disease progression (Mirelman et al., 2019). The use of gait measurements as feasible endpoints for determining disease's progressions and their cure are constrained by the rare use of quantitative gait evaluations (e.g., speeds and variability) in regular clinical examinations and lack of knowledge of the underlying processes.

The evaluation, understanding of processes, and treatment of gait deficits in PD patients have all benefited from advancements in wearable technology and imaging methods (Pardoel et al., 2021). The complicated camera-based motion-capture systems are being quickly replaced by low-cost sensor technologies, such as accelerometers and gyroscopes, which enable doctors to assess gait quantitatively in clinics and even for normal home or community based activities. These assessments offer insightful data on motor fluctuations, habitual function, and pharmaceutical response (Balaji et al., 2021). Understanding the processes driving patientrs affected with PD in terms of gait patterns has improved mainly due to neuro-images that capture brain activities during walking resulting in developments of potent treatment therapies. One of the major effects of PD is a deviation from a healthy strides and in later stages, it increases the chance of falling. The measurement of visual gait, however, may yield conflicting results in PD’s early stages. This is mostly because the sluggish walking and short stride may also be caused by old age, depression, or other diseases. This is supported by the fact that PD results in tremor, muscular rigidity, and sluggish movement (Torvi, Bhattacharya, & Chakraborty, 2018). The gait cycle (see Fig. 1) is regularly observed visually in medical settings to check for gait abnormalities, which serves as the foundation for the diagnosis of PD and the assessment of the disease's severity. These techniques are therefore semi-subjective.

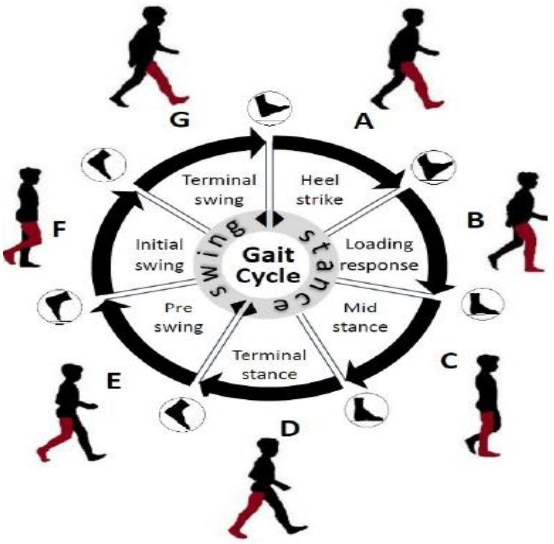

Figure 1 **shows significant gait events and breaks in a normal gait cycle.**

|Signals emanating from VGRFs (vertical ground reaction forces) were investigated as novel gait assessments for determining severity of PD in addition to binary classifications from normal gaits based on the dual-modal DLTs (Zhao et al., 2014). High accuracies were obtained in distinguishing patients with PD, normal, and other neurodegenerative disorders based on a combination of methods utilised including *k*-NNs *(K*-nearest neighbours), DTs (decision trees), RFs (random forests), NBs (Naïve Bayes), SVMs (support vector machines), *K*-means, and Gaussian mixtures. Gait categorizations of patients with PD were based on basic, kinematic, and kinetic parameters in statistical analyses which were executed using ANNs (Artificial Neural Networks). According to the findings, walking speed, knee angle, step length, and VGRFs were the most obvious traits. In a follow-up study, gait dynamics are used to predict PD using multichannel tensor decomposition (Nguyen et al., 2017). Contrarily, the average walking speed of a human is 1.4 m/s.. However, older people typically walk more slowly to increase their dynamic stability. The slower stride that comes with ageing causes more gait variance, which raises the risk of falling. Due to the degradation of neurons and the diminished capacity to regulate the locomotor and muscular systems, this is a notable trait in PD patients. Walking more slowly might reduce the metabolic energy expenditures connected to aging-related anaerobic capability. Age times speed in terms of trunk roll angle is a key indicator of the gait variability that may occur in healthy, typical ageing adults. Elderlies also have low foot clearance (Yang et al., 2014). The COP variation beneath the foot could represent this. Undoubtedly, PD has a gait disorder characterised by postural instability, delayed walking, as well as shuffles and problems commencing steps.

By utilising automatic feature extractions on raw sensor data, arbitrary characteristics were eliminated by DANNs (Deep ANNs) (Naghavi, Miller, & Wade, 2019). The investigations of PD affected gaits using DLTs like LSTM (Demrozi et al., 2019) demonstrate that these methods generate techniques for analysing gait data that are extremely important. Fundamental DANNs, however, allow clear relationships back to input data properties for models to arrive at certain conclusions. They also limit model’s predictions due to the model's opacity which makes it more challenging. They also add complexity to process classifications into groups, thus preventing systematic advancements of gait characterisations.

The current LSTM model also requires a lot of training time and is very sensitive to random weight initialization. To help, this study introduces an effective feature selection method and hybrid CNNs with LSTM for the reliable prediction of PD data.

Section 2 of the remaining research is devoted to a review of current gait traits-based Parkinson disease prediction. Section 3 outlines the methodology's recommended approach. The findings and discussion are presented in section 4. Section 5 covers the conclusion and further research.

2. Literature Review

The evaluation, processes, and therapies to enhance gait are three linked elements of gait abnormalities in PD patients that are evaluated in this section. The review identifies knowledge gaps and offers perspectives that might result in fresh insights and innovations to enhance clinical judgments and treatments.

