Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Neurological disease prediction using impaired gait analysis for foot position in cerebellar ataxia by ensemble approach

M. Shanmuga Sundari & Vijaya Chandra Jadala

AUTOMATIKA
2023, VOL. 64, NO. 3, 540–549
https://doi.org/10.1080/00051144.2023.2194097

REGULAR PAPER
OPEN ACCESS Check for updates

Neurological disease prediction using impaired gait analysis for foot position in cerebellar ataxia by ensemble approach

M. Shanmuga Sundari and Vijaya Chandra Jadala

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswarm, India

ABSTRACT
In neurological field, Cerebellar Ataxia (CA) prediction is done with Gait values of human actions. The Analysis of Gait (AoG) may lead the good treatment. The goal of this work was to develop a machine-learning-based model for predicting AoG using the poor gait patterns that occur before AoG. While executing designed AoG-provoking walking tasks, an accelerometer was connected to the lower back of 21 subjects with 12 different walking positions to gather acceleration impulses. The exercise was walking for one minute at each of 12 varied walking speeds on a split-belt treadmill in the range [0.6, 1.7] m/s in 0.1 m/s increments. To reduce the effects of weariness, the speed sequence was randomized and kept a secret from the subjects. Machine-learning algorithms like support vector machine (SVM) and k-nearest neighbours (KNN) have been tested in existing research studies. These algorithms perform well when the amount of data is little and the classification is binary. SVM, KNN, decision trees, and XGBoost algorithms have all been used in the proposed study on the CA data set. We discovered that the AdaBoost algorithm provides a more accurate categorization of the severity of CA disease.

ARTICLE HISTORY
Received 21 January 2023
Accepted 17 March 2023

KEYWORDS
Adaboost; cerebellar ataxia; neurological disease; machine-learning algorithms

1. Introduction

Although the motor symptoms of cerebellar ataxia (CA) are the most well-known, numerous non-motor symptoms have also been reported [1]. Irregular actions and the inability to suppress urges are hallmarks of the psychiatric disorders known as impulse control disorders (ICDs). A well-known area of medical specialization is neurological specialization [2]. The brain instructs the body on how to respond to events. Using this research, we can pinpoint the activity issue and determine the nervous system's capacity. A disruption in a person's activity rhythm may result in neurological diseases. Brain, spine and nerve damage are the focus of neurosurgery. Our specialists use neurosurgery to treat neurological diseases. Finding activity patterns [3] in the medical field is difficult. We must observe the patient's motions in order to pinpoint the condition [4]. Issue identification and pinpointing the issue is very difficult in the early stage of neuro disease. A patient's death could occur due to any failure in their medical care.

In cross-sectional investigations, the prevalence of ICDs and related diseases ranged from 15% to 20% [5–7]; annual incidence was estimated to be around 10% [8,9], and after five years of the disease's occurrence, the overall incidence reached over 50% [10]. These issues can also affect PD individuals whose disease has been present for longer than five years.

Neurologists will evaluate the complexity of disorders related to the nervous system. It helps to assess the disease's severity, which influences treatment. For humans, the capacity of the brain is crucial in terms of all activities. The activity pattern shows the severity of the CA condition. Recent advancements in neurosurgical critical care and neuroimaging technologies enable early patient treatment with minimally invasive techniques, which leads to death and treatment is also very tough. An interdisciplinary team [11] that consists of various neuro specialists is working towards this aim.

One of the most prevalent and incapacitating symptoms of cerebellar patients is an analysis of gait (AoG) [12]. AoG is a severe kind of gait dysfunction that is characterized by "despite the intention to walk, there is a brief, episodic lack of forward motion or a significant decline" [13]. AoG is more common when turning, travelling through confined places and starting a gait. AoG severely limits mobility, increases the risk of falling and lowers the quality of life [14–16].

The variety of diseases also increased at the same time as medical remedies advanced quickly. The improvements in medicine have raised standards in society. A neurological condition called CA causes improper coordination of body movements. Exercise coordination issues will be influenced by muscle power. When the magnitude and synchronization of the limbs to maintain posture are compromised, it will occur.

CONTACT M. Shanmuga Sundari mshanmugasundari@gmail.com Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswarm, Andhra Pradesh, India

They become physically unbalanced and are unable to perform further tasks. If the condition is discovered in the early stages, treatment is simpler. Interpreting ensembles might be more challenging. Even the finest ideas do not always succeed in persuading decision-makers. Even the finest concepts are occasionally rejected by the intended audience. Lastly, ensembles are more expensive to develop, train and use.

The health care sector greatly benefits from artificial intelligence [17]. By accurately anticipating the disease, it helps save the lives of countless people. On the activity data, numerous algorithms are performed, and the results are then examined. The result of the analysis is trustworthy and precise for forecasting. For the analysis of the necessary disease prediction, we employ the support vector machine (SVM) [18], decision tree (DT) [19], and k-nearest neighbour (KNN) [20] algorithms. To obtain more precise predictions from the machine-learning methods, we apply the ensemble technique [21].

These characteristics' combined ability to predict outcomes has not been thoroughly studied. Only three research [22] and patient-level forecasts have been reported. Researchers used a logistic regression using neuro-clinical and genetic data in all three studies and then utilized the receiver operating characteristic (ROC) curve to determine how well it predicted outcomes (ROC AUC). The accuracy of the performance outcomes was impacted by the absence of cross-validation or a replication cohort in any of these experiments [23].

We use machine-learning approaches to forecast the neuro-clinical data. In order to train and cross-validate the obtained models and see if they could be applied to the other cohort, we used two longitudinal cohorts. With knowledge of the patient's clinical history and genotyping information, the goal was to estimate the risk of ICDs at the subsequent appointment. By combining numerous models rather than relying just on one, ensemble approaches seek to increase the accuracy of findings in models. The integrated models considerably improve the findings' accuracy. Because of this, ensemble approaches in machine learning have gained prominence.

