Biomedical Signal Processing and Control 88 (2024) 105696

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Human heart health prediction using GAIT parameters and machine learning model

Pawan Singh [a], Prabhat Singh Kourav [a], Shaurya Mohapatra [b], Vikash Kumar [a], Subrata Kumar Panda [a,c,*]

[a] Department of Mechanical Engineering, National Institute of Technology Rourkela, 769008, India
[b] Department of Prosthetics and Orthotics, Chakradhara Institute of Rehabilitation Sciences, Bhubaneshwar, Odisha 752104, India
[c] UCRD, Chandigarh University, NH-95 Chandigarh-Ludhiana Highway, Mohali, Punjab, India

ARTICLE INFO

Keywords:
Cardiovascular disease
GAIT pattern
Spatio-temporal parameters
Machine learning model
Logistic regression
Artificial neural network

ABSTRACT

This work implements different machine learning models to relate the human gait parameters to understand the heart's health. The current techniques offer a non-invasive and effective way to determine each individual's heart disease risk. Initially, the numerous gait parameters such as step length, stride length, cadence and velocity are obtained experimentally using Gait System and wearing a *retro*-reflective marker. The experimentally obtained data sets are used further to train the machine learning models, and the trained models are tested further for various gait parameters. After training and testing, the Logistic regression model shows the most effective result. The current work predicts an individual's heart health early using gait parameters and cost-effectively.

1. Introduction

An increasing number of people are losing their lives due to cardiovascular disease, further complicated by the difficulties and high costs associated with diagnosing and treating it [1]. According to estimates, 17.9 million fatalities globally from CVD, or 32% of all deaths, occurred in 2019 [2]. Heart attacks and strokes caused eighty-five percent of these deaths. Heart attacks and strokes caused eighty-five per cent of these deaths. In 2019, noncommunicable diseases were responsible for 17 million deaths among those under 70, with cardiovascular diseases accounting for 38 % [2], Refer to Fig. 1. Detection of heart health in its earlier stage is still a significant task for medical science. Several techniques are available for heart health diagnosis that are quite effective, but they must understand their cost-effectiveness concerning the existing detection methodologies. Nowadays, various unconventional ways are derived for diagnosing and treating heart diseases. However, due to some complex processes and expensive procedures, the early diagnosis of malfunctioning of an individual's heart health is not common. Heart diseases are generally diagnosed in most cases after a heart stroke and/or significant pain. In this regard, the machine learning algorithms could be adopted as a special tool for accurate and timely detection. Still, the reliability of the testing data obtained from the ECG or other traditional methodologies is a significant concern. Hence, techniques [3] like ANN, CNN, SVM, Naïve Bayes, etc., have received attention for prediction purposes due to their accuracies. Especially in the medical field, several studies were performed and observed that implementing machine learning can improve the diagnosis quality of the disease, and better results can be obtained using Artificial Intelligence (AI) and Machine Learning (ML). Wang et al. [4] implemented artificial intelligence for visually impaired persons for diagnosis, resulting in earlier and more accurate eye disease diagnosis. Hence, the patient can receive appropriate treatment. More specifically, AI is also crucial for the early diagnosis of heart problems with higher accuracy and less time. There are several techniques used for the diagnosis of heart health, such as ultrasound virtual reality, ECG [5], echocardiogram [6], cardiovascular MRI [7], CTCA [8], and many more. The analysis of the data obtained using the above methods is too lengthy, complex, and time-consuming, so AI is preferred to study data obtained

Abbreviations: CVD, Cardiovascular Disease; CHD, Coronary Heart Disease; A_o, Accuracy; S_o, Sensitivity; S_P, Specificity; P_o, Precision; R_o, Recall; F_o, F1-Score; SVM, Support Vector Machine; ANN, Artificial Neural Networks; KNN, K-Nearest Neighbours; CNN, Convolution Neural Network; LP, Linear Regression; QTM, Qualisys Track Manager.

* Corresponding author at: Department of Mechanical Engineering, National Institute of Technology Rourkela, 769008, India.
E-mail addresses: pawansingh504231@gmail.com (P. Singh), prabhatsinghkourav72@gmail.com (P.S. Kourav), shauryamohapatra8249@gmail.com (S. Mohapatra), vikashkr050sagar@gmail.com (V. Kumar), call2subrat@gmail.com, pandask@nitrkl.ac.in (S.K. Panda).

https://doi.org/10.1016/j.bspc.2023.105696
Received 25 July 2023; Received in revised form 30 September 2023; Accepted 5 November 2023
Available online 8 November 2023

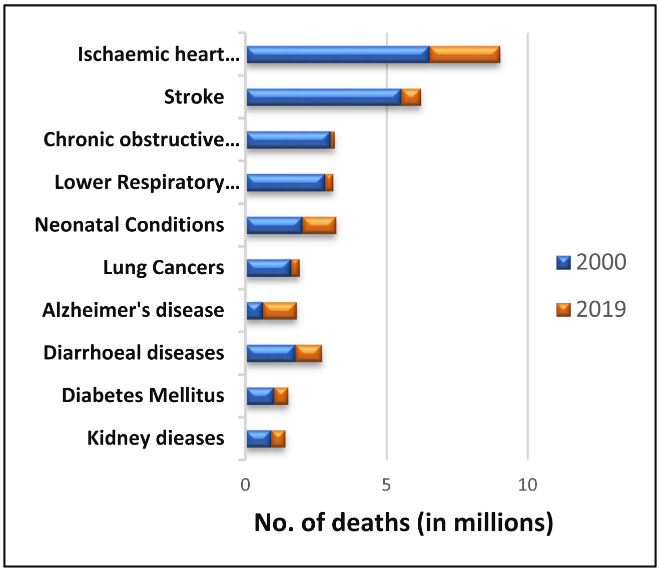

Fig. 1. Leading causes of death globally.

from the different diagnosis methods mentioned to increase the accuracy and reduce the time taken in the analysis [9].

Gait analysis is the scientific study of how people or other animals walk. It is used in primary scientific study, i.e., clinical evaluation and treatment [10]. In the gait system, several parameters are analyzed: step size, walking velocity, ground force, step count, etc. [11]. The differences observed in the gait pattern are lateral and anterior trunk deviation, abnormal hip rotation, functional leg length discrepancy, excessive knee extension, rhythmic disturbances, unnatural walking base, abnormal foot contact, etc. [12]. In the gait system, the body part's motion is observed by the camera and recorded on the system.