In order to assess gait data, El Maachi, Bilodeau, and Bouachir (2020) suggested an unique intelligent Parkinson diagnosis method based on deep learning approaches. In order to create a DNN (Deep Neural Network) classifier, they employed 1D CNNs (1D-Convnets). Eighteen 1D data from foot sensors detecting VGRFs are processed using the suggested model. The network's initial section is made up of 18 parallel 1D-Convnets that correlate to system inputs. The outputs of the 1D Convnets are concatenated in the second component, which is a fully connected network, to provide final classifications. Additionally, this algorithm was evaluated using the UPDRS (Unified PD Rating Scale) to diagnose Parkinson's condition and forecast the severity of the disease. Their tests showed that the suggested strategy for detecting Parkinson disease using gait data works quite effectively. The suggested method had a 98.7% accuracy rate. also managed to predict the severity of PD with an accuracy of 85.3%. Unsupervised approaches were substituted by feature learning techniques based on PCA (principal component analysis) and time based statistics by Mazilu et al. (2013) in their study. The study found that the latter routinely outperformed the former in terms of F1 measures for FoG (Freezing of gait) detections, Freezing Indices by 8.1%. Investigations of pre-FoG patterns, or patterns that appear before FoG occurrences could be advanced in contrast to the use of EEG, the sole method that was used. By reaching F1 measures of 56% in pre-FoG class for patients who revealed adequate gait degenerations prior to FoG, the study showed that FoG prediction ability was patient dependent, relating to "three-class" issues (FoG vs. pre-FoG vs. normal locomotion). Vidya and Sasikumar (2021) introduced gait classifications based on MCSVM (multi-class SVM) decision support systems as gait changes are initial PD symptoms. The study used VGRF dataset for kinematic analyses and extracted spatiotemporal properties. Important gait biomarkers were uncovered in this work as it used correlation based feature selections along with multi-regressions to normalize gait time series data values. By breaking the multi-class classification problem into multiple binary classification issues using 1-1 method, the proposed PD severity evaluation methodology additionally assessed performances of four SVM kernel functions using three different walking tests. According to experimental findings, the quadratic SVM classifier performs better than previous cutting-edge techniques that used gait datasets for PD diagnosis, with an average accuracy of 98.65%.

Using the AnyBody modelling framework, Eltoukhy et al. (2017) created a Kinect-driven musculoskeletal model to forecast three-dimensional GRFs during walking in PD patients. Nine PD patients underwent ground walking trials where kinematics and GRFs (ground reaction forces) were measured using Kinect v2 and force plates. Peak vertical and horizontal ground reaction forces and impulses generated throughout the braking and propulsive stages of gait cycles were assessed. To test if Kinect sensors could predict GRFs precisely and consistently throughout gait cycles, 3-D ensemble curves of GRFs with associated 90% confidence intervals (CI90) were compared. The results showed that Kinect v2 sensors were viable clinical assessment tools for predicting GRFs produced by movements in people with PD. Using MLTs (machine learning techniques), Borzì et al. (2021) suggested a wearable system that can identify the usual worsening of the walking pattern that occurs before FOG episodes and determine whether dopaminergic medication impacts the system's capacity to detect and forecast FOG. 11 PD patients were divided into two groups: those getting dopaminergic treatment (on) and those not receiving it (off). Both groups were given two inertial sensors, one on each shin, and instructed to complete a timed up and go test. and carried out an iterative process of segmenting the angular velocity data, followed by feature extractions of both time and frequencies. To identify FOG and pre-FOG events, they used a wrapper technique for feature selection and improved several MLTs. With patients both on and off treatment, the built FOG detection algorithm demonstrated good performance in a leave-one-subject-out validation. In terms of pre-FOG identification, the developed classification algorithm in patients receiving (not receiving) treatment obtained 84.1% (85.5%) sensitivity, 85.9% (86.3%) specificity, and 85.5% (86.1%) accuracy. A unique CNN-based deep FOG detection technique was presented by Xia et al. (2018) where data segments from 1 dimensional acceleration signals were inputs. CNNs effectively achieved automated feature learning and detections of FOG occurrences from routine walks and eliminated the need for hand crafted feature extractions or time consuming feature selections. The study used a dataset encompassing eight hours lab data of ten PD patients who frequently experience FOG. Their system's classification accuracy in a patient dependent contexts exceeded 99% while 80.70% in patient-independent situations. Without GPUs (graphics processing units) accelerations, the method took only 3.6 ms to classify 4s data segments. Camps et al. (2017) proposed FOG detections based on DL|Ts and signal processing methods and it was the first application of DLTs in detections of FOG. The study used information from 15 PD patients who showed FOG for their proposed model's assessments. Triaxial accelerometer, gyroscope, and magnetometer signals were captured by an inertial measuring device mounted on the left side of the waist. The method attained validation performances of 88.6% and 78% for sensitivity and specificity, respectively and equivalent to current standards.

VGFR Spectrogram Detector and Voice Impairment Classifiers were two neural network based models proposed by Johri and Tripathi (2019) for early sickness identifications in healthcare. DANNs predicted sicknesses based on voice recordings and large scale image classifications of gait signals which were converted to spectrogram images. Their experimental results demonstrated that their proposed model outperformed most methods in terms of accuracy. The classification accuracy of their VGFR Spectrogram Detector was 88.1%, while Voice Impairment Classifier scored 89.15% in accuracy of detecting PDs. Intelligent categorizations were proposed by Celik and Omurca (2019) where RFs, Extra Trees, Gradient Boosting, SVMs and LRs (Logistic Regressions) were examined. They used a total of 1208 speech data sets comprising of 26 features gathered from PD patients and non-patients for classifications. Feature spaces of datasets were widened through correlation maps where characteristics were gathered using PCA, IG (Information Gain). The classification results obtained with extended features are superior than those obtained with the data's original features, it is found.

A model based on MLTs was created by Dash (2021) utilising the five classifiers RFs, , LRs (Logistic Regressions), XGBoost, AdaBoost, and Gradient Tree Boosting to predict PD where Gradient Tree Boosting had the most impressive accuracy (98.31%) with ROC curve of 98.66%. The study demonstrated that the stated design is more accurate than the existing techniques described in the literature and that its occurrence rate is lower. In order to rate the severity of PD using gait pattern, Balaji et al. (2021) introduced a unique deep learning architecture based LSTM network. Unlike MLTs, the LSTM network learns the long-term temporal relationships in the gait cycle without the requirement for manually created features for robust diagnosis of PD. By substituting memory blocks for self-connected hidden units, the LSTM network resolves the vanishing gradient problem and can thus decide when to acquire new information. For training the LSTM network, three separate gait datasets with VGRF recordings for various walking situations are employed. The suggested method makes use of dropout and L2 regularisation techniques to prevent data overfitting. H&Y (Hoehn and Yahr) and UPDRS (universal PD rating scale) categorized severity levels of PD patients. Adam, a stochastic gradient-based optimizer solved cost functions where experimental results showed that Adam-optimized LSTM networks could effectively learn gait kinematic parameters and offer binary classification accuracies of 98.6% and 96.6% accuracy for multi-class classifications, thus improving accuracy ranges by 3.4% when compared with other relevant approaches.

