International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 16 (2024) pp. 87–100

Article

Extended Study of Gait Analysis in Identification of Cardiovascular Diseases

Rabindra Bista [1],*, Sanjog Sigdel [1] and Joao C. Ferreira [2,3]

[1] Department of Computer Science and Engineering, Kathmandu University, Dhulikhel-4, Kavre 6250, Nepal; sigdelsanjog@gmail.com
[2] Inov Inesc Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal; joao.carlos.ferreira@iscte-iul.pt
[3] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisboa, Portugal
* Correspondence: rbista@ku.edu.np

Received date: 2 April 2024; Accepted date: 6 May 2024; Published online: 24 May 2024

Abstract: Walking pattern analysis known as Gait Analysis, is one of the key indicators among various clinical parameters such as cardiovascular disease, in identifying symptoms of diseases. With sensors, the walking patterns of people can be retrieved to identify the differences between patients and healthy controls. The paper comprises an analysis of Gait datasets comprising nine Kinematic gait parameters, using classification models, the development of a data augmentation algorithm, and a proposed prediction model. Classification analysis of the Gait dataset was done using Neural Network (98.65% accuracy). An algorithm GDAA was developed for the augmentation of Gait data whose time complexity is *O(nf)*. The final result lies within a minimum, maximum range of the original dataset. The algorithm can be extended by researchers in augmenting their dataset with slight modifications. Data analysis of the augmented dataset was done on varying sizes of datasets to evaluate optimum classification results. Rigorous analysis of augmented data was done with Neural Network (97.1% accuracy). Ranking of gait features for both independent and dependent features was also done during the analysis. A prediction model is proposed which identifies whether input Gait data belongs to pathological gait or healthy gait with the help of a classification model trained on augmented data.

Keywords: gait analysis; cardiovascular diseases; explainable artificial intelligence; sensors; digital health

1. Introduction

Digital health is the convergence of digital technologies with health, healthcare, living, and society to enhance the efficiency of healthcare delivery and make medicine more personalized and precise [1]. Walking is one of the most common activities humans do every day. With the help of sensors, we can analyze the walking pattern of a person which provides useful health information [2]. Gait Analysis is a systematic study, which involves analysis of measurement, description, and assessment of qualities that characterize human locomotion [3]. Healthcare professionals can explore human gait and implement results from the gait phase recognition concept in their routine practice to identify abnormalities [4]. Gait Analysis cannot definitively diagnose or predict a particular disease. However, the analysis results can be a value add for medical professionals to understand patients' behavior. Cardiovascular disease is a leading cause of mortality worldwide, and early identification of individuals at risk is critical for timely intervention and management. The MPU-6050 sensor is a small, low-cost device that can be easily attached to the body to collect Gait data. By analyzing this data using machine learning techniques, a prediction model can be developed to identify individuals at risk of developing cardiovascular disease.

MPU-6050 sensors provide a non-invasive, cost-effective, and easily accessible method for identifying individuals at risk. This ultimately leads to better outcomes and reduced healthcare costs [5]. Gait Analysis can provide insights into an individual's physical function and mobility, and it can be useful in the diagnosis and management of various health conditions, including neurological disorders, musculoskeletal conditions, and cardiovascular disease. By analyzing Gait data collected from sensors, such as the MPU-6050, machine learning techniques can be applied to develop prediction models for identifying individuals at risk of developing certain health conditions or for monitoring the progression of an existing condition.

Gait Analysis can provide a valuable tool for improving the diagnosis and management of various health conditions, and the use of sensors can enhance the accuracy and reliability of Gait Analysis by providing objective and quantitative data.

Health data is a crucial subject to perform research in. With limited data and time constraints, an augmentation algorithm that generates abundant simulated data was developed. Experiments with augmented data with accuracy and analysis with the actual dataset were performed. Augmented data for each gait parameter lies within the maximum and minimum range of the original data. The classification accuracy of augmented data was similar to that of the real datasets, ranging from 88.4% to 99.2%. Details are in section 4. gait Features were not only identified as independent and dependent but were also ranked using Info. Gain, Gini, ANOVA, and χ^2 methods. Finally, a prediction model was proposed that predicts whether the gait pattern of the new subject is pathological or healthy gait with an average of 97%.

While several studies have investigated the potential of gait analysis in cardiovascular and other diseases, our research focused specifically on a data augmentation algorithm. The augmentation algorithm provided a solution to overcome the issue of limitation of healthcare data research. Building upon this foundation, the present study highlights a detailed research analytical framework, data preprocessing, classification, comparison and prediction model along with feature ranking among the features used for the research.

2. **Related Work**

Research work [6] explores the use of Gait Analysis and the MPU-6050 sensor in the early identification of cardiovascular diseases. The authors develop an analysis framework for Gait data and apply data augmentation techniques to improve the accuracy of their prediction model. The results of their study suggest that Gait Analysis using the MPU-6050 sensor can be a useful tool for the early identification of cardiovascular diseases. The paper [7] presents a Gait Analysis framework for predicting cardiovascular disease using data collected from the MPU-6050 sensor. The authors apply machine learning techniques to Gait data and compare the performance of various prediction models. Their results demonstrate that Gait Analysis using the MPU-6050 sensor can be an effective tool for the early identification of cardiovascular disease. The research work [8] presents a machine learning-based approach for the early identification of cardiovascular disease using Gait Analysis with the MPU-6050 sensor. The authors develop a prediction model using a combination of gait features and clinical data. Their results demonstrate that Gait Analysis using the MPU-6050 sensor can be a useful tool for the early identification of cardiovascular disease, particularly when combined with other clinical data.

This paper [9] presents a prediction model for the early identification of cardiovascular disease using Gait Analysis with the MPU-6050 sensor. The authors develop a feature extraction method and apply a Support Vector Machine (SVM) classifier to predict cardiovascular disease. Their results demonstrate that Gait Analysis using the MPU-6050 sensor can be an effective tool for the early identification of cardiovascular disease, with an accuracy of over 90%. This paper [10] presents a systematic review of studies investigating the use of Gait Analysis with the MPU-6050 sensor for the early identification of cardiovascular disease. The authors analyze the methodologies used in previous studies and compare the performance of various prediction models. Their review suggests that Gait Analysis using the MPU-6050 sensor has the potential to be a useful tool for the early identification of cardiovascular disease, but further research is needed to improve its accuracy and reliability.