In contrast, the ground reaction force is recorded by either a fixed surface machine or an insole sensor system [13]. Some researchers have also used the wired insole to measure the continuous ground reaction force instead of the long-fixed measuring device [14]. Several past studies have reported the effect of heart health on the patient's gait pattern. It is found that there is a difference in walking patterns and step length in patients with peripheral artery disease (PAD) and without PAD [15]. A few studies have also been performed which detect the health status using the data of walking patterns. In a similar context, Juen et al. [16] have developed a mobile application which collects the data from the accelerometer, analyzes it and reports the person's heart health. The study indicated the accuracy of the model and software application developed, and it also notes that the gait model is more accurate for chronic disease. Similarly, there is a difference between the gait patterns of the patients with CVD and those without CVD, which can be separated with the help of CNN models.

Moreover, the ground reaction forces observed by the feet can also be analyzed in different patients. Although every person has a different gait pattern, symmetry can be observed if categorized based on their heart health [17]. Matsuzawa et al. [18] reveal a notable difference in the gait pattern of the person at risk of cardiovascular events after myocardial infarction. Recent research has been presented based on the gait variability of individuals considering varying walking speed [1], length [2] and balance abnormalities [3]. Further, the study has been presented to improve clinical assessment and prediction of mortality and morbidity among older patients undergoing cardiac surgery [4]. Considering Deep learning techniques, AI has been adopted to improve the lives of persons with oral diseases [1] or visually impaired [2]. Further, a customized CNN-based app is developed to recognize COVID-19 patients by demonstrating the damage caused in the lungs [3].

From past studies, it can be observed that there is a noticeable difference in the gait pattern of the person with and without cardiovascular disease. Few studies have reported that improving the machine learning model results in early detection of heart disease and diagnosis compared to traditional approaches. In the present work, three machine learning models (ANN, SVM, and logistic regression models) are developed, which can differentiate the gait patterns of regular and post-stroke patients. The data sets obtained using experimentation are initially used to train the machine learning models. The trained machine learning models are used further for various gait parameters and detect cardiovascular ailment in prior phases. The developed models are also detecting cardiovascular disease in its early stages. Still, they can also provide necessary information, which, in turn, helps provide adequate treatment at the right time to save lives.

2. Methods

2.1. Study participants

The dataset (gait data) utilized for this study is obtained from two laboratories. The data of normal persons have been collected from the laboratory (Applied Mechanics laboratory) of the National Institute of Technology Rourkela. In this category, ten randomly selected participants, five were males, and the remaining five were females. Each participant in the study conducted eight trials and collected 80 data sets. The participants are 19–39; their height varies from 160 cm to 179 cm, and their weight category is between 76 and 52 kg.

Similarly, the data of post-stroke patients are recorded at the laboratory of Swami Vivekanand National Institute Rehabilitation Training and Research Centre, Cuttack. In this data category, 30 post-stroke patients (twenty-three male and seven female) are considered with single trials of each person. This category of patients belongs to the Age group of 43–62 years, and their height varies between 182 cm and 160 cm, and their weight category of 57–68 kg. Hence, the dataset consists of 80 data sets from the normal person and 30 data from post-stroke patients.

2.2. Gait analysis

Gait Analysis is the study of how the body of animals moves. There is a difference in the pattern of walking of every individual; it is still observed in some studies that there is a similarity in the gait parameters of normal cases, and it can be separated from the gait parameters of patients. Several studies have been performed on specific gait parameters with heart failure in older adults. Pulignano et al. [19] studied older adults of age >70 years and observed the significant association of 1-year mortality with gait speed. This study also suggests that the gait speed should be part of the clinical evaluation process of heart health. Also, Beatty et al. [20] have conducted a 6 min walk test and reported the observations. In this study, it was observed that the patients with lower distances in the 6-minute walk test observed higher cardiovascular events than those of higher distances. From this study also, it can be concluded that the gait speed of a person is independently related to heart health. The present study is performed considering the post-stroke patients, and it is observed that there is a notable difference in the gait parameters of a normal person and post-stroke patients. In the gait analysis, the motion of a human is recorded. Different parameters are analysed, and many different parameters can be recorded, such as hip angle, knee angle, ground reaction force in separate legs, etc. This study considers only four parameters: step length, stride length, velocity, and cadence. These parameters are used to differentiate individuals' gait patterns. The details of spatial parameters (distance parameters) and temporal parameters (time parameters) are enumerated in the subsequent lines:

- Step Length – This is the distance between both feet' corresponding successive heel contact points. The length of the right step should be proportional to the length of the left step in a normal gait. This parameter can provide significant information regarding a patient's issue.

- Stride Length – This is the distance between the heel of the same foot contacting the ground in successive positions. It is equivalent to having twice the step length when walking normally.
- Cadence – This represents the number of steps taken in a given time.
- Speed (velocity) – This is the portion of ground covered by the body in a given amount of time, and it is typically measured in meters per second. Patients experiencing issues typically walk slower to reduce the forces and moments they must manage while gaiting. If the patient's condition improves, improvement in their velocity can be observed.

3. The framework of predictive model development

The step-by-step working methodology includes collecting data using a gait system, processing data using software, and then developing and training the model, shown as a flow chart in Fig. 2. The procedure followed for the study is broadly divided into two categories: data collection and model development. Data collection is the laboratory process that includes data collection and cleaning. The data is collected from the laboratory using a gait system and QTM software. Then, the raw data obtained was cleaned, and the required parameters were calculated using MOKKA software. Then, in the second phase of the study, the machine learning models were prepared. The prepared models were trained and tested for accuracy with the collected data by categorizing the data set into training and testing datasets. The best model can be selected based on their quality parameters, such as accuracy, precision, recall and F-score. The finalized model can be used to observe the heart health of any random person using the gait parameters as input.

3.1. Problem definition

The present works use machine learning methods to distinguish between normal people and stroke survivors using gait-related parameters. The problem is framed as a classification task to identify normal versus post-stroke gait patterns. Initially, the model is trained on labelled gait data, and further Supervised learning technique is utilized to complete the task. This current model investigates various classification techniques, including Logistic regression, SVM, and ANN. The result of this model can be used as a tool for early detection and diagnosis of gait abnormalities in post-stroke patients, improving treatment planning and progress monitoring.