3. Proposed Methodology

Introduce an effective feature selection method and hybrid CNNs with LSTM in this part for the reliable prediction of PD data. And figure 2 shows the suggested model's workflow in action. The input layer of the HCNN-LSTM receives the pre-processed, segmented signals of VGRFs. Gait cycle has important spatiotemporal characteristics that can distinguish between healthy people and people with PD. Processes first eliminated noisy data followed by the use of LRPs to analyse model outputs and reveal crucial spatiotemporal gait GRF signals for predictions. The study indicated that body balances and foot landing were important gait characteristics of patients in late stages of PD, whereas heel strikes and body balances were the most indicative gait aspects for classifications of healthy gaits. Finally, Hybrid CNNs with LSTM for the reliable prediction of PD data and efficient operation.

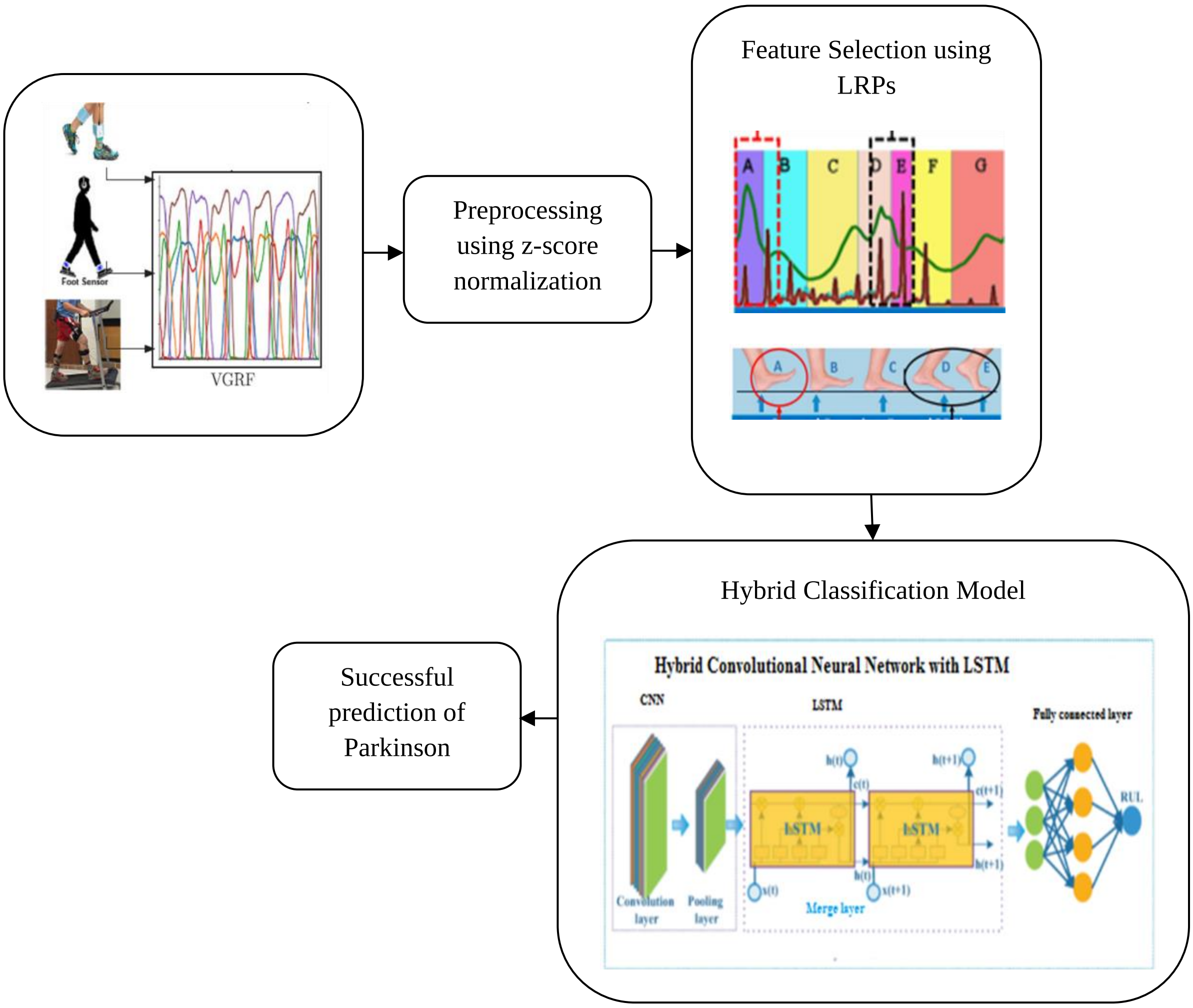

Figure 2. **The overall process of the proposed HCNN-LSTM model based detection of PD**

3.1. Preprocessing using Z-Score normalizations

Data preparations are processes of converting raw data into understandable formats and it is pertinent to assess data's quality before using MLTs or other data mining techniques for processing data. This phase is essential as quantity and calibre of accuracy are important to data and the findings were found to be more pertinent when both amount and quality of images were good. The dataset in this study was normalised using Z-score approaches. The result section includes a detailed description of the dataset.

- **Z-Score Normalizations**

The average intensity for each individual dataset was first calculated for each experiment's raw intensity data, and then the average of the averages was calculated (Patro & Sahu, 2015). The computation of the normalisation factors, which were then applied to each experiment, was based on this grand average. The overall average was the same for all subsequent normalised data averages.

Z-scores can be plotted on normal distribution curves in the range of -3 and +3 for standard deviations (σ), or at extreme lefts and rights of normal distribution curves.