The study of related work was useful in selecting the MPU-6050 sensor for the Gait Analysis. These related works were useful in validating the gait parameters and machine learning algorithms. Nine gait parameters were evaluated and the machine learning algorithms used are relevant and value-adding in Gait Analysis research. Evaluation of time and distance parameters during walking helps assess abnormal gait, to quantify improvement resulting from interventions, or to predict subsequent events such as falls [11]. The study of human gait, analysis of gait patterns with the help of different categories of gait parameters, and statistical and machine learning algorithms used in the study of Gait Analysis are studied in detail to analyze the research gap. Some relevant works are presented in the below sections.

2.1. Human Gait Analysis

Gait phases describe the entire walking period of a human being. Walking is a periodic movement of body segments and includes repetitive motions, therefore, the gait phase is used for illustration [12]. However, patients impaired by paralysis or arthritis might show deviated behavior. There is a low probability of a patient with paralysis touching his/her heel to the ground like that of healthy control. Likewise, the initial floor contact may be made by the entire foot i.e., flat foot, rather than using the forefoot to contact [13]. Sensor-based systems have assisted research works to identify healthy and pathological gait with the help of sensor-based computer systems. Research [14] performs temporal Gait Analysis to evaluate adaptive gyroscope-based algorithms. The algorithm used in the research adaptively calculated thresholds to determine Initial Contact and Terminal Contact which have some parameters for analyzing gait patterns. Authors in the paper [15] have published methods to detect the quasi-real-time gait event using uniaxial shank-attached gyroscopes. Uni-axial gyroscopes were attached on shanks measuring the mediolateral axis angular velocities which helped in identifying end contact and initial contact. It suggests that end contact and initial contact events have distinctive features in the sensor signals appearing as sharp negative peaks.

Researchers have built an inverted pendulum model of gait which posits that the leg alternates between advancing as a pendulum during the swing phase (pivot at the hip) and as an inverted pendulum during the stance phase (pivot at the foot) [16]. The use of two sensor devices Microsoft Kinect and 3DMA cameras is done to perform a comparative analysis of gait parameters. This research highlighted the gait speed, step length, and stride length possessed overall and relative agreement between two sensor devices with low percentage error (<8%). Foot swing on the other hand possessed excellent relative (r = 0.93) and overall agreement with percentage error (13%). Research on gait disorder and balance is performed by the authors [17] on patients with Alzheimer's disease. Stride and gait cycle decomposition were retrieved from the sensors for analysis. Research performed an ANOVA test which reflected a significant measurable difference in gait parameters. Gait Analysis has played an important role in analyzing the step-by-step spatiotemporal parameters of the elderly suffering from hemiparetic, parkinsonian, and choreic gait.

Research work presents a case study in the analysis of Kinetic and Kinematic gait parameters in the spastic hemiplegic patient after selective tribal neurotomy [18,19]. Research experiments on the healthy and patients with neurological diseases. Gait Analysis was performed before and after a week of patients suffering from spastic hemiplegia using 3DGA systems. The research identified the spatiotemporal factors associated with the gait of bilateral lower extremities one week after STN. Authors identified that one of the symptoms of heart-failure in old age people was gait speed [20]. It was identified as one of the parameters of the research which consisted of 331 subjects 70 years old or above. Research work introduced a new approach to achieving a more reliable method to assess the physical performance status (PPS) among different groups of cancer patients undergoing chemotherapy [21]. With the use of body-mounted inertial sensors in 23 body segments, a six-minute walk test (6MWT) was performed. Research work considered calculating the parameters such as running average, temporal median, and motion history image two weeks before chemotherapy, and two weeks after chemotherapy. It was observed that a higher gait speed with a greater range of body motion is related to an appropriate PPS. Slower gait speed resulted in a smaller range of physical motions which caused a limitation in a variety of clusters along with two or more consecutive sequences. A wearable multi-axis inertial measurement module using an MPU-6050 sensor was attached to the feet of patients with Stroke or Parkinson’s disease [22]. This research uses a tri-axial gyroscope, tri-axial accelerometer, inertial signal acquisition, and signal processing to detect the gait phase. This work will greatly assist in comparing the gait parameters with cardiovascular patients.

Nandy and Chakraborty performed Gait Analysis research using a Kinect Xbox device and cross-validated it with their in-house development of a sensor-based biometric suit, an Intelligent Gait Oscillation Detector(IGOD) [23]. Researchers captured gait signatures from joint angle trajectories of the left hip, left knee, reign hip, and right knee of the subjects’ skeleton model. Machine learning algorithms such as Naive Bayesian classifiers and K-Nearest Neighbour are used to compare IGOD Data and Kinect Sensor Data. The research concluded the Mean % Error from Kinect Sensor Data was reduced from that of IGOD Data and the result increased for the k-NN algorithm.

Allseits and the team developed a real-time algorithm for temporal Gait Analysis using inertial measurement units [24]. Shank gyroscope has been used as an Inertial Measurement Unit to identify heel-strike and toe-off. Research highlights the significance of heel-strike and toe-off which serves as a foundation for the calculation of gait parameters. In-shoe motion sensors (IMs) are used on 26 healthy subjects to estimate the temporal gait parameters concerning bilateral lower limbs using a gait event detection approach [25]. The research was carried out by using a 3D motion analysis system, the Track 3 (Vicon Motion Systems, Oxford, UK), for measuring the reference measures and a 10 Bonita B10 motion-capture cameras (Vicon Motion Systems, Oxford, UK), with five cameras on each side of a straight path was used for the motion analysis. Results from this research evaluated signal features for gait event detection, the best candidate signal feature among nine gait features. As an outcome of this work, researchers were able to establish a method for estimating temporal gait which can be used further in cross-examination or analysis of patients suffering from different diseases.

The geometrical method to estimate step length and its symmetry between two sides with the measurement from four IMUs was attached to lower limbs [26]. The author used a 3D accelerometer with a range of ±8g and a 3D gyroscope with a range of ±1000°/s to estimate the step length and gait symmetry between two legs. The root means a square error of step length estimation was computed during the data analysis. Research also computed relative errors to compare their results with previous works. Healthy subjects had 5.4 cm or RMSE in step length and 3% RMSE in gait symmetry. While subjects with gait impairment had 9.6 cm and 14.7% RMSE in step length estimation and gait symmetry estimation respectively.

2.2. Statistical and Machine Learning Algorithms

Author Zheng and team conducted a feasibility study using machine learning and statistical approaches in identifying severe disturbances of gait and gait initiation in neurodegenerative diseases [27]. Gait data published by PhysioNet was used in the research which consists of gait records from 15 patients with PD (Parkinson's disease), 20 patients with HD (Huntington's disease), 13 patients with ALS (Amyotrophic Lateral Sclerosis), and 16 healthy controls. Three supervised classification models Support Vector Machine, KStar, and Random Forest were used in the analysis with twelve measurement features like gait cycles, left stride interval, right stride interval, left swing interval, left stance interval, right stance interval, etc. These classification models across all seven binary classification problems exhibit an accuracy ranging from 73.23% to 93.96% and an AUC ranging from 0.80 to 0.93. SVM, KStar, and RF models predicted accuracy of 86.9%, 84.7%, and 84.9% with AUC of 0.91, 0.91, and 0.97 respectively.