3.2. Data collection

The data can be taken from any open-source platform where it is available in a usable format, or it can be experimentally recorded by the setup available in the laboratory. As mentioned, few studies are

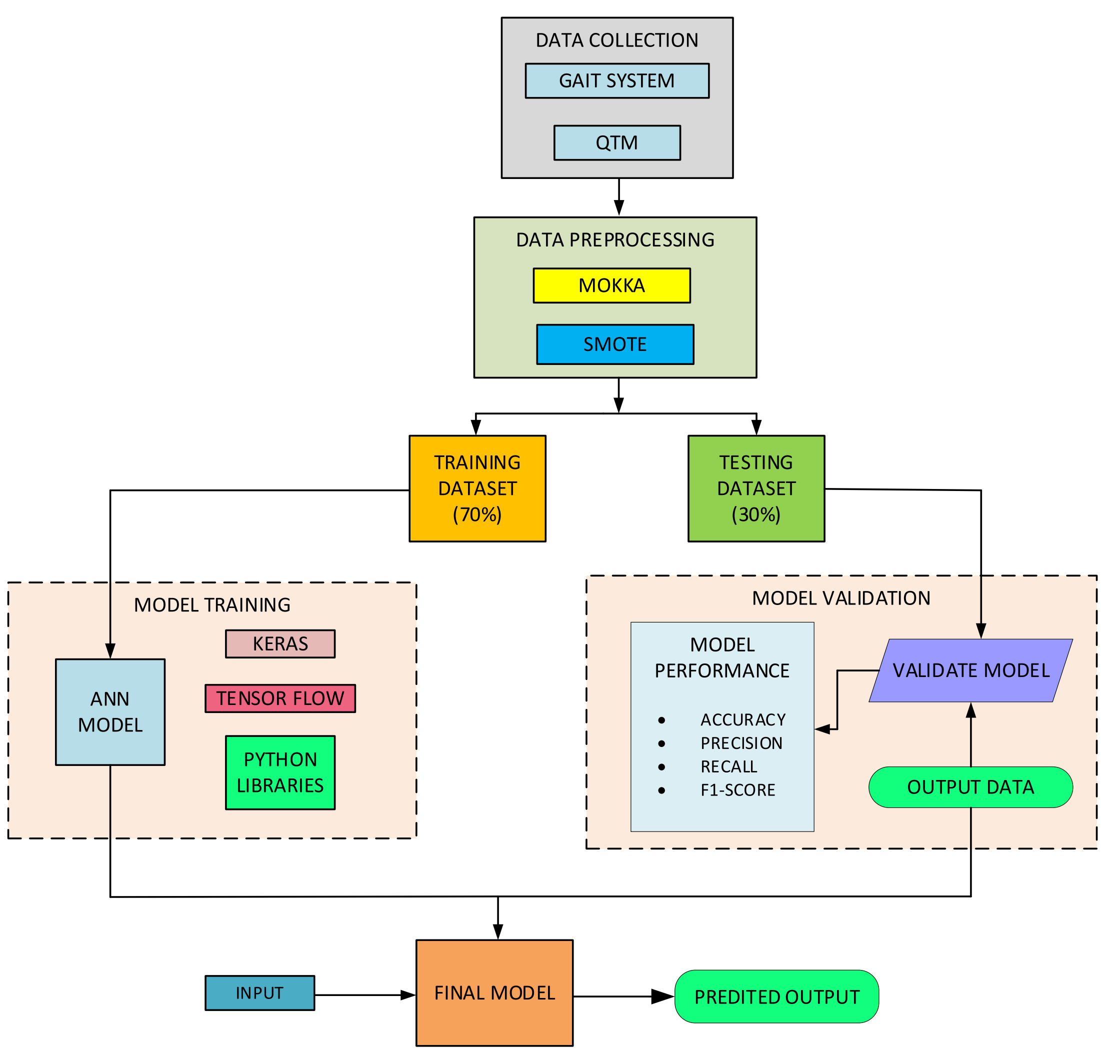

Fig. 2. Methodology of project.

available, so the dataset is unavailable on open-source platforms. Hence, the data is collected from the laboratory (Applied Mechanics laboratory) at the National Institute of Technology Rourkela. Fig. 3 depicts the experimental setup, which consists of four high-speed Oqus 5.0 motion capture cameras surrounding the 6 m walkway (integrated Kistler's force plate (Fig. 3), along with the passive *retro*-reflective markers for tracking the motion of the participants and all the equipment are integrated with the computer system utilizing a control box and software named QTM (Quality Track Manager). It collects the corresponding data at a frequency rate of 100 Hz. QTM software integrates the camera and the force plate with the computer system. It is licensed software available at the applied mechanic's lab of NIT Rourkela. Record the observation; the person is required to wear the markers, and their motion is recorded in the form of an $\times$,*y*, and *z*-coordinate system to time. Each subject has to walk on a walkway wearing the markers. In this study, the data is recorded by taking eight trials of each person, every four trials, with each left and right leg on the force plate. The recorded data from the lab have some noise introduced during recording, so it is filtered with a second-order Butterworth filter (low-pass with a cut-off frequency of 10 Hz) to eliminate the noise and get usable numbers [21].

The second set of data, i.e., the data of post-stroke patients, is collected from the motion laboratory of Swami Vivekananda National Institute of Rehabilitation Training and Research Centre (SVNIRTAR) Olatpur, Cuttak. This laboratory also consists of a walkway surrounded by cameras. It includes passive *retro*-reflective markers for tracking down the motion of the participants, integrated with a BTX force plate (Type – 600, dimension- 600 × 500 × 50 mm). The data set (post-stroke patients) is obtained using the steps discussed in the above sections. Further, the recorded data is processed and converted into the desired usable format.

3.3. Data pre-processing

Data pre-processing ensures that the data is correct, comprehensive, and usable for machine learning employing a series of operations, including cleaning, transformation, and selection. To create an effective machine learning model, pre-processing of data is essential. Firstly, the experimental data collected from the laboratories needs to be separated into two categories, i.e., the data associated with normal and abnormal cases. Furthermore, the model is trained to learn the differences between these two categories. The steps used in data preparation are enumerated in subsequent lines:

(1) The data collected from the gait system can be obtained in multiple formats. The data files in the present work are in c3d file format, which is made further readable utilizing MOKKA software.
(2) The 3D cartesian location of selected markers and the time are extracted in CSV file format using the MOKKA software. Each marker location is extracted according to the marker locations as per necessary calculations required for gait parameters. The values for every 0.01 s and marked sensors, i.e., x, y, and z-locations, are collected for the input.
(3) In the case of normal person data, the location of three markers is extracted for the study, which is in red dots, as shown in Fig. 4. Further, the manual data analysis is done to obtain the different gait parameters for the study. In the case of post-stroke patients, all the gait parameters are directly extracted from the software.
(4) The location of the marker at a particular time is considered a point in space, and the distance between two locations in 3D space can be calculated using the vector length formula:

$$\overrightarrow{AB} = B(x,y,z) - A(x,y,z) \tag{1}$$

Based on Eq. (1), the step length and stride length can be calculated while the cascade can be computed as:

$$\text{Cadence} = \frac{\text{Number of Steps}}{\text{End Time - Start Time}} \tag{2}$$

$$\text{Velocity} = \frac{\text{Final location - Initial location}}{\text{Final time - Initial time}} \tag{3}$$

(5) The dataset is combined and randomly shuffled so that data from both types can be used for training. Combining and randomly shuffling the dataset can ensure that data from all categories are included. It eliminates the possibility of interference or bias introduced by the original order of the data by shuffling it around.
(6) Now, the data is visualized to understand how it is organized, and the relationships between the attributes are seen. This is accomplished with the help of Seaborn, a Python library that generates a matrix of scatter plots displaying the pairwise relationships between all variables in a given dataset. Four attributes are mainly used here; the matrix will be four rows and four columns.