The paper residue has the following structure: We covered the literature evaluation and the drawbacks of works in Section 2 that are linked. The data set and experimental setup are described in Section 3. Section 3 presents a mathematical analysis of machine learning. The ensemble approach is described in Section 4. In Section 5, the problem statement is defined. Section 6 is a report on the experimental analysis. Sections 7 and 8 provide an explanation of the comparison study and outcome analysis. Section 9 talks about the results and upcoming projects.

2. Literature survey

On the alumni data set, the DT technique was applied by Daniela Alejandra Gomez Cravioto et al. [24]. The alumni data set was analysed and categorized using the WEKA tool. The tool for machine-learning algorithms is WEKA. Based on the choice of data set, it provides the result using established algorithms. They used the CRISP-DM approach and the WEKA tool. They contrasted different machine-learning methods, including all DT algorithms. The best method for creating a classification system is determined by comparing machine-learning algorithms like DT. The analysis of the alumni data set revealed that the random forest algorithm outperformed other algorithms.

The most common neurological condition that affects the central nervous system is Parkinson's disease (PD). The number of its sufferers has increased significantly, especially in underdeveloped countries. Trembling, diminished mental reaction and poor posture are the earliest signs of PD. Almost ten million individuals have been diagnosed with it as of this point, and it is a terrible medical ailment that is common in developed and developing countries. The disease's primary origin is still unknown, however, based on the indications and symptoms it exhibits, it can be treated if caught in its early stages. There is currently no known cure or preventative measure for PD, and it is unclear if the condition is genetic or natural. Many clinical and blood investigations have been provided to aid in the PD diagnosis. Correctly diagnosing PD can be challenging, especially in the early stages. Occasionally medical professionals will request blood testing or brain scans to rule out the potential of other illnesses.

The logistic regression (LG), Naive Bayes (NB), stochastic gradient descent (SGD), adaptive boosting, bagging, DT, and random forest (RF) classifier are eight machine-learning methodologies that the author applied to assess analysis utilizing the WEKA tool [25]. They anticipated the advent of cancer sickness. The curve displayed the applicable algorithms' anticipated rate. Random forest provided the most accurate outcome in this study. The confusion matrix and other indicators were used to calculate the values. Protein and blood metrics are just a few examples of patterns that can be recognized using machine learning. These are the contributing elements to COVID- 19 [26] severe forecast.

In their study, the author used the Gini algorithm to forecast migraines and the intensity of migraines. The feature was recorded using artificial neural network ideas. The dispersed delay values were recorded using vector data. These tactics' effectiveness has been assessed against the results of earlier study models. In this work, the author used aged people and a machine-learning approach to predict how headaches will turn out [27]. The model's sensitivity displayed the optimum value in the resultant table for diseases with low prevalence. Machine-learning researchers can choose and sample features with the help of the WEKA tool.

The various ensemble approaches were discovered by the author [28]. They applied bagging, stacking and voting techniques using different machine-learning methods. They used ensemble learning techniques to propose a number of promising frameworks. The prediction metrics are AUC, accuracy, F-measure, recall, precision and MCC. The author presented the predicted values in the conclusion part.

The heel-knee-shin test and the rapid alternating movements test were two of the tests used by the author [29] as she focused on limb motor function assessment. The following traits are present in the daily activities that have an impact on the limb joints. The human activities are calculated using repetition and rhythm. The person needs to keep a straight posture in order to carry out these responsibilities.

They received a lot of experimental data from these body sensor networks. These complicated data also include patterns of human behaviour and environmental information [30]. Activity mining is the most effective method for developing efficient human activity patterns. The usage of utility pattern mining (UPM) is widespread among organizations and applications. This study employed the wireless multipath propagation of signals to examine human patterns.

Wireless multipath signal transmission [31] in confined environments, which includes direct radiation, reflection, diffraction and scatter propagation, is the basis of wireless sensor systems.

Multipath propagation is used to overcome interior walls, flooring, furniture and other obstacles. The pattern of wireless signal transmission is always changing, and this is reflected in human behaviour [32]. It alters the wireless signal's characteristics after being received. Human actions may result in multipath environmental information depending on human behaviour.

The importance of studying fundamental neuroscience cannot be emphasized enough. Due to the complexity of the nervous system, a problem's diagnosis [33] might be challenging. Additionally, the medical community observes novel virus varieties for which a cure is required. We need to provide patients with faster care in today's society. Many researchers used Bayesian quadratic discriminant analysis to test the kinematic and diagnostic aspects of CA patients. To anticipate the disease, the author calculated finger chase movement values. The study clarified how machine-learning methods using density-based mining and root mean value probability work. Backpropagation [34] was employed in the analysis to locate and create the pattern.

The random forest algorithm's k mean value is used to calculate the human activity pattern. The author employed classifiers to identify the deviation after taking human activity into account as the pattern. To comprehend the pattern styles, they used various graphical data. Bitmaps were employed in UPM, and the output was shown as visual graphs. The majority of instances of ALS are caused by a deadly neurological ailment that is sporadic, and this chronic disease's origin will evolve at different phases. It is quite challenging to pinpoint the mutation's genesis.

Machine learning aids in the early diagnosis of disease. Based on an action, the process of mining [35] is evaluated. The technique for coordinating the order of operations is called activity mining. The movement is tracked and recorded to create a data collection. The author employed the truncal and collateral measures related to signal values. Machine learning gives the better accuracy for the prediction problems.

The Naive Bayes algorithm [36] was employed by the author to forecast current events. On the activity pattern data set, SVM activities are utilized to forecast the values. The metrics were calculated and assessed using this method. Every step of the process was calculated and measured using the process mining concept.