Concretely, let xi (i = 1, 2, · · · , D) denote the i-th component of each feature vector x ∈ R D. We first compute the mean and the standard deviation of these D components:

$$\mu_x = \frac{1}{D}\sum_{i=1}^{D} x_i, \sigma_x = \sqrt{\frac{1}{D}\sum_{i=1}^{D}(x_i - \mu_x)^2} \quad (1)$$

Z-score normalization is then applied as

$$x^{(zn)} = ZN(x) = \frac{x - \mu_x 1}{\sigma_x} \quad (2)$$

According to these computations, the original feature vectors are first projected along the 1 vector to a hyperplane that encompasses the origin and is perpendicular to √ 1. The resultant normalised vectors are then scaled to have the same length as D, such that they lie on a hypersphere of radius √ D. The method of feature selection is carried out after preprocessing the provided data, as explained in the section below.

3.2. Feature Selections

The most used artificial intelligence approach for gait analysis is supervised learning, by far. Such MLTs use supervised mapping of ground truth input data to an output label to train a network in order to identify the gait pattern. The actual designation in this case is either healthy or PD. The models pick up on the various gait degradation phases during this process.

- **LRPs**

LRPs are backward propagation techniques that show which ANN input vector segments are most important for model prediction. In this study, we measure the contribution of a single input xi component (in our example, a sensor signal at a particular time frame) to the prediction of $f_c(x)$ (*c*) signify a class of PD severities rating, based on gaits). Through back propagations, intermediate nodes obtain fresh copies of predicted PD severity rating up to input layers (Binder et al., 2016). LRPs highlight signal segments that contribute most to model's predictions like for example data segments with most variability create "heat maps" over original signals. Neural networks encompass several layers of neurons (feature maps in the case of a convolution layer) and where neurons are activated using:

$$a_j^{1+l} = \sigma\left(\sum_i a_i^l\, \omega_{ij}^{(l,l+1)}\right) + b_j^{(l+1)} \tag{3}$$

Here, a_i^l denote the activation of a neuron *i* in the previous layer in forward direction; $\omega_{ij}^{(l,l+1)}$ denote the weight received in forward direction by neuron *i* from neuron *j* in the previous layer; a_i^l $\omega_{ij}^{(l,l+1)}$denote the contribution of neuron *i* in layer *l* to the activation of the neuron j in layer *l* + 1. The sum is computed over all neurons that are connected to neuron *j*. The function σ is a nonlinear monotonously increasing activation function, a rectified linear unit 'ReLU (max(0,x)') and $b_j^{(l+1)}$ is a bias term. The DCNN picks up these biases, weights, and activations during supervisory training. The parameters $(\omega_{ij}^{(l,l+1)} + b_j^{(l+1)})$ are updated via back-propagating during training utilising model error for the latter and calculations on categorical cross entropy.

The LRPs approach decomposes the DCNN output for a given prediction function of gait class *c* as f_c for input x_i and generates a "relevance score" R_i^l for *i*th neuron of layer *l* and R_j^{l+1} from the *j*th neuron in the previous layer, where the relevance conservation principle is satisfied as:

$$\sum_i R_{i\leftarrow j} = \sum_i R_j = f_c(x) \tag{4}$$

LRPs start at DCNN's output layers after Softmax layers are removed. This technique eliminates other classes and uses gait classes C as inputs to LRPs (healthy gaits or one of the three PD severity ratings). The back propagations necessary to un-pool pooling layers needs to be computed by sending signals to neurons for which activations are computed in forward passes. When average-pooling is used, the neuron with the average activation value is taken into account.

As generalisation, have a look at output neuron I in one of the model layers. From a neuron j in the higher layer, which represents the model's output, it receives a relevance score Rj (class c). The activation function of neuron j, which is created in the forward pass and updated by back-propagating during training, is used to redistribute the scores among the connected neurons throughout the network layers based on the contribution of the input signals xi. The latter will retain a certain relevance score and communicate its value to following neurons in the opposite direction based on its activation function. After that, the system calculates relevance scores for each sensor signal during a certain time period. High relevance scores for particular time frames emphasise the regions that influenced the model classifications the most in a heat map created from these scores. The following is a representation of the relevance propagation rule Ri for layer l:

$$R_i^l = \sum_j \frac{a_i^l \omega_{ij}^{(l,l+1)}}{\sum_j a_i^l \omega_{ij}^{(l,l+1)}} R_j^{(l+1)} \tag{5}$$

There are other propagation rule such as $(\alpha\beta - rule)$

$$R_i^l = \sum_j (\alpha.\frac{(a_i^l \omega_{ij}^{(l,l+1)})^+}{\sum_j (a_i^l \omega_{ij}^{(l,l+1)})^+} - \beta.\frac{(a_i^l \omega_{ij}^{(l,l+1)})^-}{\sum_j (a_i^l \omega_{ij}^{(l,l+1)})^-} R_j^{(l+1)} \tag{6}$$

where $(a_i^l \omega_{ij}^{(l,l+1)})^+$ and $(a_i^l \omega_{ij}^{(l,l+1)})^-$ denote the positive and negative part of $a_i^l \omega_{ij}^{(l,l+1)}$ respectively, such as that $(a_i^l \omega_{ij}^{(l,l+1)})^+ + (a_i^l \omega_{ij}^{(l,l+1)})^- = a_i^l \omega_{ij}^{(l,l+1)}$in eq 4, and the parameters $\alpha\ and\ \beta$ are chosen so that $\alpha - \beta = 1\ and\ \beta \geq 0$. Other stabilizing terms can be used to avoid divisions by zero.

Human gait differs from person to person, and even within a single individual, therefore models must be reliable and tolerant of changing input data. When employed for the LRPs analysis, the interpretation of relevant input data points must be resistant to noise and volatility in the input data stream. The sounds in the input data stream are effectively reduced by the normalising approach.

The normalising method efficiently reduces the noises in the incoming data stream. By progressively reducing the greatest relevance scores provided by the best LRPs technique selected using the aforementioned method, this process is completed in order to assess the significance of DCNN model design. Next, the model performance is tested by re-predicting on the test data for each model. The models most suited to taking use of LRPs are those whose performance significantly declines after just a few perturbation steps. This is because the performance drop makes it possible to claim that the few regions that were deleted were essential for correct classification performance, and as a result, it is an indication of significant connections between the input patterns and learned classes.