Fig. 3. Gait System 6 m long platform and subject wear *retro*-reflective marker [17].

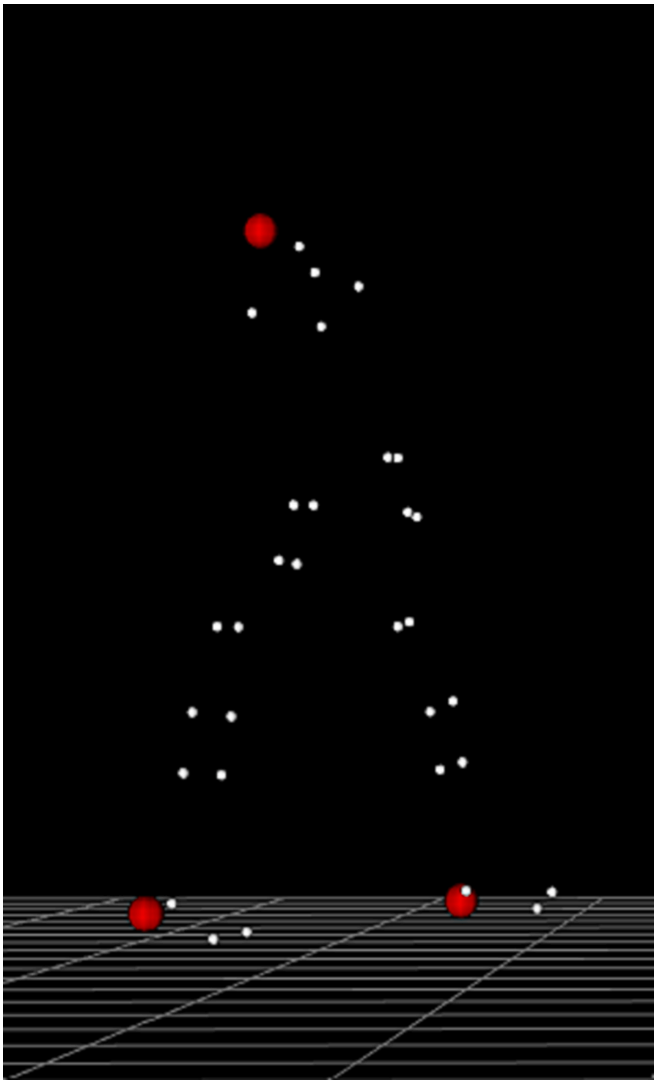

Fig. 4. Visualization of c3d file in MOKKA software.

The histograms of each variable in the dataset are displayed along the diagonal of the matrix.

(7) Initially, some dataset is required for training and testing the model's quality. In this respect, the dataset is split into training and testing subsets in the provided code by calling the train-test-split function in the sklearn model selection module. The training set is the data set from which the model learns. A testing set is used to check the accuracy, precision, recall, and F1-score to check the quality of the model.

Finally, the data will be split up so that a machine learning model can be trained on the training subset and tested on the testing subset to determine how well the model performed. Overfitting can be avoided using this approach, which occurs when a model is overly complex and performs well on the training data but poorly on new data that has yet to be seen. Initially, there is a total of 110 data points, and then it is split into datasets having 88 and 22 data points.

4. Different machine learning models

Machine learning is a subfield of AI that uses algorithms to train computers to improve a given task over time through observation and experience rather than manual instruction. There are several applications of different machine learning models, out of which many are used to solve real-world problems. Based on the type of problem and requirements, different machine learning models can be preferred after analyzing their quality parameters. Recently Wang et al. [22] developed a web application for the COVID-19 recognition. The results of the study show good results in diagnostic prescription as well as in medical education. The diversity in the application of artificial intelligence can be recorded and utilized for different purposes in various fields. Huang et al. [23] used deep learning to analyze oral health data for image segmentation and recognition tasks. The study found promising results and suggested further improvement in the work. The present work uses supervised learning to train the model based on labelled data. A training dataset is utilized to learn a model, and then the model is validated with the help of a testing dataset. The three machine learning models were proposed for the prediction and tested in the present work, shown in Fig. 5.

4.1. Logistic regression model

Logistic regression is a statistical model which is used for binary classification problems, especially in which the target variable can take only one of two possible values. In this research project, the gait parameters are analyzed with logistic regression to determine whether they should be considered normal or abnormal. The logistic regression model estimates the likelihood that a given input will be assigned to a particular category. Establishing a threshold value converts this probability into a decision regarding binary classification. For instance, if the threshold value is set to 0.5, any input with a chance greater than or equal to 0.5 is considered abnormal.

In contrast, any input with a possibility that is lower than 0.5 is considered normal as in [24]. Fitting a logistic function to the input data makes logistic regression work. The logistic function converts the data that is fed into it into a probability value that falls somewhere between 0 and 1. The logistic function is defined as:

$$p(a) = \frac{1}{(1 + e^{(-b)})} \tag{4}$$

where $p(a)$ is the probability of the input being classified as abnormal, 'e' is the exponential function, and 'b' is a linear function of the input features.

The performance of the trained logistic regression model is then assessed using the test data. Accuracy, precision, recall, and F1-score are used further to evaluate the models in this study.

4.2. Support vector machine model

A powerful machine learning technique, SVM excels and resolves various classification and regression issues. This study used SVM as a binary classification model to reliably categorize gait parameters as normal or abnormal. SVM performs its function by constructing a hyperplane that divides the input data into distinct classes. This is done with the assistance of a boundary designed to maximize the difference between the two classes. The margin is the distance between the hyperplane and the data points belonging to each class that is closest to it, and the SVM algorithm attempts to determine which hyperplane is the

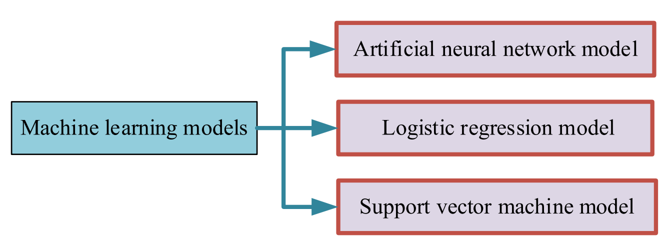

Fig. 5. Different machine learning models proposed in the work.

one that maximizes this margin. It is highly versatile and can be applied to both linear and nonlinear classification problems; this enables it to capture complex patterns in the data. After training the SVM model, its efficacy is measured with testing data. A confusion matrix was created by comparing the predicted labels to the actual labels for the testing data, and accuracy, precision, recall, and score were used as evaluation metrics.