The geographical approach and values were used by the author to assess the road colocation pattern [37]. The author used an artificial neural and collocation path to find an activity forecast for the assessment of his project. For humans, the brain acts as the major role. The body as a whole is activated using the brain as a model. The actions are well-synchronized and cooperate with the brain. The pattern displays negative numbers when the actions fail.

The author employed a variety of machine-learning approaches to conduct misleading phishing attacks that sought out sensitive data. The spoofed emails and actual messages are recognized as spam and input mails, respectively. The emails contained [38] threats, which they classified using computer learning techniques.

The approaches employed in the pertinent principles and drawbacks are explained in Table 1.

3. Process flow of disease prediction

This paper aims to present a revolutionary disease prediction method [39] with five key phases, including

- data preprocessing,
- noisy data reduction,
- extract feature,
- select the proper feature, and
- final classification.

Table 1. Literature survey concepts.

Ref.	Approaches	Methods used	Drawbacks
[1]	IOT devices	Microprocessor unit, Matlab	Low accuracy using sensors
[2–4]	Machine learning	SVM, KNN, DT	It was only applicable to one subject and did not apply to several ailments
[5]	Machine learning	Classification algorithms	A single notion was used to identify the best classifier
[6,7]	Machine learning, Ensemble method	AdaBoost	This study established the effectiveness of the Gaussian-RBF kernel. However, this is not used in any programme
[8]	Machine learning	SVM and DT	Small database used
[9–12]	Machine learning	Classification algorithms	They only employed one sample, and in real-time circumstances, classification did not flow dynamically
[13,14,15]	Machine learning	Tree structured algorithm	Decision-making processes are lacking in some subjects that this research needs to include
[18,23,25]	Machine learning, Ensemble method	SVM, encryption and fuzzy logic system	Lacking to make grouping and integrate the self-adaptive features
[28–30]	Activity process mining	Alpha algorithm	Real-time dataset is missing
[31,32,38]	Machine learning	Naive Bayes	This study did not look into concerns with development and mobility

- The preparation phase [40] receives the input data initially, and it is during this phase that the normalization of data process is carried out.
- The imbalance processing phase, which implements an enhanced a plan of action to address the class imbalance, is then completed.
- Following the resolution of the imbalanced problem, the features are retrieved, comprising the original feature, greater statistical features, entropy, correlations, enhanced similarity measurement and statistical features.
- Additionally, in the feature selection phase, relevant features are chosen from the extracted features, and this is done using an improved relief technique.
- These chosen characteristics are used in the ensemble classifiers, which comprise the NN, RNN, RF and k-NN models, during the classification phase.
- The output of NN, RNN and RF serves as the input for k-NN in this scenario. The suggested new models perfectly calibrate the NN and RNN weights to improve the system's ability to anticipate diseases.
- The ultimate result is then efficiently and precisely obtained.

Figure 1 explains the overall the block diagram flow in the system. So the binary classification of disease is got using the above model.

4. Processes of disease prediction

We used CA data parameters as the training (discovery) cohort. We used cross-validation technique [41], which is shown in Figure 1, to objectively measure the performance analysis of the models. We divided the PPMI participants at random, placing 80% in the training set and the remaining 20% in the testing sample for the outer loop. We used a five-fold cross-validation approach in the inner loop to enhance the model hyper-parameters on the training set. The algorithms' ability to fit the training set of data is controlled by these hyper-parameters.

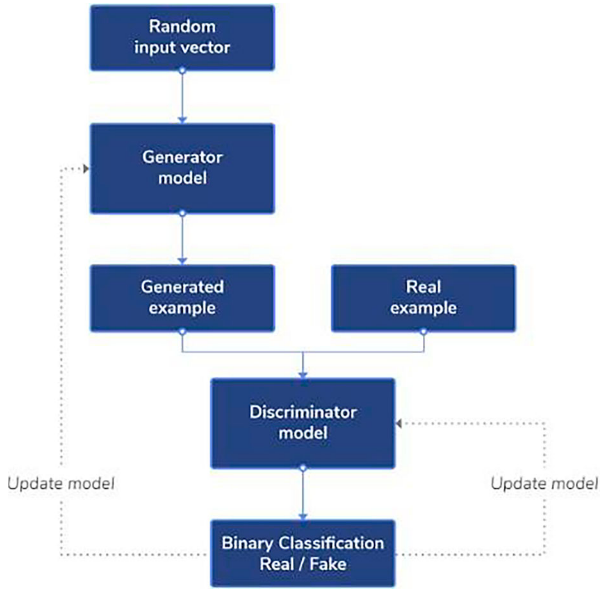

Figure 1. Architecture of the prediction.

4.1. Proposed system

The proposed system was used to process genomic data and extract variations of interest that met inclusion requirements. Using the Python packages pandas [42] and NumPy [43], the various text-like files were processed. Forward-filling was the technique used to impute missing values, and it required using the most current non-missing values for the attribute and object in question. The baseline values from the training set were used to impute baseline missing values. Because it can be used on any scale without any training and just a tiny portion of the data from a small selection of variables was ascribed, we chose to employ this straightforward method.

A linear combination is considered convex if the weights all add up to one and are not negative. The weights [44] show the relative importance of each calculation. When a visit is given a weight of 1, then a visit is given a weight of 1, the "summary" visit is just the baseline visit. Otherwise, it is the most recent visit. The most recent visits can be given higher weights if they are thought to be more significant than previous visits, or identical weights might be applied so each visit proportionally counts to this "summary" visit.

For the outer loop, we randomly assigned 70% of the subjects to the training set and the remaining 30% to the test set. To improve the model hyper-parameters on the training dataset, we employed a five-fold subject-level cross-validation strategy in the inner loop. The degree to which the algorithms adhere to the training set of data is controlled by these hyper-parameters.

We divide 70–30% training and test data. We used five-fold cross-validation approach for enhancing the parameters. The inner loop is used to create a model for making them to regulate the algorithms. The hyper-parameters are considered as the training level models.