3.3. Hybrid Classification model

Using efficient operations and reliable predictions of PD data using hybrid CNNs with LSTMs, the suggested models can be helpful for identifying changes in postural balance and grading PD severities since they are robust to noises and classify big datasets efficiently.

3.3.1. LSTM

Long-term data storages and access are made possible by replacements of typical nodes in hidden layers with memory cells in LSTM. LSTM networks are extensively used in time series predictions for applications including machine translations, air pollution/weather forecasts and speech recognitions (Zheng et al., 2020). Fig. 3 depicts memory cells and blocks that make up LSTM's hidden layers. There are three gate elements namely input, forget and output gates. The multiplicative gate units are used to prevent the detrimental consequences caused by unrelated inputs.

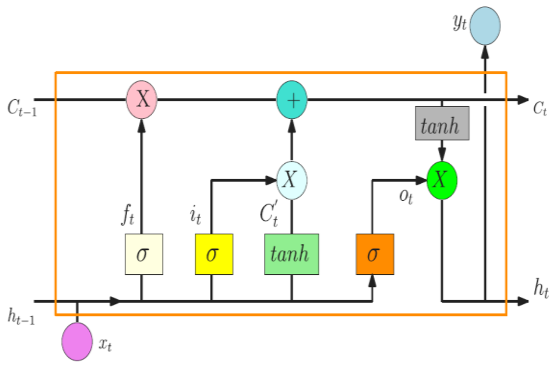

Fig. 3. **LSTM block diagram with gating mechanism.**

The amount of data to be preserved or forgotten by the memory cell is determined by the forget gate. The forget gate's activation function is calculated as shown below.

$$f_t = \sigma(W_{hf}h_{t-1} + W_{xf}x_t + b_f) \quad (7)$$

where b is the bias vector, h_{t-1} is the previous block output, x_t is the input sequence; W_{hf} and W_{xf} represent the weight matrices for the output vector of precedent cell and input vector of current cell, respectively, in the forget gate. σ is the sigmoid function given by

$$\sigma(x) = (1 + e^{-x})^{-1} \quad (8)$$

The input layer determines what information can be stored in the memory cell from the current input vector as follows.

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}x_t + b_i) \quad (9)$$

The output gate layer, which decides which output can be passed in the current time step, is defined as

$$O_t = \sigma(W_{ho}h_{t-1} + W_{xo}x_t + b_o) \quad (10)$$

Integrating the forget gate and input, the current cell state is computed by

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t' \quad (11)$$

where C_t is the cell state at time step t and $\odot$ is the Hadamard product that indicates element-wise multiplication of vector. $f_t \odot C_{t-1}$ and $i_t \odot C_t'$ decide the information to be inherited from the precedent cell state and current input, respectively. Hence, C_t' is determined based on the tanh activation function.

$$C_t' = tanh(W_{hC}h_t + W_{xC}x_t + b_C) \quad (12)$$

The hidden state is determined by multiplying the output gate with the current cell state.

$$O_t = h_t \odot \tanh(C_t) \quad (13)$$

Overfitting issues can arise in DNNs with many learnable parameters, especially when they are trained on a short dataset. Because of this, the DNN model is unable to categorise the new test sample. Less training data, noisy input data, and a high dimensional input space are potential causes of overfitting in deep learning systems. As a result, the overfitting problem in the DNN model is minimised using the current regularisation strategies. The L2 regularisation and dropout techniques have been applied in this application to alleviate the overfitting problem.

An excessively complicated model may overfit more frequently, as was discussed in L2 regularisation. As a result, by deleting layers from the model, we may immediately lower its complexity and shrink the size of the suggested model. By reducing the number of neurons in the fully connected layers, complexity may be further diminished. Additionally, the model should be sufficiently complicated to strike a balance between underfitting and overfitting for this purpose.

3.3.2. CNNs with Modified Baysein Optimization

CNNs are the most effective DLTs and use input, hidden layers, and output layers. Layers are composed of synthetic neurons that mimic brain neurons. The corresponding weights of neurons enable data to pass via input hidden layers to output layers (Chauhan, Ghanshala, & Joshi, 2018). These weights are regularly adjusted using activation functions that take sums of input weights as inputs. Networks repeat themselves to reduce mistake. A deep neural network is created by including more hidden layers. But unlike a conventional ANN, CNNs can accept full pictures as input, and they scale effectively (Figure 4).

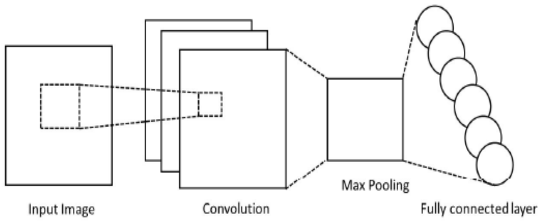

Figure 4. **CNNs**

The bulk of the layers in a CNN's design are the input layer, the convolution layer, the rectified linear units (ReLU) layer, the pooling layer, and the fully connected layer. It is common practise to utilise an input image that has the following dimensions: height, width, and number of channels. For instance, an RGB image has three colour channels. Convolution layers are the core elements of CNNs because they contain convolution filters, also called as kernels. Each of these filters creates an activation function that responds to particular components like edges and colours by convoluting the entire image. The ReLU activation function is then often utilised by a ReLU layer to accelerate the training process. In order to avoid overfitting, the pooling layer gradually downsamples the input picture by removing unnecessary data. Convolution and pooling layers are followed by the fully linked layers. To detect huge patterns, all the neurons in these layers are linked to all of the activation mechanisms from the layer before. The appropriate class is determined by feature combination in the top layer. However, based on the application or data, its design may vary. Thus, networks may consist of 1 or 2 convolution layers or could also be complex network with hundreds of fully linked convolution layers.