4.3. Artificial Neural Network (ANN)

A machine learning algorithm (ANN) mimics the structure and behaviour of the human brain. It is built from many interconnected nodes, which act as artificial neurons, and processes data by applying a series of mathematical transformations, as shown in Fig. 6. Data is taken in by the input layer, processed, passed on to the next layer, and so on until the output layer generates the final result. The vital step in machine learning is training. The prepared data is fed into the machine learning model during training to discover patterns and generate predictions. As a result, the model learns from the data and completes the assigned goal. The model improves in predicting over time as it is trained. The model has one hidden layer with 100 nodes by default, and its input and output layers are of variable size depending on the training data and the desired outcomes. Rectified linear unit (ReLU) is a nonlinear activation function used in the hidden layer of the model. The logistic sigmoid activation function is the output layer's function when performing binary classification tasks.

Adam, the default optimization algorithm, adjusts the training speed of each parameter according to the gradient's momentum. The model is given multiple opportunities to go through the training data, known as "epochs," during the training process. At each time step, the model adjusts its settings following the gradient of the loss function relative to those settings. Doing so iteratively teaches the model to improve its predictions and reduces the loss. The model uses L_2 regularisation with a regularisation strength of 0.0001 to prevent overfitting by punishing highly high weights. The random state parameter is not explicitly set, so it takes an arbitrary value during model initialization. The training data is shuffled before each epoch by default, which can help improve the learning process.

4.4. Classification of evaluation metrics

After each model is prepared, its effectiveness is measured primarily through accuracy, precision (positive predictive value), recall (sensitivity or true positive rate), and F-score. The metrics are derived in this work as follows:

- Accuracy is the proportion of correctly predicted data from the training dataset found in the testing dataset relative to the total number of observations.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{5}$$

- A model's precision is evaluated by accurately estimating the number of positive cases. It indicates how many of the predicted positives actually turned out to be positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

- The number of positive samples that were accurately detected by the classifier out of the total number of samples that could have been positive is referred to as the recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

- The F-score, also called the F1-score, measures a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative.'

$$F - \text{Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{8}$$

where TN represents the total of patients who have determined correctly that they do not have CHD. FN is the total of all patients who were wrongly told they did not have CHD. TP denotes the total of people who have their CHD diagnosis confirmed. FP is the total number of patients improperly diagnosed with CHD.

5. Results

The study addressed the challenges posed by expensive and sometimes uncomfortable heart health monitoring techniques such as ECG, echocardiogram, and X-rays. It explored the potential of utilizing gait patterns as a novel and cost-effective approach for heart health prediction.

Data collection involved recording the gait patterns of both normal individuals and those with heart-related issues. The data included in the study were obtained from two sources: the gait patterns of normal

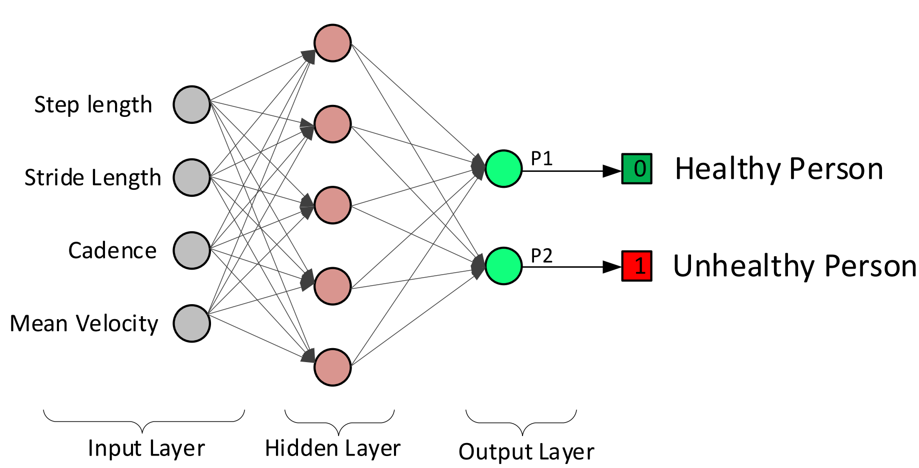

Fig. 6. Workflow of the ANN Mode.

individuals were recorded at the National Institute of Technology Rourkela biomechanics laboratory, and data from post-stroke patients were acquired from the Swami Vivekananda National Institute of Rehabilitation Training and Research Laboratory. Few data collected and processed are shown in Figs. 7, 8 and 9. The compiled data file is visualized using the MOKKA software, the visualization of the markers is obtained on the visual screen, and the location of the markers extracted in the file and sample of extracted data is shown in Fig. 7. The collected data is combined and randomly distributed in the file to avoid biased training. Fig. 8 is a representation of the datasets in combined and random form. The scatter plot is drawn by utilizing the data sets for visualization purposes. The scatter plots demonstrate how each of the variables in the dataset is connected, as shown in Fig. 9. Every dot on the figure represents a single data point, and its location reflects the corresponding values of the two variables for the specific data point. A diffuse and dispersed pattern shows a weak association, and a collection of densely packed dots indicates a strong connection. The red dots represent patients' data, while the blue one represents the data of healthy persons. The figure shows that the patient's data of each parameter are diffuse and dispersed, while the data of healthy persons are intensely dense and related to each other. The highly thick dots represent the strong relationship between the parameters, meaning the variation of one parameter will significantly affect another. The figure shows that for the healthy person's data, each parameter is strongly related to the other.

In contrast, the patients' data showed less dense plots due to the wide range of study participants. From the plot, it can also be observed that each figure follows more or less the same pattern. It is one of the key findings that can be observed in scatter plots.

Upon combining the data from healthy and unhealthy individuals, notable differences in gait patterns were observed. However, manual differentiation of these patterns proved to be challenging. Consequently, machine learning techniques were implemented to address this issue. The dataset was divided into testing and training sets for model development and evaluation.

During the training phase, the machine learning models learned to differentiate between individuals with and without cardiovascular disease by observing the patterns in the data. Subsequently, during the testing phase, the models were evaluated using metrics such as accuracy, precision, recall, and other quality-checking factors. Fig. 10 shows the data validation results. After the training, inputting the data into the model calculates the overall impact on both the parameters on 1 and 0. Based on this value, heart health is displayed as a healthy case or not. Similarly, Fig. 11 compares the accuracy obtained by the model during its training and validation phase. It can be clearly seen from the graphs that in the initial state of both degrees, the accuracy of the training phase is much more than the validation phase. However, as the training is completed, the training and validation accuracy is comparable.

The analysis of the models revealed that logistic regression outperformed the other two models, achieving an impressive F1 score of 96.87%, an accuracy of 96.87%, precision of 97.05%, and recall of 96.87%. Support vector machines (SVM) also demonstrated admirable performance, with an F1 score of 95.31%, accuracy of 95.45%, precision of 95.72%, and recall of 95.45%. However, artificial neural networks (ANN) exhibited slightly lower accuracy, precision, recall, and an F1 score of 90.26%, with an accuracy of 90.90% and precision of 91.91%. The various parameters of different machine learning models are also listed in Table 1.