Researchers used the 3d-accelerometer and a 3d-gyroscope to identify the spatial gait parameters (foot angle, stride length, and stride width) and temporal parameters (heel stride, toe-on contact time, stance time, stride time, and swing time) [28]. The research used a publicly available benchmark dataset and estimated stride length, width, and mediolateral change in foot angle up to −0.15 ± 6.09 cm, −0.09 ± 4.22 cm, and 0.13 ± 3.78 degrees respectively. Research particularly focuses on the extraction of gait parameters to experiment with the results with Deep Convolutional Neural Networks. Error evaluation on the training set for 90/10% train/test split of the dataset was done. Two models were prepared and they estimated the error function to be 0.02 and 0.1 respectively.

The accuracy of optimized solutions was evaluated for measuring spatiotemporal gait parameters [29,30]. Subjects were equipped with two wireless Inertial Measurements Units composed of a tri-axial accelerometer, triaxial gyroscope, and a triaxial magnetometer on their feet. The result considered temporal parameters HeelStrike (HS), ToeStrike (TS), HeelOff (HO), and ToeOff (TO) & spatial parameters StrideLength (LS). ANOVA test evaluated the mean and standard deviation of the relative error for stride length, gait time, and gait speed for the left and right foot. The overall mean and standard deviation error expressed in percentage for stride length, gait time, and gait speed were 1.1 ± 0.7 & 0.8 ± 0.5, 1.7 ± 0.9 & 1.0 ± 1.0, 3.4 ± 1.8 & 2.2 ± 1.1 respectively. Proposed statistical model for the evaluation of accuracy, research gained an overall accuracy of 1.1 ± 07% and 0.8 ± 0.5% for left and right sensors respectively during the evaluation of stride length. Recent work on gait monitoring and measurements has been carried out for Parkinson's disease using Artificial Intelligence Based wearable gait monitoring [31]. Research focuses on monitoring and assessment of gait in patients with Parkinson's disease and also proposes a wearable physiograph for qualitative and quantitative gait assessment by performing bilateral tracking of the foot biomechanics and unilateral tracking of the arm balance. AI-based decision support is built using gait parameters in Convolutional Neural Network (CNN) with 95% accuracy, 90% sensitivity, and 95% precision.

The gait pattern analysis research work has been carried out to understand its association with diseases such as Parkinson's disease, Alzheimer, Peripheral Arterial Disease, and Neuro-degenerative diseases. Research work on Gait Analysis to analyze the foot movement of elderly citizens has been carried out. However, inadequate research was carried out on cardiovascular disease and gait pattern analysis.

Review of related work recommended a handful of works done on cardiovascular diseases and gait parameters together. The gait pattern analysis research work has been carried out to understand its association with diseases such as Parkinson's Disease, Alzheimer, Peripheral Arterial Disease, and Neuro-degenerative diseases. Research work on Gait Analysis to analyze the foot movement of elderly citizens has also been carried out. However, very little research was carried out on cardiovascular disease and gait pattern analysis. A research gap was identified in the analysis of cardiovascular disease with gait patterns. Limited work has considered the kinematic gait parameters for Gait Analysis. Very few datasets were available for exploring the Gait data. Among the literature which carried research on Gait Analysis and diseases, very few research considered the analysis with neural networks, and the ranking of features that were important in their research work wasn't mentioned. This literature review shares different research gaps in data analysis, machine learning approaches, and technology algorithm development for doing research work in cardiovascular diseases.

3. Methodology

A framework for Gait Analysis was designed to perform a scientific study on Gait Analysis and how one can identify whether the new subjects were healthy controls or cardiovascular patients. The research framework of Gait Analysis in Figure 1, involves several steps, including data collection, data preprocessing, feature extraction, data analysis, and interpretation:

Data Collection: The first step involves collecting Gait data from individuals using sensors such as the MPU-6050, which can provide accurate measurements of gait parameters [29]. Data Preprocessing: The collected data may contain noise or artifacts that need to be removed before analysis. Research collected sensor data which was preprocessed using signal filtering technique. Feature Extraction: Gait Analysis involves extracting meaningful features from the preprocessed data, such as stride length, step width, and walking speed. These features can be extracted using signal processing techniques and statistical methods. Data Analysis: The extracted features are then analyzed using machine learning algorithms to identify patterns and relationships between the gait parameters and specific health conditions.

Interpretation: Finally, the results of the analysis are interpreted to provide insights into the individual's physical function and mobility. The interpretation of the results can be used for the diagnosis and management of various health conditions in cardiovascular disease.

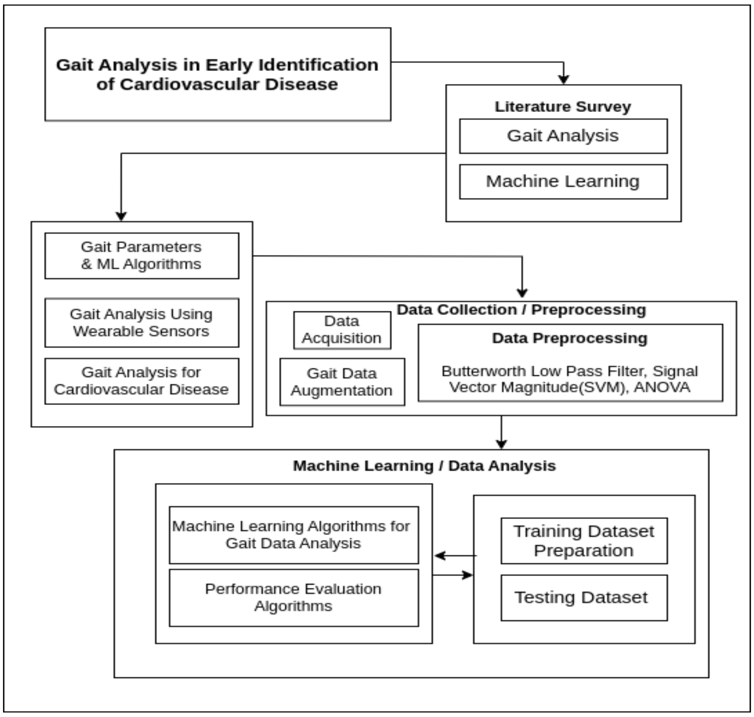

Figure 1. Research Framework of Gait Analysis.