Individual differences in walking patterns [45], the impact of bodily sensations [46] or various disease states could all be contributing factors to the variation in prediction accuracy. The MI scores can be used to identify various walking patterns, with cadence being the most useful characteristic for all patients save one.

4.2. Architecture of proposed system

Figure 2 explains the workflow of the project. It started with dataset collection and was followed by data cleaning which explains in Section 5. Then, the weight value is calculated and applied to different machine-learning algorithms. The resultant accuracy of prediction is the final result of our research.

5. Data preprocessing

5.1. Remove outliers

The current system is used to eliminate outliers [47] and we randomly pick a collection of authentic experimental evidence from a healthy individual in a neurological test, as well as a subcarrier. Figure 2 displays the subcarrier's signal curve, the original signal and outliers for average people. When a typical person does the Romberg test, the body shaking should be within a

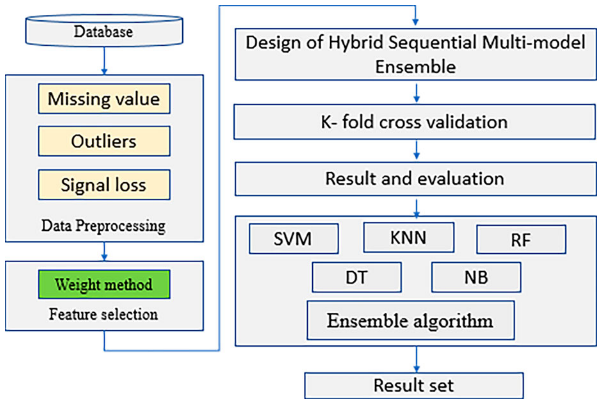

Figure 2. The roadmap of the framework.

specific scope, and the subcarrier signal curve should be steady. The signal curve has many burrs, and the volatility is high. To complete eliminating the outliers of the initial signal, we might alternatively utilize the classification function.

5.2. Denoising for signals

The noise presents the outliers in the received data that will be filtered out and have been eliminated. The majority of conventional filters [48] are linear and non-linear, including the mean filter and Wiener filter. The disadvantages of the conventional denoising method include the increase in entropy following signal translation, the inability to characterize the non-stationary features of the signal, and the failure to access the correlation of the signal. The wavelet transform is used to get around these problems.

5.3. Feature extraction

Selecting the right subcarriers is required prior to feature extraction. We are aware that more information will be contained in a batch of data when the variance is higher. We choose the 10th subcarrier of the third antenna for Romberg's test and the subcarrier of the second antenna for attribute distribution in accordance with the principle of maximum variance.

5.4. Feature extraction from ensemble test data

Normal people can maintain their balance even when they blink or close their eyes; nevertheless, small variations may occur due to environmental noise and an object's respiration. Patients with sensory ataxia who can sustain posture and balance when the blinking stage and vigorously move their bodies during eyes closing stage create a waveform that is largely stable in the blinking stage and unsteady in the closing eyes.

The attribute physical relevance is as follows. The signal's energy is reflected by the signal's mean square value, its stability is described by the signal's mean value, its degree of dispersion is reflected by the standard deviation, its impact characteristics are reflected by the signal's kurtosis, and its asymmetry is reflected by the signal's skewness.

The peak factor can be used to determine whether the signal has been impacted. The waveform factor, which has a value larger than or equal to 1, has a physical significance in the electronics sector that can be predicted as the ratio of the original AC signal to the DC signal of equal power.

The data set includes the ground reaction forces as well as the marker positions of the markers that were placed on the shoes above the first (FM1), second (FM2) and fifth metatarsal (FM5) heads and on the aspect of the Achilles tendon insertion on the calcaneus (FCC) for dynamic tracking.

The markers are named according to the foot side, anatomical position and the direction: [L/R] [Position] [x/y/z] and occur as columns in the following order: L FCC x, L FM1 x, L FM2 x, L FM5 x, R FCC x, R FM1 x, R FM2 x, R FM5 x, L FCC y, L FM1 y, L FM2 y, L FM5 y, R FCC y, R FM1 y, R FM2 y, R FM5 y, L FCC z, L FM1 z, L FM2 z, L FM5 z, R FCC z, R FM1 z, R FM2 z, R FM5 z.

The laboratory coordinate system coincided with the following anatomical directions: *x*, the posterior–anterior plane; *y*, right-to-left movement, *Z*, the downward and upward (vertical) directions.

6. Experimental algorithm

We used the same CA data set to apply the AdaBoost algorithm [49]. The AdaBoost approach employs the three classifiers mentioned above. The ensemble performed numerous evaluations and then extracted the outcome. Accuracy [50] is displayed along with the outcome as AdaBoost, which is a DT-like display. However, the other associated values outperform the other approaches.

The AdaBoost algorithm is used as an ensemble method mostly used in machine learning. Higher weights are given to situations that were mistakenly identified when the weights are reassigned to each instance. The term for this is adaptive boosting. In order to prioritize previously unclassified points by prior classifiers, AdaBoost modifies subsequent weak learners. It might occasionally be more resistant to the overfitting issue than other machine-learning methods. Individual learners might not be strong, but if the final model performs better than random guessing, it will become a potent learner.

It is an ensemble of pre-trained models. It is associated with a categorization. Every model produces predictions, often known as votes. A machine-learning model's voting method [51] involves choosing the top pre-trained models and averaging their projections to get a final prediction output.