6. Discussion

The results of this study have important implications for the field of cardiovascular health monitoring. Traditional diagnostic methods like ECG, echocardiograms, and X-rays are not only expensive but can also be uncomfortable for patients. Pursuing alternative, more accessible, and patient-friendly diagnostic techniques is crucial in healthcare. In this context, the study's exploration of gait pattern analysis and machine learning models is noteworthy. It opens up the possibility of early detection of heart disease, a critical factor in improving patient outcomes. These models demonstrate their potential as reliable diagnostic tools by achieving high accuracy, precision, and recall values, particularly in logistic regression and support vector machines.

The study's success in utilizing gait patterns for heart disease diagnosis highlights the evolving role of machine learning in healthcare. Machine learning models showcased their ability to discern intricate gait patterns associated with heart disease, a testament to their capability to learn and adapt. This suggests that machine learning can play a pivotal role in augmenting the diagnostic capabilities of healthcare professionals.

However, while the results are promising, further research is warranted. Exploring more advanced machine learning algorithms, optimizing model parameters, and integrating additional features or data sources could lead to even more precise diagnostic tools. Collaborating between researchers and healthcare professionals remains crucial for validating these models in real-world clinical settings. The insights and expertise of clinicians are indispensable in bridging the gap between research findings and practical healthcare applications.

7. Conclusion

The feasibility of using machine learning methods for diagnosing heart disease has been investigated in the present paper. For the dataset of 110 data points, distributed evenly between 30 negative and 80 positive cases, three models are tested, i.e., logistic regression, SVM, and ANN. The necessary findings are outlined in the subsequent section:

- The logistic regression and SVM models performed well with high accuracy, precision, recall, and F_1 scores, indicating their effectiveness in differentiating the gait pattern of post-stroke patients from that of normal individuals. However, the ANN model performed worse on all measures due to the smaller dataset size. The ANN model's performance can be improved with larger datasets.

Time	R_ICT			R_FM1			L_FM1		
s	mm	mm	mm	mm	mm	mm	mm	mm	mm
	X	Y	Z	X	Y	Z	X	Y	Z
0.01	-712.853	140.138	965.875	-334.597	246.195	131.315	-790.825	299.275	42.0101
0.02	-707.399	139.749	964.185	-327.614	246.799	130.816	-790.702	299.267	42.0137
0.03	-701.669	139.439	962.546	-320.563	247.026	130.341	-790.577	299.279	42.0472
0.04	-696.04	139.207	960.732	-313.621	247.188	129.66	-790.523	299.229	42.0704
0.05	-690.181	138.899	959.116	-307.12	247.141	128.668	-790.478	299.124	42.0125
0.06	-684.213	138.626	957.692	-300.427	246.858	127.66	-790.496	298.975	41.7131
0.07	-678.32	138.476	955.938	-294.238	246.468	126.099	-790.616	298.825	41.5092
0.08	-672.18	138.122	954.578	-288.23	246.108	124.427	-790.718	298.665	41.2014

Fig. 7. Extracted location of markers sample from MOKKA software in CSV format.

S.No	Step length	Stride length(mm)	Cadence (steps/min)	Mean Velocity (m/s)	Target \nClass
1	641.500000	1278.071525	104.651128	1.150198	0
2	643.250000	1280.656000	104.651163	1.150198	0
3	632.000000	1287.499000	104.046243	1.145443	0
4	676.000000	1403.557000	107.142857	1.282100	0
5	617.000000	1226.707000	83.333300	0.889630	0

Fig. 8. Sample dataset.

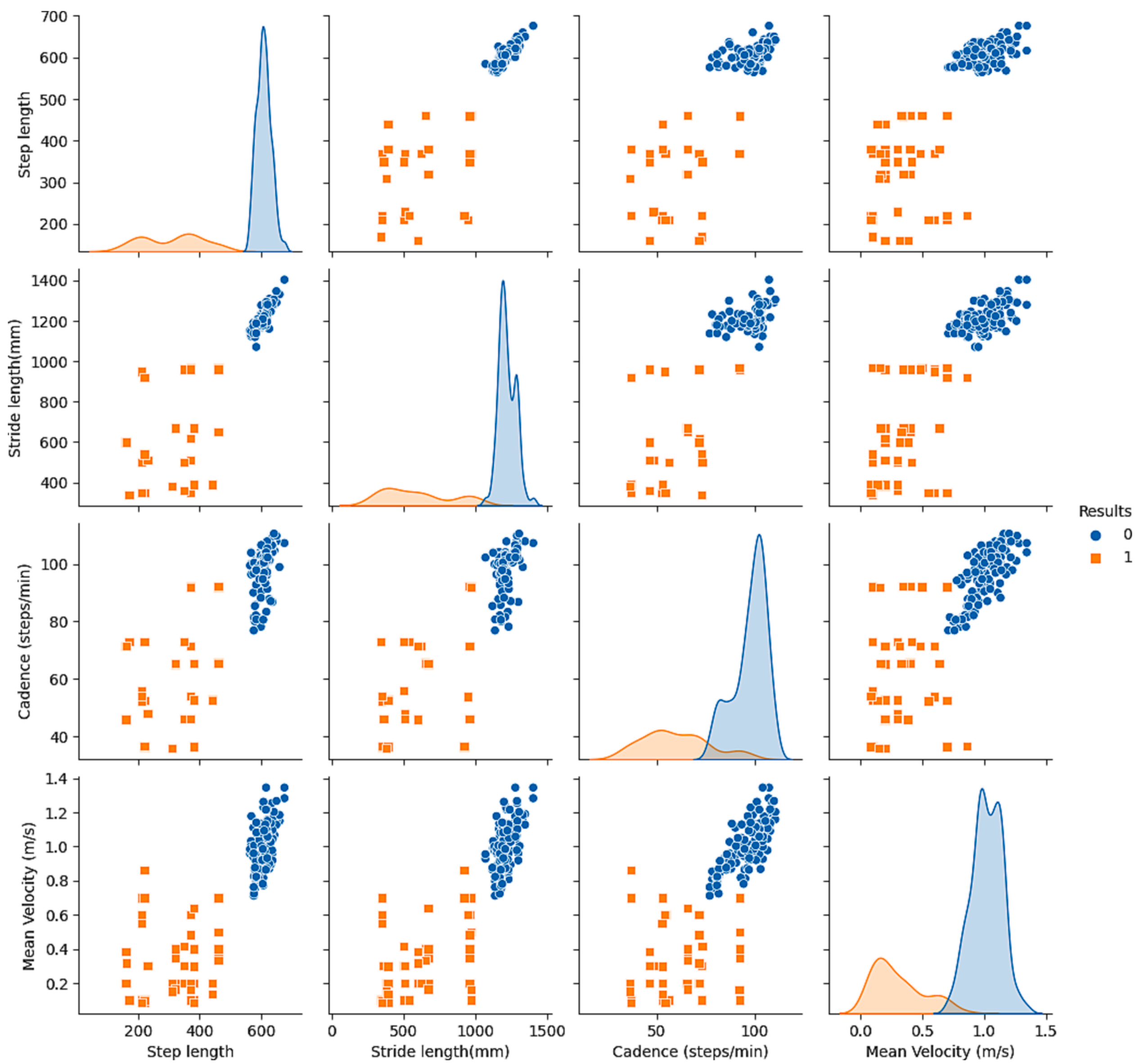

Fig. 9. Scatter plot of the dataset.