- **Bayesian Optimizations**

Bayesian optimizations are efficient MLTs and best choices for an expensive objective functions. By using a probabilistic model, the black-box method of Bayesian optimization seeks to reduce or maximise any given objective function (Frazier, 2018). A loss function plus a probabilistic model make up a Bayesian optimization. It seeks to describe the distribution of the goal function f by modelling it.

$$x_{new} = arg_x X^{max/min} f(x) \quad (14)$$

where X represents any particular design space of interest. This sequentially updated model is employed for generating effective sampling selections. It also uses an acquisition function to keep up capabilities for both exploration and exploitation. The next selection is made using the relevant candidates chosen by the acquisition function. Many times, a Gaussian process is chosen to get some of the objective function's necessary parameters. The loss function illustrates the running sequence's efficiency.

In BO, a Bayesian prior distribution that assesses policy performance encodes the objective function's uncertainty. Since the approach is Bayesian, the estimated values' explicit encoding of uncertainty. A posterior distribution across the objective is generated following policy executions, and this posterior is utilised to direct the exploration procedure. The generalisation performance of Gaussian process models, and as a result, the performance of the BO technique, are significantly influenced by the definition of both the GP mean function and the kernel function storing relatedness between points in the function space.

3.3.3. Gaussian Distribution process

The effectiveness of BO as a Bayesian approach greatly depends on the calibre of the modelling effort. The nature of the posterior and, thus, the generalisation capabilities of the surrogate representation are determined by the prior distribution's specification. We decide to use the Gaussian process to model the objective function

$$\eta(\theta) \sim GP(m(\theta), k(\theta, \theta')) \quad (15)$$

A mean function $m(\theta)$ and a covariance function $k(\theta, \theta')$ create GP models. The anticipated value at a given location θ and θ' is specified by the mean function, $m(\theta) = E[\eta(\theta)]$ Similar to this, the covariance function calculates the covariance as $k(\theta, \theta') = E[\eta(\theta) - m(\theta)\eta(\theta') - m(\theta'))]$ The correlation between the objective values at locations and'is represented by the kernel function. The information about the underlying class of functions is encoded by both of these functions.

The posterior distribution at additional sites must be determined in order to determine the improvement function mentioned above. This posterior has a simple form in the GP model. The conditional posterior distribution is Gaussian with mean given the data D1:n.,

$$\mu(\eta(\theta_{n+1})|D_{1:n}) = m(\theta_{n+1}) - k(\theta_{n+1}, \theta)K(\theta, \theta)^{-1}(y - m), \quad (16)$$

where m is a vector of size n with elements m(θ_1),...... m(θ_n) and variance,

$\sigma^2(\eta(\theta_{n+1})|D_{1:n}) = k(\theta_{n+1}, \theta_{n+1}) - k(\theta_{n+1}, \theta)^t K(\theta, \theta)^{-1}(y - m)$ (17)

Define y to be the column vector of observed performances such that $y_i = \eta(\theta_i)$. Define $K(\theta, \theta)$ to be the covariance matrix with elements $K_{i,j} = k(\theta_i, \theta_j)$.Define $k(\theta_{n+1}, \theta)$ to be the column vector of correlations such that the ith element is $k(\theta_{n+1}, \theta_i)$ $k(\theta_{n+1}, \theta)^t$is the transpose of this vector). So the proposed hybrid classification model efficiently identify the gait based PD.

4. Results and Discussion

Gait patterns are collections represented by 18 columns of data, containing 16 VGRF sensor signals and 2 cumulative values from left and right foot sensors. About 13,000 examples for a single force-sensitive resistor may be found in the PD dataset. during a period of 2 minutes of gait recording (FSR). Since there are 173 patients for whom the sensor data from the VGRFs are available, there are a total of 40,482,000 gait samples (18 x 13,000 x 173). First, the first 10 seconds of data and the last 20 seconds of data are deleted to lessen the impacts of gait commencement and end-up. The dataset link is given as https://nist.mni.mcgill.ca/multi-contrast-pd25-atlas/. In the context of a data mining issue, the following definitions apply to the entries in the confusion matrix: The model identified four incorrect predictions: the correct negative prediction, also known as true negative (TN), as failed; the incorrect positive prediction, also known as false positive (FP), as passed; the incorrect negative prediction, also known as false negative (FN), as failed; and the correct positive prediction, also known as true positive (TP), as passed. The following formulas are used to determine the performance measures based on this confusion matrix.

Precisions are ratios of correctly found positive observations to total expected positive observations.

$Precision = TP/(TP + FP)$ (18)

Sensitivities or Recalls are defined as ratios of correctly identified positive observations to over-all observations.

$Recall = TP/(TP + FN)$ (19)

F – measures are defined as weighted averages of Precisions and Recalls and hence takes false positives and false negatives.

$F - measure = 2 * (Recall * Precision)/(Recall + Precision)$ (20)

Accuracies are computed in terms of positives and negatives as follows:

$Accuracy = (TP + FP)/(TP + TN + FP + FN)$ (21)

Table 1. **Comparison table between the proposed and existing methods**

Performance metrics	MC-SVM	CNNs	1D-Convnet	LSTM	HCNN-LSTM
Accuracy	71.41	83.54	89.26	93.9	95.67
Precision	69	76	84	91	97
Recall	74	79	85	91	93.1
F-measure	67	76	81	94	97.5

The table 1. tabulates the performance comparison values between the proposed and existing methods.

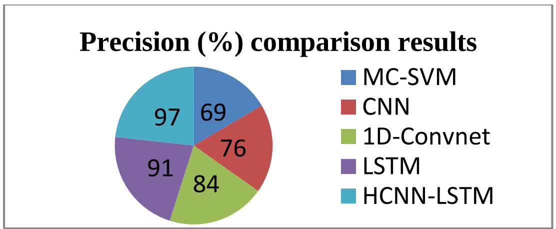

Fig. 6. **Results of a precision comparison of the proposed and current methods for identifying data related to PD**

The accuracy comparisons between suggested and existing methods for categorising PD are shown in Fig. 6. Overall, the results showed that the HCNN-LSTM model performed better on the provided datasets when compared to MLTs under consideration. These results are consistent with the earlier error rate and can be attributed to the rule sets produced by the proposed classification model. According to the findings, the suggested HCNN-LSTM strategy outperforms other current classification methods in terms of precision.