6.1. Adaboost pseudocode

Step 1: Input dataset; Initialize base learner and number of learning iteration

Step 2: Initialize equal weight to all training samples

Step 3: Start loop to compute weight

- Train a base learner using training samples
- Compute error using Equation (1)

$$\in_t = \sum_{i=1}^{n} w_{i,t} \tag{1}$$

- Compute the weight of base learner explained in (2)

$$w_{i+1} = w_{i,t} e^{-y\alpha th_i(x_i)} \text{ for } i \text{ in } 1 \ldots\ldots. n \tag{2}$$

$$Re \text{ normalize } w_i^{-y\alpha thi(xi)} \text{ for } i \text{ in } 1 \ldots . . n \tag{3}$$

- Normalize using Equation (3)
- Set the new weight value

Step 4: Normalize and predict the value for classification

7. Statistical analysis

7.1. Training the AoG prediction model with tagging techniques

To distinguish between pre-AoG gait and regular walking, boosting of pruned C4.5 trees was chosen because it outperformed other strategies in terms of detecting AoG. Before being fed into the classifier, the features were normalized to the range [0,1]. To distinguish pre-AoG from regular walking, only features from the pre-AoG and phases of regular walking were used in the model training.

In two different schemes—one patient-dependent and the other patient-independent—we assessed and contrasted the model's performance. A 10-fold validation strategy was used for every subject in scheme 1. In each fold, 30% of the dataset was utilized for the training set and 70% for the testing set. All 12 participants in scheme 2 underwent cross-validation using the leave-one-subject-out technique

$$\text{Sensitivity} = (\text{TP}/\ (\text{TP} + \text{FN})) * 100\% \tag{4}$$

$$\text{Specificity} = (\text{TN}/\ (\text{TN} + \text{FP}) * 100\% \tag{5}$$

$$\text{Accuracy} = ((\text{TP} + \text{TN})/\ (\text{TP} + \text{TN} + \text{FP} + \text{FN})) \tag{6}$$

where TP (true positive) denotes the total number of pre-AoG windows that were successfully identified; FP (false positive) denotes the total number of normal walking windows that were incorrectly classified; FN (false negative) denotes the total number of incorrectly classified pre-AoG windows; and TN (true negative) denotes the total number of successfully identified normal walking windows explained in (4)–(6).

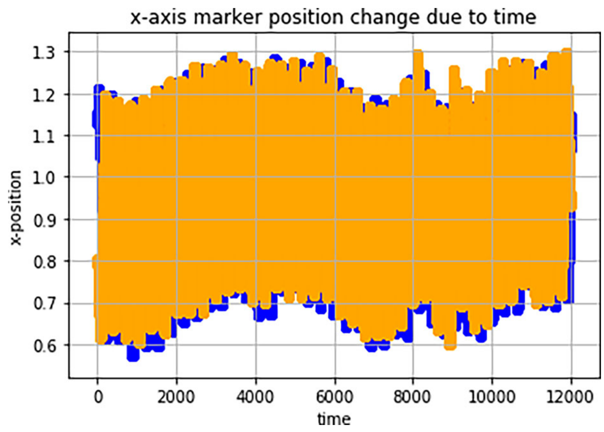

Figure 3. The gait value analysis with *x*-axis.

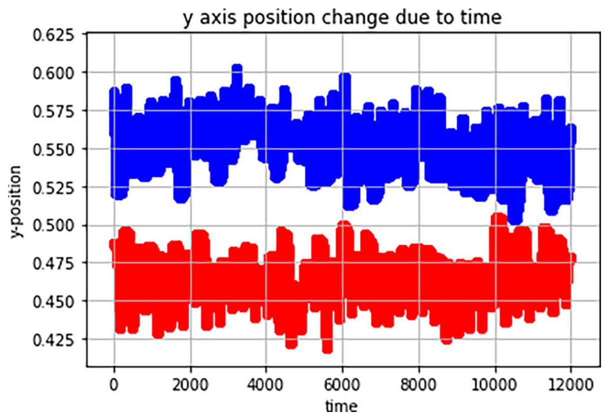

Figure 4. The gait value analysis with *y*-axis.

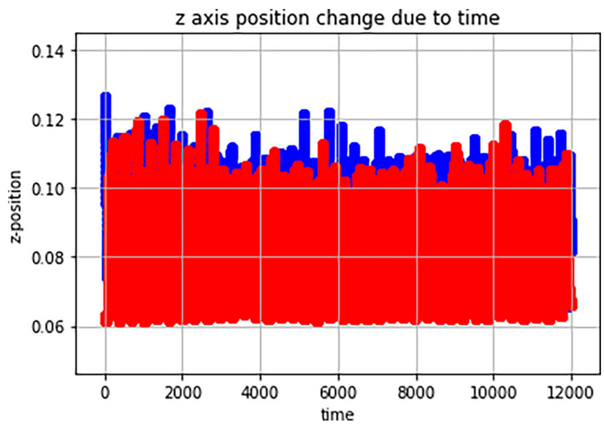

Figure 5. The gait value analysis with *z*-axis.

7.2. Conventional AoG detection features extraction and selection

Pre-stops' gait patterns were different from pre-AoGs' patterns, as was the case (Figure 4). In pre-stops, variation in the VT exhibited a trend that was slightly different from that in pre-AoG, and the cadence tended to be steadier. Similar patterns [52–54] were seen in pre-AoG periods for other parameters, but they tended to improve or drop one or two strides after and were milder. The FP rate could be decreased and pre-AoG and pre-stops could be distinguished using these various trends.

The study's main weakness is the patient-independent model's low sensitivity, which may be caused by the intrinsic differences across patients. It can be increased by enlisting more patients for the training of the AoG prediction model. Additionally, the suggested evaluation of the prediction algorithm is offline, necessitating an online experiment to confirm it and determine the impact of the predictable cues on actual AoG events.

8. Results for AoG values with coordinates

We choose sensor data to visualize with the both right and left foot. Figure 3 shows that the *x*-axis shows two different colours in the graph. The variations are shown in yellow colour. The foot position of CA patients is completely different from that of normal patients.