- Machine learning techniques have significant potential to diagnose heart disease. These techniques offer a non-invasive and effective method for determining which individuals are at risk of heart disease. More research is required to investigate the possibility of using various models and datasets to enhance the effectiveness of machine learning models in diagnosing heart disease.
- Machine learning models such as logistic regression, support vector machines, and artificial neural networks (ANN) can potentially be effective in diagnosing heart disease. The present research has significant consequences for the medical field, where it can be used as a helpful resource for diagnosing and treating heart disease at an earlier stage and ultimately leads to better health of patients.

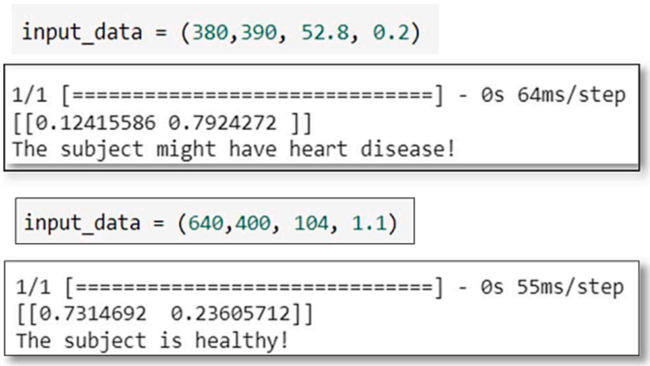

Fig. 10. Data validation outputs.

Funding

This research received no external funding.

Ethical and informed consent statement

Informed consent was obtained from all subjects involved in the study.

CRediT authorship contribution statement

Pawan Singh: Conceptualization, Methodology, Formal analysis, Investigation. **Prabhat Singh Kourav:** Conceptualization, Methodology, Formal analysis, Investigation. **Shaurya Mohapatra:** Conceptualization. **Vikash Kumar:** Formal analysis, Investigation. **Subrata Kumar Panda:** Conceptualization, Supervision.

Table 1
Evaluation parameter results of different machine learning models.

Parameters	Logistic Regression (%)	SVM (%)	ANN (%)
Accuracy	96.87	95.45	90.90
Precision	97.05	95.72	91.91
Recall	96.87	95.45	90.90
F1 Score	96.87	95.31	90.26

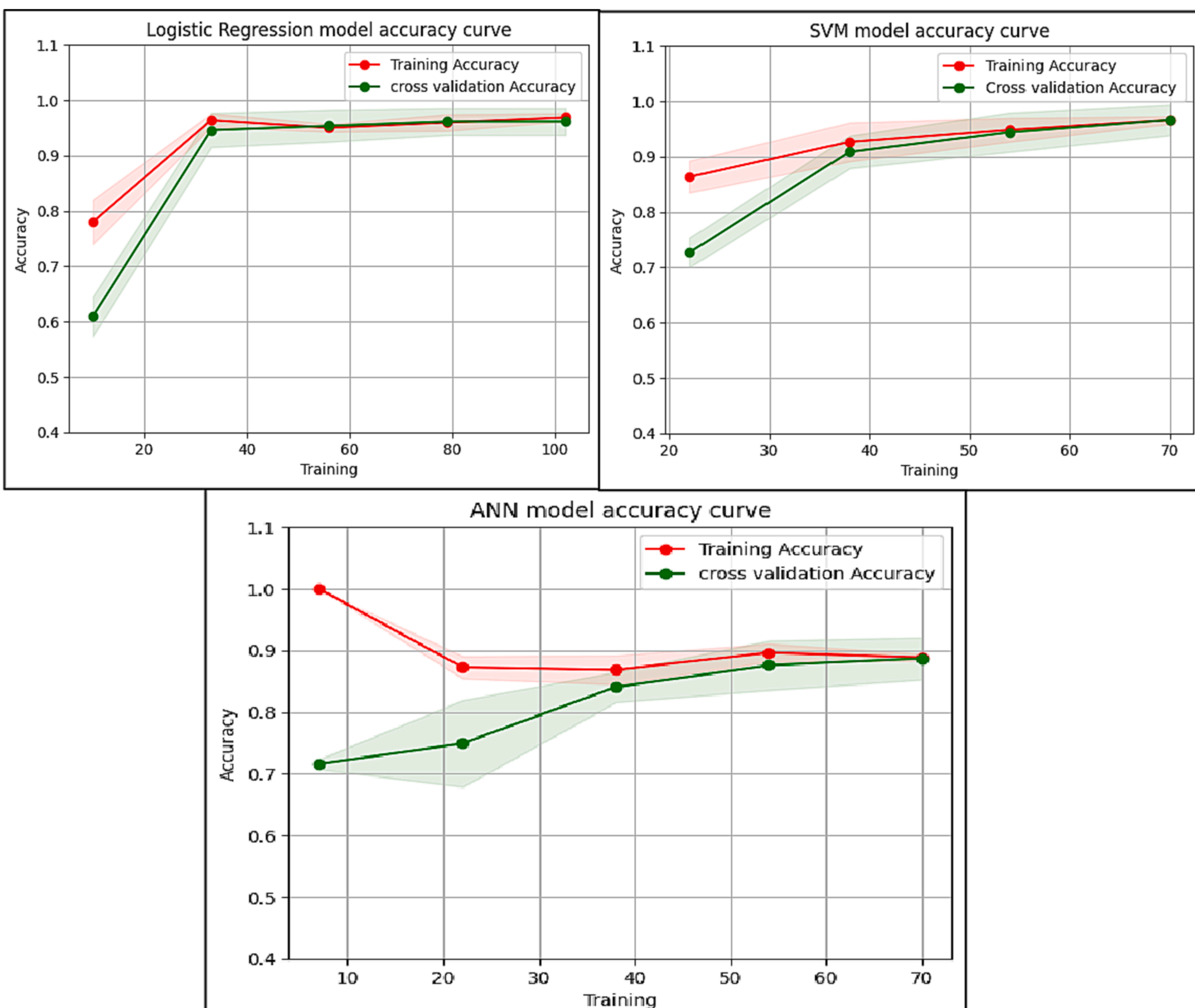

Fig. 11. The Accuracy Comparison curve of the Logistic Regression model, SVM model, and ANN model for Training and Validation set.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors would like to express special thanks to **Prof Thirugnanam** for providing the data of healthy cases from the biomechanics laboratory of NIT Rourkela. The authors are further thankful to **DR. Saurya Mohapatra** for his contribution to the study by providing the data of post-stroke patients from the laboratory of SVNITRAR and also expressing gratitude to **DR. Shreeshan Jena** for his kind support in understanding and implementing this research.