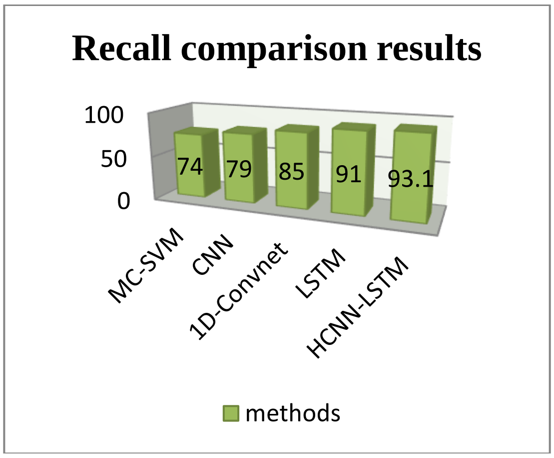

Fig.7 **shows the classification of data for Parkinson disease using the proposed and existing methods and the recall comparison findings**

The memory comparison of the proposed and current methods for categorising the Parkinson disease data is shown in Fig. 7. The data utilised in this study primarily focuses on classifying people who have symptoms of Parkinson disease with a number of characteristics that often influence the diagnosis. As a result, the prediction model is viewed as a classification issue that arises from having PD or not. As a consequence, the suggested supervised models were used for the assigned task, and the outcomes were examined and assessed.

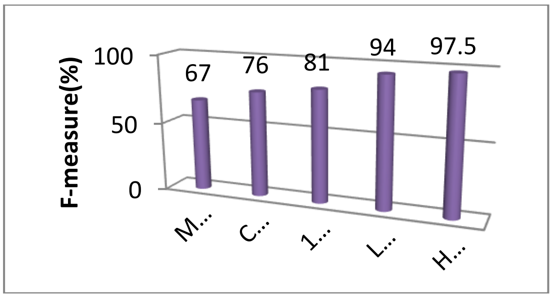

Fig. 8. **The suggested and current methods for categorising the data for Parkinson disease were compared using the F-measure.**

The F-measure comparative values of suggested and existing methods for identifying data on PD are shown in Fig. 8 where it is demonstrated that the suggested model has the greatest accuracy rate measurement in each database when compared to other MLTs,. When compared to the other approaches, the utilised database yields the best f-measure findings. It can be seen from the graph that the proposed HCNN-LSTM model outperforms the current approaches in terms of f-measure.

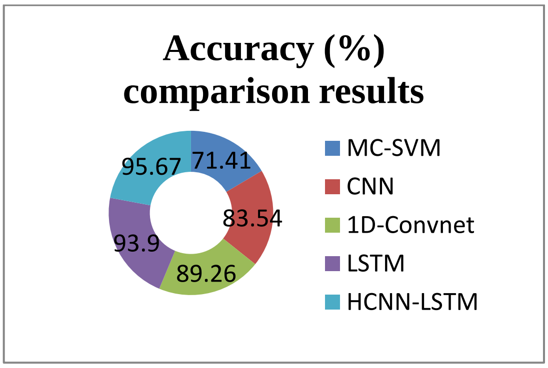

Figure 9 **shows accuracy comparison findings for the proposed and current methods of identifying data related to PD**

The accuracy comparisons between the suggested and existing methods for categorising PD are shown in Fig. 9. A supervised MLT that can accurately forecast the goal and generalise new instance predictions is considered successful. Typically, accuracy has two kinds, sensitivity and specificity which may be used to determine the validity of a model. According to the simulation findings, the suggested HCNN-LSTM model has a high accuracy rating of 95.67%, compared to the current LSTM model's 93.9%, the 1D-Convnet model's 89.26%, the CNNs model's 83.54%, and the MC-SVM model's 71.41%. As a consequence of the findings, it can be said that the suggested HCNN-LSTM approach provides good accuracy outcomes when compared to existing classification techniques.

5. Conclusion

In this study, hybrid classification models are created, examined, and found to be effective at categorising gait impairment in PD patients. This study employed VGRFs gait time series datasets to discriminate between PD patients and healthy controls. The HCNN-input LSTM's layer takes in the pre-processed, segmented data from VGRFs. Gait cycle contains significant spatiotemporal features that can differentiate between persons with PD and those who are healthy. LRPs are used to interpret the results of the models and provide information about the features of the spatiotemporal gait GRF signals that are most crucial for the accuracy of the algorithmic predictions. The PD data predictions made by the hybrid CNNs with LSTM are accurate, and they operate quickly. The models outperform prior manual feature selection methods and show resilience to perturbation noise, overcoming individual differences in stride length. According to the LRPs research, body balance, which refers to the extent to which a patient's condition impairs their ability to walk without falling, is a key indicator for the diagnosis of Parkinson's disease (PD). Additionally, whether making diagnostic conclusions based on visual observation criteria or bespoke features built from data from other GRFs sensors, healthcare professionals may find it helpful to identify the most significant gait cycle events. In terms of future approaches, the methods put forward in this paper can assist in creating a plan for individually tracking the severity of PD as it progresses.

References

[1] Abou, L., Peters, J., Wong, E., Akers, R., Dossou, M. S., Sosnoff, J. J., & Rice, L. A. (2021). Gait and Balance Assessments using Smartphone Applications in Parkinson's Disease: A Systematic Review. *Journal of medical systems, 45*(9), 1-20. https://doi.org/10.1007/s10916-021-01760-5

[2] Balaji, E., Brindha, D., Elumalai, V. K., & Umesh, K. (2021). Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease. *Medical Engineering & Physics, 91*, 54-64. https://doi.org/10.1016/j.medengphy.2021.03.005

[3] Balaji, E., Brindha, D., Elumalai, V. K., & Vikrama, R. (2021). Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Applied Soft Computing, 108*, 107463. https://doi.org/10.1016/j.asoc.2021.107463

[4] Binder, A., Bach, S., Montavon, G., Müller, K.-R., & Samek, W. (2016). Layer-wise relevance propagation for deep neural network architectures. In *Information science and applications (ICISA) 2016* (pp. 913-922). Springer. https://doi.org/10.1007/978-981-10-0557-2_87

[5] Borzì, L., Mazzetta, I., Zampogna, A., Suppa, A., Olmo, G., & Irrera, F. (2021). Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. *Sensors, 21*(2), 614. https://doi.org/10.3390/s21020614