8.1. Y-axis of LRFM1 sensor data to see different variations of sensor data

The foot *y*-axis positions are mentioned in Figure 4 and the left leg is represented in blue colour and the right leg is shown in red colour. The coordination between the left and right legs is producing different values which are mentioned in Figure 4.

8.2. Z-axis change of LRFM2 sensor

LRFM sensor is used to collect the *z*-axis values for the left and right legs that are reflected in Figure 5. The blue colour is represented the left leg and the red colour represented the right leg. The values are captured during the time mentioned and depicted in Figure 5.

8.3. Correlation matrix for foot axis position

8.4. Visualize all axis in 3D graph

The left foot is represented with red colour and the right leg is represented with blue colour in Figure 6. The 3D visualization is coded using the matplotlib library. The CA patient's leg position has different values and gives the highest time period to keep them in the normal position. The figure above shows us how a human's right and left foot move (in 0.6 m/s, GP1).

Figure 7 shows the correlation between the left and right positions for normal people. The actual coordination should be properly aligned between both legs. When there is a mismatch in the legs axis then we have to identify and find the problem behind that. In Figure 7, both legs alignment is mentioned with the *x*, *y* and *z* axes. The alignment is clearly depicted in the graph.

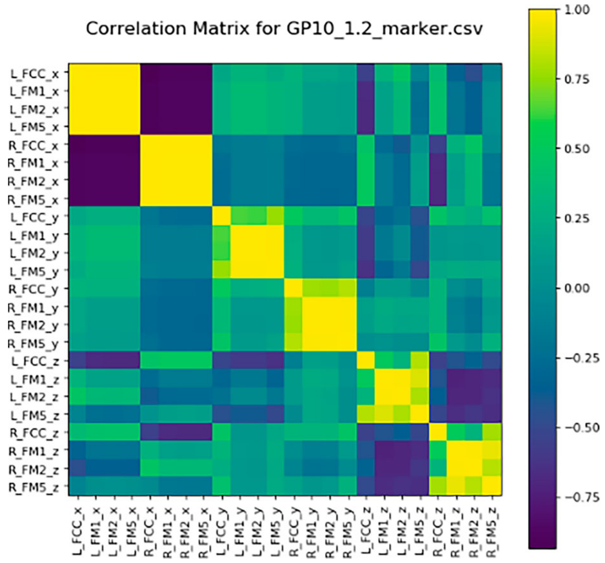

Figure 6. Correlation matrix for gait analysis.

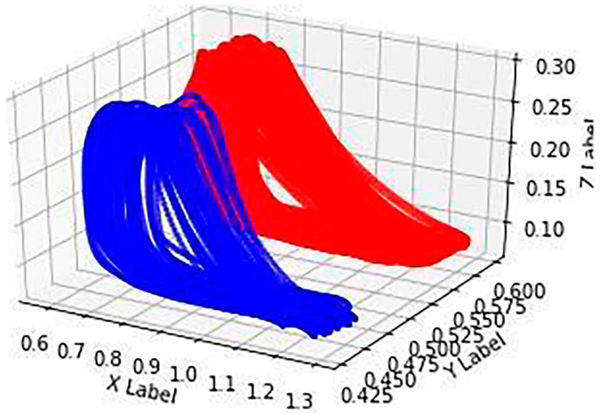

Figure 7. 3D Right-Left foot image.

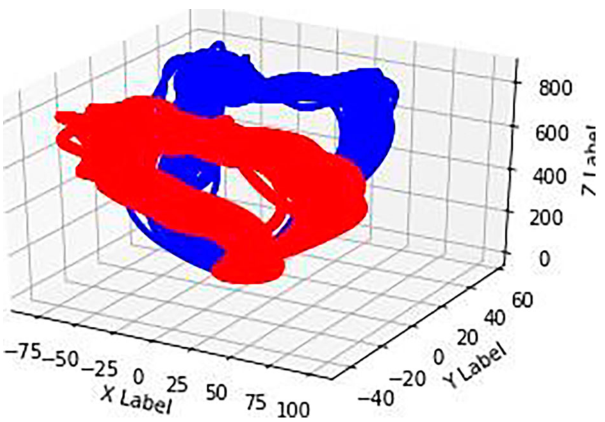

Figure 8. 3D Left-Right foot image.

Figure 8 shows that alignment changes in legs. It will be giving an idea about the patient and their disease. When the neurological disease occurs that time only the patient will suffer this kind of misalignment of the legs.

Table 2. Comparative study for prediction of CA disease.

S.No.	Algorithm	Prediction accuracy (%)
1	SVM	97.5
2	Naïve Bayes	93.8
3	Logistic regression	98.6
4	AdaBoost	99.6

9. Comparative analysis

Table 2 contains a dataset on neurological diseases. Using machine-learning algorithms, this study achieves 99.6% accuracy in identifying the activity patterns that indicate neurological disease. The dataset for CA disease produced a better prediction using Adaboost. Other measurements, including root mean squared values and error values, are used to create this forecast. In the end, this examination of different neural network techniques yields a better prediction answer. AdaBoost, an exciting one, is used in the comparative analysis. This technology's strength is its ability to identify CA illness at an early stage, which will benefit physicians.

10. Conclusion

Neurological conditions like sensory ataxia and CA have an important impact on the quality of life of patients; for this reason, early detection is crucial and essential. Wireless sensing without contact has been suggested in this research as a way to distinguish between the symptoms of the diseases. The benefits include increased comfort, reduced self-consciousness, and other things. The system's convenience and cost advantage are its key benefits.

This work put forth a technique for creating machine-learning models that use individualized labelling and impaired gait patterns to predict AoG accurately and quickly. This strategy produced models that performed better than other prediction literature models. Using artificial intelligence methods and clinical illness characteristics, we can design more effective AoG prediction models that give individuals with PD a method to stop the upcoming AoG via supplying anticipatory cueing.