References

[1] S. Pathberiya, Patient autonomy and decision making, Rapid Repson. Mak. Evid. based Med. Work Individ. patients (2016). https://www.bmj.com/content/353/bmj.i2452/rr-17,.
[2] WHO, No Title, Top 10 Causes Death, 2020.
[3] A. Rajkumar, G.S. Reena, Diagnosis of heart disease using datamining algorithm, Glob. J. Comput. Sci. Technol. 10 (10) (2010) 38–43.
[4] S. Nandy, M. Adhikari, V. Balasubramanian, V.G. Menon, X. Li, M. Zakarya, An intelligent heart disease prediction system based on swarm-artificial neural network, Neural Comput. Appl. 1 (2021), https://doi.org/10.1007/s00521-021-06124-1.
[5] A. Rosiek, K. Leksowski, The risk factors and prevention of cardiovascular disease: the importance of electrocardiogram in the diagnosis and treatment of acute coronary syndrome, Ther. Clin. Risk Manag. 12 (2016) 1223–1229, https://doi.org/10.2147/TCRM.S107849.
[6] S. Ghosh, B.P. Chattopadhyay, R.M. Roy, J. Mukherjee, M. Mahadevappa, Estimation of echocardiogram parameters with the aid of impedance cardiography and artificial neural networks, Artif. Intell. Med. 96 (February) (2019) 45–58, https://doi.org/10.1016/j.artmed.2019.02.002.
[7] N. van der Velde, et al., Detection of subclinical cardiovascular disease by cardiovascular magnetic resonance in lymphoma survivors, JACC CardioOncol. 3 (5) (2021) 695–706, https://doi.org/10.1016/j.jaccao.2021.09.015.
[8] W.B. Meijboom, et al., Pre-operative computed tomography coronary angiography to detect significant coronary artery disease in patients referred for cardiac valve surgery, J. Am. Coll. Cardiol. 48 (8) (2006) 1658–1665, https://doi.org/10.1016/j.jacc.2006.06.054.
[9] S. Malakar, S. D. Roy, S. Das, S. Sen, J. D. Velásquez, R. Sarkar, Computer Based Diagnosis of Some Chronic Diseases: A Medical Journey of the Last Two Decades, no. Ml. Springer Netherlands, 2022. doi: 10.1007/s11831-022-09776-x.
[10] A. Zaroug, A. Garofolini, D. T. H. Lai, K. Mudie, R. Begg, Prediction of gait trajectories based on the Long Short Term Memory neural networks, PLoS One 16 (8) (2021), doi: 10.1371/journal.pone.0255597.
[11] S. Jena, G.M. Sakhare, S.K. Panda, A. Thirugnanam, Implementation of multiple regression technique for detection of gait asymmetry using experimental gait data, J. Med. Biol. Eng. 41 (1) (2021) pp, https://doi.org/10.1007/s40846-020-00533-8.
[12] WEJ, Gait analysis, vol. 4, no. 1. Butterworth-Heinemann Ltd, 1984. doi: 10.1016/S0737-0806(84)80104-5.
[13] M. Iwai, et al., The validity of spatiotemporal gait analysis using dual laser range sensors: a cross-sectional study, Arch. Physiother. 9 (1) (2019) 1–8, https://doi.org/10.1186/s40945-019-0055-6.
[14] S.J.M. Bamberg, A.Y. Benbasat, D.M. Scarborough, D.E. Krebs, J.A. Paradiso, Gait analysis using a shoe-integrated wireless sensor system, IEEE Trans. Inf. Technol. Biomed. 12 (4) (2008) 413–423, https://doi.org/10.1109/TITB.2007.899493.
[15] M. Szymczak, P. Krupa, G. Oszkinis, M. Majchrzycki, Gait pattern in patients with peripheral artery disease, pp. 1–7, 2018, doi: 10.1186/s12877-018-0727-1.
[16] J. Juen, Q. Cheng, V. Prieto-centurion, J. A. Krishnan, B. Schatz, Health Monitors for Chronic Disease by Gait Analysis with Mobile Phones 20(11) (2014) 1035–1041. doi: 10.1089/tmj.2014.0025.
[17] S. Jena, G. Mohan, S. Subrata, K. Panda, Implementation of multiple regression technique for detection of gait asymmetry using experimental gait data, J. Med. Biol. Eng. 1 (0123456789) (2020) pp, https://doi.org/10.1007/s40846-020-00533-8.
[18] Y. Matsuzawa, M. Konishi, E. Akiyama, Association between gait speed as a measure of frailty and risk of cardiovascular events after myocardial infarction, JAC 61 (19) (2013) 1964–1972, https://doi.org/10.1016/j.jacc.2013.02.020.
[19] G. Pulignano, et al., Incremental value of gait speed in predicting prognosis of older adults with heart failure: insights from the IMAGE-HF study, JACC Hear. Fail. 4 (4) (2016) 289–298, https://doi.org/10.1016/j.jchf.2015.12.017.
[20] M.A. Beatty, L. Alexis, N.B. Schiller Whooley, Six-minute walk test as a prognostic tool in stable coronary heart disease: data from the Heart and Soul Study, Mol. Cell. Biochem. 23 (1) (2012) 1–7, https://doi.org/10.1001/archinternmed.2012.2198.Six-minute.
[21] S. Jena, G.M. Sakhare, S.K. Panda, A. Thirugnanam, Evaluation and prediction of human gait parameters using univariate, multivariate and stepwise statistical methods, J. Mech. Med. Biol. 17 (5) (2017) 1–13, https://doi.org/10.1142/S0219519417500762.
[22] J. Wang, S. Wang, Y. Zhang, Artificial intelligence for visually impaired, Displays 77 (February) (2023), 102391, https://doi.org/10.1016/j.displa.2023.102391.
[23] C. Huang, J. Wang, S. Wang, Y. Zhang, A review of deep learning in dentistry, Neurocomputing 554 (April) (2023), 126629, https://doi.org/10.1016/j.neucom.2023.126629.
[24] P. Chinnasamy, S. Arun Kumar, V. Navya, K. Lakshmi Priya, S. Sruthi Boddu, Machine learning based cardiovascular disease prediction, Mater. Today Proc. (xxxx) (2022), https://doi.org/10.1016/j.matpr.2022.04.907.