3.1. Data Collection

An Inertial Measurement Unit (IMU) consisting of Raspberry PI 3 Model B, MPU-6050 Tri-axial Gyroscope, & Tri-axial Accelerometer was developed. The hardware setup was inspired by the experimentation model set up by Fitriani and the team [32]. IMU developed was fixed on the shoes of research participants which included 16 healthy controls and 53 patients with cardiovascular diseases. Subjects were asked to walk 10 meters in order to collect their Gait data. The characteristics of subjects who participated in the Gait Analysis are shown in Table 1.

Table 1. Characteristics of participants of Gait Analysis

Parameters	Cardiovascular Patients (N=53)	Healthy Controls (N = 16)
Men/Women(n)	28/25	5/11
Height(cm)	161.79 ± 11.53	162.88 ± 10.85
Weight(kg)	63.57 ± 11.52	62.19 ± 10.23
BMI(kg/m^2)	24.30 ± 4.24	23.49 ± 3.64
Age(years)	44.58 ± 17.01	32.44 ± 6.20

Total of nine features: Number of strides, Stride length, Stride time, Walking time, Stride frequency, Stride velocity, Stride cadence, Stance time, and Swing time were considered in this research. The Data acquisition method and initial analysis result is published at IEEE Conference [29]. Raw data collected by the IMU went through data/signal processing using a Butterworth low-pass filter [29].

In receiving data from various wireless-sensor networks, one important part is to follow protocols and methodologies for data integrity. The actual data received from the sensor should be preserved and aggregated as suggested by the research paper [33]. To ensure data can be understood well as a part of signal processing Butterworth low pass filter algorithm is used which helps in reducing the high-frequency data and the walking trembles. Equation (1) illustrates the formula given below:

$$\acute{G}(\omega) = \frac{1}{\sqrt{1+\omega^{2n}}} \tag{1}$$

where ω is the angular frequency in radians per second and n is the number of poles in the filter equal to the number of reactive elements in a passive filter. Figure 2 displays the graph representing data received from the IMU after signal calibration. After the noise cancellation, the tri-axial data were calibrated by applying Signal Vector Magnitude on the tri-axial data received from the sensor. Equation (2) is shown below:

$$\mathrm{SMV}\omega(\mathrm{k}) = \sqrt{[\omega\mathrm{lx(k)}2 + \omega\mathrm{ly(k)}2 + \omega\mathrm{lz(k)}2]}, \tag{2}$$

where, ωlx(k), ωly(k), and ωlz(k) are the angular velocities of X, Y, and Z-axis in K time.

Data was stored in a database and treated as a DataFrame in Pandas, a library available in Python programming language [34]. After data calibration, a dataset was prepared which has 69 rows and nine features, Number of strides, Stride length, Stride time, Walking time, Stride frequency, Stride velocity, Stride cadence, Stance time, and Swing time. The Gait dataset consists of a proportion of data between healthy and cardiovascular patients that is 25:75 in percentage.

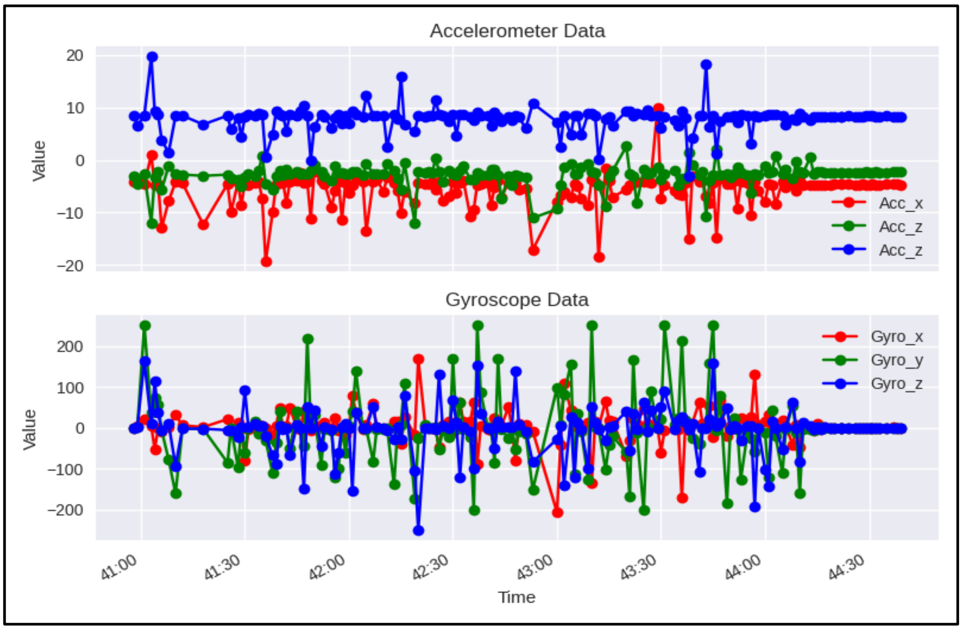

Figure 2. Graph of tri-axial MPU-6050 sensor data after calibration.

3.2. Gait Data Augmentation Algorithm

Analysis and results of the dataset collected from IMU are shared in Results Section 4. Next, a Gait Data Augmentation Algorithm(GDAA) (Algorithm 1) was developed with the assumption of minimum and maximum values for each parameter to be that from the previous dataset. A literature study presented that researchers used dataset available in different data portals and implemented them in their research as a secondary source of data [11,12,23]. Augment requires the same proportion of Healthy Controls and Patients. The Gait parameters should be within the range of values in the original dataset. The data augmentation algorithm is presented below:

Initially, the dependent and independent gait features are highlighted in their respective lists, minimum and maximum of independent parameters are calculated. For a range defined by the integer k, for all the independent parameters data is generated using a random function generator. The minimum and maximum values of each parameter define the range of random function generators. Those minimum and maximum values are calculated from the input file's data range. Generation of independent Gait data using a random function generator takes place in steps 8–11. For all four independent parameters, random data is generated and appended to the list respectively. The implementation of this random data generation involves four functions that call random functions separately using their min and max data range. As four functions do the same task repeatedly steps 8–11 in the algorithm are condensed.