The data are initially preprocessed by taking away outliers and applying wavelet transform filtering, after which the required features are identified, and finally the model is trained using BP neural network, SVM and RF machine-learning methods. The experimental results demonstrate the effectiveness of the technical plan outlined in this study by demonstrating that most of the algorithms can reach approximately 90–98% prediction accuracy, accurately distinguishing between sensory ataxia and CA. But, Adaboost is giving 99.6% accuracy to predict the neurological disease. It is giving the best result compared to other machine-learning algorithms. Next, a further investigation of C-Band wireless sensing technology's application in healthcare will take place, and we'll suggest more clinical application programmes to make clinical detection more precise, dependable and intelligent in order to ease the burden on doctors and patients. In future, the approach can be enhanced with the assistance of transfer learning.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Xia T, Yang J, Chen L. Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning. Autom Constr. 2022;133:103992.
[2] Gunjan VK, Vijayalata Y, Valli S, et al. Machine learning and cloud-based knowledge graphs to recognize suicidal mental tendencies. Comput Intell Neurosci. 2022;2022:1–10.
[3] Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. Briefings Bioinf. 2018 Nov;19(6):1236–1246.
[4] Adnan M, Kalra S, Cresswell JC, et al. Federated learning and differential privacy for medical image analysis. Sci Rep. 2022 Dec;12(1):1–10.
[5] Hao M, Li H, Luo X, et al. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. IEEE Trans Ind Informat. 2020 Oct;16(10): 6532–6542.
[6] Liu Y, Yu JJQ, Kang J, et al. Privacy-preserving traffic flow prediction: A federated learning approach. IEEE Internet Things J. 2020 Aug;7(8):7751–7763.
[7] Wu Q, He K, Chen X. Personalized federated learning for intelligent IoT applications: a cloud-edge based framework, IEEE Open J Comput Soc. 2020;1:35–44.
[8] Zhao Y, Liu H, Li H, et al. Semi-supervised federated learning for activity recognition. arXiv:2011.00851. 2020.
[9] Shanmuga Sundari M, Sudha Rani M, Ram KB. Acute leukemia classification and prediction in blood cells using convolution neural network. International Conference on Innovative Computing and Communications. Singapore: Springer; 2023. p. 129–137.
[10] Tharini VJ, Shivakumar BL. An efficient pruned matrix aided utility tree for high utility itemset mining from transactional database. Int J Intell Sys Appl Eng. 2023;11(4s):46–55.
[11] Wu Q, Chen X, Zhou Z, et al. Fedhome: cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans Mobile Comput., Early Access. 2020 Dec 16. doi:10.1109/TMC.2020.3045266.
[12] Rieke N, et al. The future of digital health with federated learning. Npj Digit Med. 2020 Dec;3(1):1–7.
[13] Kumar R, et al. Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging. IEEE Sensors J. 2021 Jul;21(14):16301–16314.
[14] Warnat-Herresthal S, et al. Swarm learning for decentralized and confidential clinical machine learning. Nature. 2021;594(7862):265–270.
[15] Sundari MS, Nayak RK. Process mining in healthcare systems: a critical review and its future. Int J Emerging Trends Eng Res. 2020;8(9):5197–5208.
[16] Yang Z, Zhong S, Carass A, et al. Deep learning for cerebellar ataxia classification and functional score regression. In: Wu G, Zhang D, Zhou L, editors. Machine learning in medical imaging. Cham: Springer; 2014. p. 68–76.
[17] Chang Z, et al. Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and machine learning. Sci Rep. 2020 Dec;10(1):1–10.
[18] Kashyap B, Pathirana PN, Horne M, et al. Quantitative assessment of speech in cerebellar ataxia using magnitude and phase based cepstrum. Ann Biomed Eng. 2020 Apr;48(4):1322–1336.
[19] Tran H, Pathirana PN, Horne M, et al. Quantitative evaluation of cerebellar ataxia through automated assessment of upper limb movements. IEEE Trans Neural Syst Rehabil Eng. 2019 May;27(5):1081–1091.
[20] Shanmuga Sundari M, Jadala VC. Improved performance analysis for cerebellar ataxia disease classification using AdaBoost. NeuroQuantology. 2022;20(6): 9488–9497.
[21] Lee J, Kagamihara Y, Kakei S. A new method for functional evaluation of motor commands in patients with cerebellar ataxia. PLoS ONE. 2015 Jul;10(7): e0132983.
[22] Hohenfeld C, et al. Application of quantitative motor assessments in Friedreich ataxia and evaluation of their relation to clinical measures. Cerebellum. 2019 Oct;18(5):896–909.
[23] Chen Y, Ghannam R, Heidari H. Air quality monitoring using portable multi-sensory module for neurological disease prevention. 2019 UK/China Emerging Technologies (UCET), 2019 Aug 21, IEEE, p. 1–4.
[24] Rukhsar S. Discrimination of multi-class EEG signal in phase space of variability for epileptic seizure detection using error-correcting output code (ECOC). Int J Inf Technol . 2018;14(2):1–13.
[25] Nasser IM, Al-Shawwa M, Abu-Naser SS. Artificial neural network for diagnosing autism spectrum disorder. Int J Acad Inf Syst Res (IJAISR). 2019;3(2):930–933.
[26] Shanmuga Sundari M, Sudha Rani M, Kranthi A. Detect traffic lane image using geospatial LiDAR data point clouds with machine learning analysis. In: Intelligent system design. Singapore: Springer; 2023. p. 217–225.
[27] Yang X, Member S. Activity pattern mining for healthcare. IEEE Access. 2020;8:56730–56738. doi:10.1109/ACCESS.2020.2981670.
[28] J. V. Chandra, G. Ranjith, A. Shanthisri, etal, A framework for implementing machine learning algorithms using data sets, International Journal of Innovative Technology and Exploring Engineering (IJITEE). 2019 Oct;8(11):155–160. doi:10.35940/ijitee.K1263.0981119.
[29] Ker JI, Wang Y, Hajli MN, et al. Deploying lean in healthcare: evaluating information technology effectiveness in US hospital pharmacies. Int J Inf Manage. 2014;34(4):556–560.
[30] Lin HT, Li L. Support vector machinery for infinite ensemble learning. J Mach Learn Res. 2008;9:285–312.
[31] Shanmuga Sundari M, Samyuktha P, Kranthi A, Das, S. Evaluating performance on COVID-19 tweet sentiment analysis outbreak using support vector machine. In: S Das, editor. Smart intelligent computing and applications, Vol. 1, Singapore: Springer; 2022. p. 151–159.
[32] Lu H, Gao H, Ye M, et al. A hybrid ensemble algorithm combining AdaBoost and genetic algorithm for cancer classification with gene expression data. IEEE/ACM Trans Comput Biol Bioinforma. 2021;18(3):863–870. doi:10.1109/TCBB.2019.2952102.
[33] Kshatri SS, Singh D, Narain B, et al. An empirical analysis of machine learning algorithms for crime prediction