[6] Camps, J., Samà, A., Martín, M., Rodríguez-Martín, D., Pérez-López, C., Alcaine, S., Mestre, B., Prats, A., Crespo, M. C., & Cabestany, J. (2017). Deep learning for detecting freezing of gait episodes in Parkinson's disease based on accelerometers. In *International Work-Conference on Artificial Neural Networks* (pp. 344-355). Springer. https://doi.org/10.1007/978-3-319-59147-6_30

[7] Celik, E., & Omurca, S. I. (2019). Improving Parkinson's disease diagnosis with machine learning methods. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1-4). Ieee. https://doi.org/10.1109/EBBT.2019.8742057

[8] Chauhan, R., Ghanshala, K. K., & Joshi, R. (2018). Convolutional neural network (CNN) for image detection and recognition. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 278-282). IEEE. https://doi.org/10.1109/ICSCCC.2018.8703316

[9] Dash, A. K. (2021). Prediction of Parkinson's Disease using Hybrid Feature Selection based Techniques. In *2021 4th International Conference on Computing and Communications Technologies (ICCCT)* (pp. 46-51). IEEE. https://doi.org/10.1109/ICCCT53315.2021.9711843

[10] Demrozi, F., Bacchin, R., Tamburin, S., Cristani, M., & Pravadelli, G. (2019). Toward a wearable system for predicting freezing of gait in people affected by parkinson's disease. *IEEE journal of biomedical and health informatics, 24*(9), 2444-2451. https://doi.org/10.1109/JBHI.2019.2952618

[11] El Maachi, I., Bilodeau, G.-A., & Bouachir, W. (2020). Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Systems with Applications, 143*, 113075. https://doi.org/10.1016/j.eswa.2019.113075

[12] Eltoukhy, M., Kuenze, C., Andersen, M. S., Oh, J., & Signorile, J. (2017). Prediction of ground reaction forces for Parkinson's disease patients using a kinect-driven musculoskeletal gait analysis model. *Medical engineering & physics, 50*, 75-82. https://doi.org/10.1016/j.medengphy.2017.10.004

[13] Frazier, P. I. (2018). Bayesian optimization. In *Recent advances in optimization and modeling of contemporary problems* (pp. 255-278). Informs. https://doi.org/10.1287/educ.2018.0188

[14] Johri, A., & Tripathi, A. (2019). Parkinson disease detection using deep neural networks. In *2019 Twelfth international conference on contemporary computing (IC3)* (pp. 1-4). IEEE. https://doi.org/10.1109/IC3.2019.8844941

[15] Mazilu, S., Calatroni, A., Gazit, E., Roggen, D., Hausdorff, J. M., & Tröster, G. (2013). Feature learning for detection and prediction of freezing of gait in Parkinson's disease. In *International workshop on machine learning and data mining in pattern Recognition* (pp. 144-158). Springer. https://doi.org/10.1007/978-3-642-39712-7_11

[16] Mirelman, A., Bonato, P., Camicioli, R., Ellis, T. D., Giladi, N., Hamilton, J. L., Hass, C. J., Hausdorff, J. M., Pelosin, E., & Almeida, Q. J. (2019). Gait impairments in Parkinson's disease. *The Lancet Neurology, 18*(7), 697-708. https://doi.org/10.1016/S1474-4422(19)30044-4

[17] Naghavi, N., Miller, A., & Wade, E. (2019). Towards real-time prediction of freezing of gait in patients with Parkinson's disease: addressing the class imbalance problem. *Sensors, 19*(18), 3898. https://doi.org/10.3390/s19183898

[18] Nguyen, H., Lebel, K., Bogard, S., Goubault, E., Boissy, P., & Duval, C. (2017). Using inertial sensors to automatically detect and segment activities of daily living in people with Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26*(1), 197-204. https://doi.org/10.1109/TNSRE.2017.2745418

[19] Pardoel, S., Kofman, J., Nantel, J., & Lemaire, E. D. (2019). Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: a review. *Sensors, 19*(23), 5141. https://doi.org/10.3390/s19235141

[20] Pardoel, S., Shalin, G., Lemaire, E. D., Kofman, J., & Nantel, J. (2021). Grouping successive freezing of gait episodes has neutral to detrimental effect on freeze detection and prediction in Parkinson's disease. *Plos one, 16*(10), e0258544. https://doi.org/10.1371/journal.pone.0258544

[21] Patro, S., & Sahu, K. K. (2015). Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*. https://doi.org/10.48550/arXiv.1503.06462

[22] Torvi, V. G., Bhattacharya, A., & Chakraborty, S. (2018). Deep domain adaptation to predict freezing of gait in patients with Parkinson's disease. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1001-1006). IEEE. https://doi.org/10.1109/ICMLA.2018.00163

[23] Vidya, B., & Sasikumar, P. (2021). Gait based Parkinson's disease diagnosis and severity rating using multi-class support vector machine. *Applied Soft Computing, 113*, 107939. https://doi.org/10.1016/j.asoc.2021.107939

[24] Xia, Y., Zhang, J., Ye, Q., Cheng, N., Lu, Y., & Zhang, D. (2018). Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients. *Biomedical Signal Processing and Control, 46*, 221-230. https://doi.org/10.1016/j.bspc.2018.07.015

[25] Yang, Y., Wang, Y., Zhou, Y., Chen, C., Xing, D., & Wang, C. (2014). Validity of the Functional Gait Assessment in patients with Parkinson disease: construct, concurrent, and predictive validity. *Physical therapy, 94*(3), 392-400. https://doi.org/10.2522/ptj.20130019

[26] Zhao, S., Rudzicz, F., Carvalho, L. G., Márquez-Chin, C., & Livingstone, S. (2014). Automatic detection of expressed emotion in Parkinson's disease. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4813-4817). IEEE. https://doi.org/10.1109/ICASSP.2014.6854516

[27] Zheng, X., Chen, W., You, Y., Jiang, Y., Li, M., & Zhang, T. (2020). Ensemble deep learning for automated visual classification using EEG signals. *Pattern Recognition, 102*, 107147. https://doi.org/10.1016/j.patcog.2019.107147