Generation of dependent Gait data takes place in steps 12–15. For all the dependent gait parameters, data is calculated using their respective formula. The dependent gait parameters formula includes some independent gait features generated at that instant of for loop steps 8–11. Calculation of dependent Gait data functions in steps 12–15 includes five functions for each dependent gait parameter. Since the calculation of five different parameters is a similar task, the steps are condensed in the algorithm. Each Gait data is appended in their respective lists. At the end of a single iteration, all parameters will have a value, and those values are stored in the respective lists which are initially empty. As the loop ends, all lists are grouped i.e., zipped into a single list. That list is then converted into a new data frame which is then exported as a CSV file in the desired directory.

To evaluate the significance of augmented data from the above algorithm One Way Analysis of Variance test was chosen [9,14]. One Way ANOVA test ensures that the data of Healthy Controls and Cardiovascular patients differ significantly from each other compared with individual gait parameters. With the statistical analysis, the gait parameters of patients with Cardiovascular Diseases deferred from that of HCs. The p-value which examined the level of significance was less than 0.05 (5% level of significance).

Algorithm 1. Gait Data Augmentation Algorithm (GDAA)

Input: Gait Dataset file as a Dataframe
List: Empty list of all features from Gait Dataset
Integer: k
Output: Augmented Gait Dataset

1: **Start**
2: Convert Input file as a DataFrame.
3: Independent[] ←Independent gait Features
4: Dependent[] ←Dependent gait Features
5: Calculate Minimum and Maximum of the Independent[] gait parameters
6: **For** values in Range(0, k):
7: Append the subject title to the empty list
8: **Function Calls:** *GenerateStrideTime(); GenerateStrideCount();GenerateSwingTime(); GenerateStanceTime();*
9: Generate Stride Time, Generate Stride Count, Generate Swing Time & Generate Stance Time
10: Append Stride Time, Stride Count, Swing Time & Stance Time to the list
11: **End Function**
12: **Function Calls:***CalculateStrideLength(); CalculateStrideFrequency(); CalculateWalkingTime(); CalculateStrideFrequency(); CalculateStrideCadence();*
13: Calculate Stride Length, Calculate Stride Frequency, Calculate Walking Time, Calculate Stride Frequency, Calculate Stride Cadence
14: Append Stride Length, Stride Frequency, Walking Time, Stride Frequency, Stride Cadence to the list.
15: **End Function**
16: **End For**
17: Zip the lists which have generated data
18: Create a dataframe from the zipped list
19: Store final value as CSV to the directory
20:**End**

4. Experiments and Results

This section describes the results experimented with the machine learning models for the dataset generated by the IMU and augmented data. Analysis was done using k-Nearest Neighbors, Classification and Regression Tree, Support Vector Machine, Naive Bayes, and Logistic Regression models. Data analysis was conducted using a Neural Network with 100 neurons in hidden layers and ReLu (Rectified Linear Unit) as the activation function for 80 no. of iterations maximum. The experimentation was done first with the original dataset where the accuracy calculation listed in Table 2 below was performed with ten cross-fold validation.

Table 2. Accuracy calculation of Classification Models (A).

Model	CA	F1	Precision	Recall
k-Nearest Neighbors	0.93	0.931	0.936	0.93
Classification & R Tree	0.95	0.95	0.951	0.95
Support Vector Machine	0.973	0.973	0.975	0.973
Naive Bayes	0.882	0.888	0.915	0.882
Logistic Regression	0.955	0.954	0.955	0.955
Neural Network	0.986	0.986	0.986	0.986

Table 3 is the results of accuracy calculation of the augmented dataset of the same size that is 69 rows. A dataset of sizes 69, 138, 207, 345, and 414 was generated with the augmentation algorithm where the proportion is 75:25.

Table 3. Accuracy calculation of Classification Models (B).

Model	CA	F1	Precision	Recall
K-Nearest Neighbors	0.928	0.931	0.945	0.928
Classification & R Tree	0.942	0.943	0.946	0.942
Support Vector Machine	0.928	0.93	0.936	0.928
Neural Network	0.971	0.972	0.974	0.971
Naive Bayes	0.899	0.904	0.929	0.899
Logistic Regression	0.884	0.886	0.89	0.884

Comparing the result of the Augmented dataset and the original dataset of size 69 it was identified that besides Naive Bayes there wasn't any gain in the classification results. However, the results are above 90% on average, which can be considered for accepting augmented data. Also, these results have improved with the increase of the augmented dataset in the section below. The classification results of the augmented dataset were near to that of the real dataset. Classification Analysis of Augmented Data for all dataset sizes was performed. The result favored optimum results for an augmented dataset of size 276(212P-64H). The result is shown in Table 4.

Table 4. Classification Analysis of Augmented Data. Data Size: 276(212P-64H).

Model	CA	F1	Precision	Recall
K-Nearest Neighbors	0.953	0.954	0.957	0.953
Classification & R Tree	0.978	0.978	0.979	0.978
Support Vector Machine	0.975	0.975	0.977	0.975
Neural Network	0.982	0.982	0.983	0.982
Naive Bayes	0.96	0.961	0.966	0.96
Logistic Regression	0.942	0.942	0.943	0.942

The gain in Classification accuracy, precision, recall, and F1 score were reflected in models like k-Nearest Neighbors, Classification and Regression Tree, and Naive Bayes. The last gain was observed with the model Support Vector Machine. The optimum results for the augmentation data were observed for the data size 276(212P-64H). The optimum result for neural network analysis was observed after experimenting with the analysis in several iterations shown in Table 5.

Table 5. Neural network analysis with varying iterations.

Model	CA	F1	Precision	Recall
80	0.975	0.975	0.977	0.975
90	0.975	0.975	0.977	0.975
100	0.975	0.975	0.977	0.975
110	0.978	0.979	0.98	0.978
120	0.978	0.979	0.98	0.978
130	0.978	0.979	0.98	0.978
140	0.982	0.982	0.983	0.982
150	0.982	0.982	0.983	0.982
160	0.982	0.982	0.983	0.982

Increasing the number of iterations to 140 iterations the optimum classification results were 98.2% accuracy, 98.2% F1 Score, 98.3% precision, and 98.2% recall. Research [35] suggests explainability as a powerful tool for justifying AI-based decisions. Extended work is carried out by classifying three different feature sets: (a) Dependent Gait Features, (b) Independent Gait Features, and (c) Dependent and Independent Features. This explains the importance of Gait features in this research. Different measures such as ANOVA score, Information Gain, Gini, and chi-square test (χ^2) have been calculated for the feature ranking purpose [15,36–38]. Recent research on feature ranking analysis such as Distribution Shapley values [39] could be another measure for the feature ranking. Since prominent literature has used Info. gain, Gini, ANOVA, and χ^2 Distribution Shapely values method will be a work for future research publication. Table 6 highlights the scores of features based on different measures in descending order.