using stacked generalization: an ensemble approach. IEEE Access. 2021;9:67488–67500. doi:10.1109/ACCESS.2021.3075140.
[34] Shanmuga Sundari M, Nayak RK. Deviation and cluster analysis using inductive alpha miner in process mining. In: Rudra Kalyan Nayak, editor. Communication, software and networks. Singapore: Springer; 2023. p. 451–458.
[35] Eldardiry H, Neville J, Rossi RA. Ensemble learning for relational data. J Mach Learn Res. 2020;21:49–41.
[36] Li J, Ling T, Huaijun W, et al. Segmentation and recognition of basic and transitional activities for continuous physical human activity. IEEE Access. 2019;7:42565–42576. doi:10.1109/ACCESS.2019.2905575.
[37] Gopi AP, Jyothi R, Narayana VL, et al. Classification of tweets data based on polarity using improved RBF kernel of SVM. Int J Inf Technol . 2020;15(2):1–16.
[38] Cravioto DAG, Ramos RED, Galaz MAZ, et al. Analysing factors that influence alumni graduate studies attainment with decision trees. Proc 2020 Int Conf Comput Sci Softw Eng CSASE 2020. 2020: 44–49. doi:10.1109/CSASE48920.2020.9142069.
[39] Martins SF, Fonseca RA, Silva LO, et al. Numerical simulations of laser wakefield accelerators in optimal Lorentz frames. Comput Phys Commun. 2010;181(5):869–875.
[40] Laatifi M, et al. Machine learning approaches in COVID-19 severity risk prediction in Morocco. J Big Data. 2022;9(1). doi:10.1186/s40537-021-00557-0.
[41] Çelik U, Yurtay N, Yılmaz Z. Migraine diagnosis by using artificial neural networks and decision tree techniques, AJIT-e online acad. J Inf Technol. 2014;5(14):79–90. doi:10.5824/1309-1581.2014.1.005.x.
[42] Matloob F, et al. Software defect prediction using ensemble learning: a systematic literature review. IEEE Access. 2021;9(July):98754–98771. doi:10.1109/ACCESS.2021.3095559.
[43] Horne M, Power L, Szmulewicz DJ. Quantitative evaluation of cerebellar ataxia through automated assessment of upper limb movements. IEEE Trans Neural Syst Rehabil Eng. 2019;27(5):1081–1091.
[44] Akhila G, Madhubhavana N, Ramareddy NV, et al. A survey on health prediction using human activity patterns through smart devices. Int J Eng Technol. 2018. doi:10.14419/ijet.v7i1.1.9472.
[45] Al-janabi S, Salman AH. Sensitive integration of multilevel optimization model in human activity recognition for smartphone and smartwatch applications. Big Data Mining and Analytics. 2021;4(2):124–138. doi:10.26599/BDMA.2020.9020022.
[46] Sundari MS, Nayak RK. Efficient tracing and detection of activity deviation in event log using ProM in health care industry, 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC), 2021, p. 1238–1245.
[47] Vijaya Chandra J, Challa N, Pasupuletti SK. Machine learning framework to analyze against spear phishing. Int J Innov Technol Explor Eng. 2019;8(12):3605–3611. doi:10.35940/ijitee.L3802.1081219.
[48] Chandra JV, Challa N, Pasupuleti SK. Advanced persistent threat defense system using self-destructive mechanism for cloud security. Proc 2nd IEEE Int Conf Eng Technol ICETECH 2016. 2016; (March):7–11. doi:10.1109/ICETECH.2016.756918.
[49] Zhou M, Ai T, Zhou G, et al. A visualization method for mining colocation patterns constrained by a road network. IEEE Access. 2020;8:51933–51944. doi:10.1109/ACCESS.2020.2980168.
[50] Narasimham C, Sai K, Pasupuleti SK.. Detection of deceptive phishing based on machine learning techniquesIn Smart Technologies in Data Science and Communication: Proceedings of SMART-DSC. Springer Singapore2019 Dec. p. 13–22.
[51] Serrao M, Chini G, Bergantino M, et al. Dataset on gait patterns in degenerative neurological diseases. Data Brief. 2017;16:806–816. doi:10.1016/j.dib.2017.12.022.
[52] Chandra JV, Pasupuleti SK. Machine learning methodologies for predicting neurological disease using behavioral activity mining in health care. 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1, IEEE, 2022 Mar, p. 1035–1039.
[53] Sundari MS, Jadala VC, Pasupuleti SK. Prediction of activity pattern mining for neurological disease using convolution neural network. 2022 7th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2022 Jun, p. 1319–1324.
[54] Raju CSK, Pranitha K, Samyuktha P, et al. Prediction of COVID 19-chest image classification and detection using RELM classifier in machine learning. 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, IEEE, 2022 Mar, p. 1184–1188.