Table 6. Feature Ranking of Dependent Gait Parameters.

Dependent Features	ANOVA	χ^2	Info. gain	Gini
Stride Velocity	518.0020	112.3012	0.5382	0.2606
Avg. Stride length	299.1685	83.2432	0.3771	0.1723
Walking Time	164.7313	112.3012	0.5382	0.2606
Stride Frequency	90.3104	57.2965	0.2834	0.1077
Stride Cadence	90.3104	57.2965	0.2834	0.1077

Stride Count, Avg. Stance time, Stride time, and Avg. Swing Time is an independent gait feature present in the Gait dataset. Table 7 highlights the scores of features based on different measures in descending order.

Table 7. Feature Ranking of Independent Gait Parameters.

Independent Features	ANOVA	χ^2	Info. gain	Gini
Stride Count	0.4010	0.1792	174.2995	89.0274
Avg Stance Time	0.3140	0.1285	110.0343	74.2617
Stride Time	0.2834	0.1077	83.2225	57.2965
Avg Swing Time	0.1320	0.0478	19.6426	15.6768

Among Independent gait features Stride Count has higher Information Gain, Gini score, ANOVA, and χ^2. Followed by the Stride Count, Average Stance Time has second ranking then Stride Time is third and fourth Average Swing Time. Among dependent gait features Walking Time has higher Info. Gain, Gini, and χ^2. Followed by Walking Time, Stride Velocity has the second-ranking, Average Stride length has the third rank, Stride Frequency has the fourth, and finally, Stride Cadence has the fifth ranking.

Third feature ranking i.e., ranking all features as a whole is shown in Table 8.

Table 8. Feature Ranking of entire gait features.

Independent Features	ANOVA	χ^2	Info. gain	Gini
Walking Time	0.5382	0.2606	164.7313	112.3012
Stride Velocity	0.5382	0.2606	518.0020	112.3012
Stride Count	0.4010	0.1792	174.2995	89.0274
Avg. Stride length	0.3771	0.1723	299.1685	83.2432
Avg. Stance Time	0.3140	0.1285	110.0343	74.2617
Stride Time	0.2834	0.1077	83.2225	57.2965
Stride Cadence	0.2834	0.1077	90.3104	57.2965
Stride Frequency	0.2834	0.1077	90.3104	57.2965
Avg. Swing Time	0.1320	0.0478	19.6426	15.6768

Results after ranking the features as a whole are slightly different from ranking independent and dependent features separately. Average Stance Time held the second rank while evaluated among independent features. Average Stance Time decreased its rank to the fifth position. Stride count which was the first in independent feature evaluation is now in the third rank. Research also includes feature importance using Explained Variance Ratio of the principal components [40]. Figure 3 illustrates the variance ratio of Gait features.

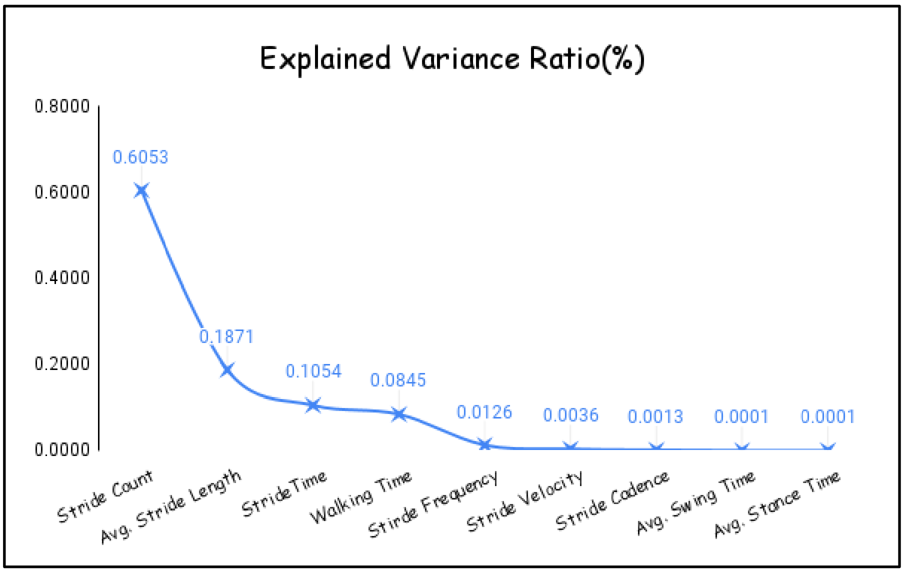

Figure 3. Explained Variance Ratio of Principal Components.

Stride Count is the most important gait feature followed by Avg. Stride Length, Stride Time and Walking Time. Rest of the features are least significant. Research incorporates all nine gait features for classification analysis, feature ranking and significance of data. Explainable AI(XAI) suggests an Example-based explanation as one of the techniques for making results interpretable [34]. Research presents such an example to describe the minimum conditions in Tables 9 and 10, that would have led to an alternate decision(pathological or healthy gait). Such an approach is presented as an unconditional counterfactual explanation because we won't need to describe the full logic of the algorithm [41].

Table 9. Testing Dataset for Prediction.

No. of Stride	Avg. Stride Length	Stride Time	Walking Time	Stride Frequency	Stride Velocity	Stride Cadence	Avg Swing Time	Avg Stance Time
11	0.909	2.175	23.928	0.459	0.417	27.582	0.854	0.865
16	0.625	1.547	24.76	0.646	0.403	38.771	0.894	0.636
6	1.66	0.985	5.910	1.015	1.691	60.908	0.669	0.327
8	1.25	1.111	8.889	0.899	1.124	53.998	0.701	0.305

Table 10. Prediction Results of Test Data.

Test No.	Neural Network	Naive Bayes	kNN	Logistic Regression	Tree	SVM
1	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00→0
2	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0	1.00 : 0.00 →0
3	0.00 : 1.00 →1	0.00 : 1.00 →1	0.00 : 1.00 →1	0.02 : 0.98 →1	0.02 : 0.98 →1	0.01 : 0.99 →1
4	0.00 : 1.00 →1	0.00 : 1.00 →1	0.00 : 1.00 →1	0.52 : 0.48 →0	0.02 : 0.98 →1	0.31 : 0.69 →1

Result 0 in the prediction Table 10 refers to pathological gait and 1 means healthy gait. When a new subject is provided in the prediction model, prediction undergoes the trained model and predicts whether the subject has a normal gait or a pathological gait. The proposed model predicts the result with an accuracy of 97%.

5. Conclusions and Future Work

This research performed rigorous work on analyzing the difference between the gait patterns of healthy versus patients with cardiovascular diseases. An Inertial Measurement Unit was developed for collecting data. The collected data went through calibration and signal processing with the Butterworth Low Pass algorithm and Signal Vector Magnitude algorithm to ensure noise was reduced and accurate data was received. The significance of data received from IMU was ensured via the One-Way ANOVA test. Analysis of generated data was done via multiple machine learning algorithms, with accuracy ranging from 88.2% to 98.6%.

Development of Gait data augmentation residing within the limitation of the real dataset, the performance of classification accuracy of augmented data, identifying independent and dependent gait parameters and their ranking using Info. Gain, Gini, ANOVA, and χ^2 methods are the key highlights of this research work. In addition to these, a prediction model was proposed that predicts whether the gait pattern of the new subject is pathological or healthy gait with an average of 97%. This research included all nine features in the classification analysis regardless of their feature importance identified by principal component analysis. Research focused on identifying significance of dependent and independent gait features in this research. Research can be extended by conducting dimensionality reduction.

This research can be further extended by experimenting with the classification accuracy of the Gait dataset increasing the gait parameters. We can also conduct the same experiment with the same gait features but classify disease categories that fall under cardiovascular diseases. This will give us a perspective on which cardiovascular disease has a greater effect on gait movement. This can also create a standard dataset for machine learning models that want to experiment and analyze Gait data.

Author Contributions

Conceptualization, R.B., S.S. and J.C.F.; Methodology, R.B., S.S., and J.C.F.; Software, S.S. R.B. and J.C.F.; Validation, R.B. and J.C.F.; Formal analysis, R.B. and J.C.F.; Investigation, R.B. and J.C.F.; Resources, R.B.; Data curation, R.B., S.S. and J.C.F.; Writing—original draft preparation, R.B.; Writing—review and editing, R.B., S.S. and J.C.F.; Visualization, S.S., R.B. and J.C.F.; Supervision, R.B.; Project administration, S.S., R.B. and J.C.F. All authors have read and agreed to the published version of the manuscript.

Funding

We would like to express our sincere gratitude to Nepalese Education in E-health Masters (NEEM) project under GA101083048 at Health Informatics Lab in Kathmandu University for supporting this research. This work was supported by the project "BLOCKCHAIN.PT(RE-C05-i01.01-Agendas/ Alianças Mobilizadoras para a Reindustrialização, Plano de Recuperação e Resiliência de Portugal na sua componente 5—Capitalização e Inovação Empresarial e com o Regulamento do Sistema de Incentivos "Agendas para a Inovação Empresarial", aprovado pela Portaria N.° 43-A/2022 de 19.01.2022).

Conflict of Interest Statement

The authors declare no conflicts of interest.

Data Availability Statement

Dataset will be made available for research purposes. Researchers can reach out to the Correspondence via email for the dataset request.

References

1. Bhavnani, S. P., Narula, J., & Sengupta, P. P. "Mobile technology and the digitization of healthcare."European Heart Journal. 2016, Volume 37(18), pp. 1428–1438. doi:10.1093/eurheartj/ehv770
2. Nilpanapan, Thirawut, Kerdcharoen, and Teerakiat. Social data shoes for gait monitoring of elderly people in smart homes. The 2016 Biomedical Engineering International Conference (BMEiCON-2016), 2016.
3. Ghoussayni, S., Stevens, C., Durham, S., Ewins, D. Assessment, and validation of a simple automated method for the detection of gait events and intervals. Gait Posture. 2004, Volume 20, pp. 266–272.
4. Lai, D. T. H., Begg, R. K., & Palaniswami, M. Computational Intelligence in gait Research: A Perspective on Current Applications and Future Challenges. IEEE Transactions on Information Technology in Biomedicine. 2009, Volume 13(5), pp. 687–702. doi:10.1109/titb.2009.2022913
5. Tanwar, R.; Nandal, N.; Zamani, M.; Manaf, A.A. Pathway of Trends and Technologies in Fall Detection: A Systematic Review. Healthcare 2022, 10, 172. https:// doi.org/10.3390/healthcare10010172
6. Rana, P., & Aggarwal, A. (2019). Gait Analysis for early identification of cardiovascular diseases using the MPU-6050 sensor. Journal of Medical Systems, 43(8), 226.
7. Wang, X., Wu, H., & Xie, Y. (2021). A Gait Analysis framework for cardiovascular disease prediction using MPU-6050 sensor data. IEEE Access, 9, 55759-55771.
8. Yao, Y., Song, M., & Lu, Z. (2021). Early identification of cardiovascular disease using Gait Analysis with the MPU-6050 sensor. IEEE Journal of Biomedical and Health Informatics, 25(9), 3352-3362.
9. Zhang, Y., & Yang, H. (2020). A prediction model for early identification of cardiovascular disease using Gait Analysis with the MPU-6050 sensor. Journal of Medical Systems, 44(7), 151.
10. Zhang, H., Li, Y., & Liu, B. (2019). Early identification of cardiovascular disease using Gait Analysis with the MPU-6050 sensor: A systematic review. IEEE Access, 7, 62171-62182.
11. Aminian, K., Najafi, B., Büla, C., Leyvraz, P., and Robert, P. Spatiotemporal parameters of gait measured by an ambulatory system using miniature gyroscopes. Journal of Biomechanics, Volume 35(5), 2002, pp. 689–699. doi:10.1016/s0021-9290(02)00008-8
12. Nandy, A. and Chakraborty, P., "A new paradigm of human Gait Analysis with Kinect," 2015 Eighth International Conference on Contemporary Computing (IC3),2015, pp. 443-448, doi: 10.1109/IC3.2015.7346722.
13. Huang, C.; Fukushi, K.; Wang, Z.; Nihey, F.; Kajitani, H.; Nakahara, K. Method for Estimating Temporal Gait Parameters Concerning Bilateral Lower Limbs of Healthy Subjects Using a Single In-Shoe Motion Sensor through a Gait Event Detection Approach. Sensors 2022, 22, 351. https://doi.org/10.3390/s22010351
14. Mileti, I., Taborri, J., Alvia, L., Parisi, S., Ditto, C. M., Peroni, L. C., Rossi, S., Fusaro, E., Prete, D. Z., Palermo, E. Accuracy Evaluation and Clinical Application of an Optimized Solution for Measuring Spatio-Temporal gait Parameters. 2020 IEEE International Symposium on Medical Measurements and Applications. 2020.
15. Huang, C.; Fukushi, K.; Wang, Z.; Nihey, F.; Kajitani, H.; Nakahara, K. Method for Estimating Temporal Gait Parameters Concerning Bilateral Lower Limbs of Healthy Subjects Using a Single In-Shoe Motion Sensor through a Gait Event Detection Approach. Sensors 2022, 22, 351. https://doi.org/10.3390/s22010351
16. Wang, L., Sun, Y., Li, Q., & Liu, T. Estimation of Step Length and Gait Asymmetry Using Wearable Inertial Sensors. IEEE Sensors Journal, 2018, Volume 18(9), pp. 3844–3851.doi:10.1109/jsen.2018.2815700
17. Hsu, Y. L., Chung P., Wang H., Pai M., Wang C. Wang j. gait and Balance Analysis for Patients With Alzheimer's Disease Using an Inertial-Sensor-Based Wearable Instrument. IEEE Journal of Biomedical and Health Informatics. 2014, Volume 18, pp. 1822-1830.g
18. Ippei, K., Hidetaka, A., Kikuta, K. Kinetic and kinematic Gait Analysis in a spastic hemiplegic patient after selective tibial neurotomy: A case report. Neurology Asia. 2015, Volume 20(4): pp. 395-399

19. Stebbins J, Harrington M, Stewart C. Clinical gait analysis 1973-2023: Evaluating progress to guide the future. J Biomech. 2023 Oct 10;160:111827. doi: 10.1016/j.jbiomech.2023.111827. Epub ahead of print. PMID: 37844470.
20. Petrone, N., Vanzetto, D., Marcolin, G., Bruhin, B., & Gilgien, M. The effect of foot setting on kinematic and kinetic skiing parameters during giant slalom: Journal of Science & Medicine in Sport. 2020.
21. Pulignano, G., Sindaco, D. D., Lenarda, A. D., Alunni, G., Senni, M., Tarantini, L., Uguccioni, M. Incremental Value of gait Speed in Predicting Prognosis of Older Adults With Heart Failure. Insights from the IMAGE-HF study, 2016, pp. 289-298.
22. Hsui.H, Yang Y-L., Lin J-C. and Wu Z-H. A Wearable Inertial Measurement System With Complementary Filter for Gait Analysis of Patients With Stroke or Parkinson's Disease Changi. IEEE Access, 2016, doi:10.1109/ACCESS.2016.2633304
23. Clark, R. A., Bower, K. J., Mentiplay, B. F., Paterson, K., & Pua, Y.-H. Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. Journal of Biomechanics. 2013, Volume 46(15), pp. 2722–2725. doi:10.1016/j. jbiomech.2013.08.011
24. Allseits, E., Lučarević, J., Gailey, R., Agrawal, V., Gaunaurd, I., and Bennett, C. The development and concurrent validity of real-time algorithm for temporal Gait Analysis using inertial measurement units. Journal of Biomechanics. 2017, Volume 55, pp. 27–33. doi:10.1016/j.jbiomech.2017.02.016
25. Rancho Los Amigos Medical Center. Observational Gait Analysis Handbook. The Professional Staff Association of Rancho Los Amigos Medical Center, 1989.
26. Sina, A., Fazel, N., David, S., Golshah, N., and Morteza, A. Objective Clinical Gait Analysis Using Inertial Sensors and Six Minute Walking Test, Pattern Recognition, 2016, doi:10.1016/j.patcog.2016.08.002
27. Yang, S., Laudanski, A., & Li, Q. Inertial sensors in estimating walking speed and inclination: an evaluation of sensor error models. Medical & Biological Engineering & Computing, 2012, Volume 50(4), pp. 383–393. doi:10.1007/s11517-012-0887-7
28. Lee, J. K., & Park, E. J. Quasi-real-time gait event detection using shank-attached gyroscopes. Medical and Biological Engineering and Computing, 2011, Volume 49(6), pp. 707–712.
29. S. Joshi, S. Khanal, and R. Bista, "Spatiotemporal Gait Analysis for Cardiovascular Disease," 2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Phnom Penh, Cambodia, 2022, pp. 181-186, doi: 10.1109/SKIMA57145.2022.10029510.
30. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115
31. Fitriani, D. A., Andhyka K, W., Kom. S., Kom, M., Risqiwati, D. T. Design of Monitoring System Step Walking With MPU-6050 Sensor Based Android. Journal of Informatics, Network, & Computer Science, 2017, pp. 1-8. doi:10.21070/ joincs.v1i1.799
32. Bista, R., & Chang, J.-W. Privacy-Preserving Data Aggregation Protocols for Wireless Sensor Networks: A Survey. Sensors, 2010, 10(5), 4577–4601. doi:10.3390/s100504577
33. Pandas Dataframe, Available Online: pandas.pydata.org/DataFrame (Accessed on Aug 13, 2022)
34. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 1–1. doi:10.1109/access.2018.2870052
35. Downey, CA. Rastegari, E., Azizian, S., and Ali, H. Machine Learning & Similarity Network Approaches to Support Automatic Classification of Parkinson's Diseases Using Accelerometer-based Gait Analysis. Proceedings of 52nd Hawaii International Conference on System Sciences, 2019, pp. 4231-4242.
36. Zheng, H., Yang, M., Wang, H., & McClean, S. Machine Learning Statistical Approaches to Support Discrimination of Neuro-degenerative Diseases Based on Gait Analysis. Intelligent Patient Management, 2009, pp. 57–70. doi:10.1007/ 978-3-642-00179-6_4
37. Kwon, Y., Rivas, A. M., Zou. J., "Efficient computation and analysis of distributional shapley values" Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, PMLR 130:793-801, 2021.
38. Zhang, H., Guo, Y., & Zanotto, D. Accurate Ambulatory Gait Analysis in Walking and Running Using Machine Learning Models. IEEE Transactions on Neural Systems & Rehabilitation Engineering, 2019.
39. S. Wachter, B. Mittelstadt, and C. Russell. (2017). ''Counterfactual explanations without opening the black box: Automated decisions and the GDPR.'' [Online]. Available: https://arxiv.org/abs/1711.00399
40. X. Yuan, P. He, Q. Zhu, and X. Li. (2017). ''Adversarial examples: Attacks and defenses for deep learning.'' [Online]. Available: https://arxiv.org/abs/1712.07107
41. Ilesan, R.R.; Cordos, C.-G.; Mih ˘ail ˘a, L.-I.; Fles,ar, R.; Popescu, A.-S.; Perju-Dumbrav ˘a, L.; Faragó, P. Proof of Concept in Artificial-Intelligence-Based Wearable gait Monitoring for Parkinson's Disease Management Optimization. Biosensors 2022, 12, 189. doi.org/10.3390/bios12040189.