MSc Data Science Project Report

Clinical movement patterns and predictive models of long COVID

prepared by

Harikrishnan Mohanan

Project Supervisor: Peter Ellison

Date: 20/09/2024

No. of Words: 5932

2023/24 Entry Cohort

MSc Data Science: MSc Project

Declaration of Help Received

Name of Student: Harikrishnan Mohanan

Project Supervisor(s): Peter Ellison, Dr. Adar Pelah

This form should be completed by each student and submitted as part of the report. The information provided allows proper acknowledgement of the help you have received and will help the markers in evaluating your performance.

1. Did anyone other than the project supervisor(s) assist you in carrying out your project?

   Tick the appropriate box:

   ☐ Yes — ***Complete Section 2***

   ☑ No — ***Go to section 3***

2. Detail the help received (Suggested categories are given below the table):

Complete in block capitals

Name of Helper	Category (i to v)	Details (if necessary)

(i) Day to day help
(ii) Initial training in a technique or operation
(iii) Specific assistance (e.g. correction of code)
(iv) Help with interpretation of results
(v) Any other category of help

3. I have read the Assessment Information in the MSc Handbook on academic misconduct and plagiarism. I have completed the Academic Integrity Tutorial.
   I declare that this assessment is a presentation of original work, I am the sole author and I am not attempting to pass off work created by generative AI content tools as my own.

Name of student (in lieu of signature)

Harikrishnan Mohanan
…………………………………………………………………………………….Date…………20/09/2024…………

Clinical movement patterns and predictive models of long COVID

1. Abstract

This study looks into using gait analysis as a way to diagnose and monitor long COVID, a condition where symptoms last long after a COVID-19 infection. Researchers studied 170 participants, including long COVID patients and healthy controls, to find small changes in how people move. They used several machine learning models, such as Logistic Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine (SVM), and a Voting Classifier, to classify people based on their walking features like speed, step length, and step width. The Voting Classifier performed the best, with 68% accuracy, while Random Forest and XGBoost were close behind at 67%. These models were somewhat effective in telling apart those with and without COVID-19, with AUC values ranging from 0.58 to 0.74. The study found slight differences in gait, like faster walking speed and narrower step width in the COVID group, hinting at possible neuromuscular problems. These results suggest that gait analysis could help identify long COVID, but more research with larger datasets is needed to improve accuracy and find more markers for diagnosis.

2. Introduction

COVID-19 is a complex disease that impacts multiple organ systems and physiological processes. Although it is primarily a respiratory tract and lung infection, COVID-19 can involve the central nervous system and lead to neurologic symptoms and disorders as well. Recent data show that neurological effects of COVID-19, such as encephalitis, seizures, necrotizing encephalopathy, and strokes, may result in disabling aftereffects or ongoing symptoms known as long-haul COVID or long COVID (Balcom, Nath, and Power 2021). It is a range of symptoms that can last for weeks or months and can affect anyone who has had COVID-19 with no established diagnostic criteria (Phillips and Williams 2021). It is characterized by symptoms extending beyond the initial illness phase, long COVID includes a wide variety of severe physiological and psychological indications. Commonly reported symptoms of long COVID include fatigue, breathlessness, anxiety, cognitive impairment and brain fog which affect concentration and memory (Lechner-Scott et al. 2021). Musculoskeletal issues such as fatigue, muscle pain and joint pain were reported after severe COVID-19 (Godaert et al. 2020). A few studies also mentioned that long COVID symptoms can persist for up to 3 to 6 months (Abdullahi et al. 2020; Karaarslan, Güneri, and Kardeş 2022). These symptoms after severe COVID-19 which mainly affect the respiratory system can affect a person's gait (Keklicek et al. 2022). Long COVID is estimated to impact a substantial proportion of COVID-19 survivors, posing a significant challenge to both patients and healthcare systems. Addressing this condition is crucial for improving the quality of life for those affected and reducing the long term burden on healthcare resources.

Gait is an important routine function which reflects the final outcome of a person's neuromotor control. Human gait involves load, motion and cycle. This includes variations in speed, gait width and lower limb loading times (spatial and temporal parameters) as well as changes in motion ranges (kinematic parameters) or ground reaction forces (Jamari et al. 2022). Figure 1 illustrates key gait parameters such as step length, stride length, and step width, which are critical in assessing gait patterns and potential abnormalities. Gait analysis provides valuable insights into how the musculoskeletal system interacts with other internal systems (Hausdorff and Alexander 2005). It has been reported that various neurological damages following COVID-19 may impact gait by disrupting cerebellar coordination (Helms et al. 2020). In a study, it was reported that coordination and gait disturbances were detected in an individual after acute COVID-19 even though he did not have a history of neurological disease in the past (Pistoia et al. 2021). Understanding the movement patterns linked to long COVID is important for several reasons. First, many people with long COVID experience problems with mobility, balance, and coordination, which can seriously affect their daily lives and overall well-being. Second, identifying these movement patterns can help us learn more about the disease itself and the reasons behind ongoing symptoms.
Studying these patterns can also help create models that predict who might develop long-lasting symptoms, allowing for early treatment and customized rehab plans. Additionally, this knowledge can improve patient care, guide treatments, and lead to better long-term outcomes. By focusing on movement patterns, researchers and healthcare providers can gain a clearer picture of the full impact of long COVID, leading to more effective care for those affected.

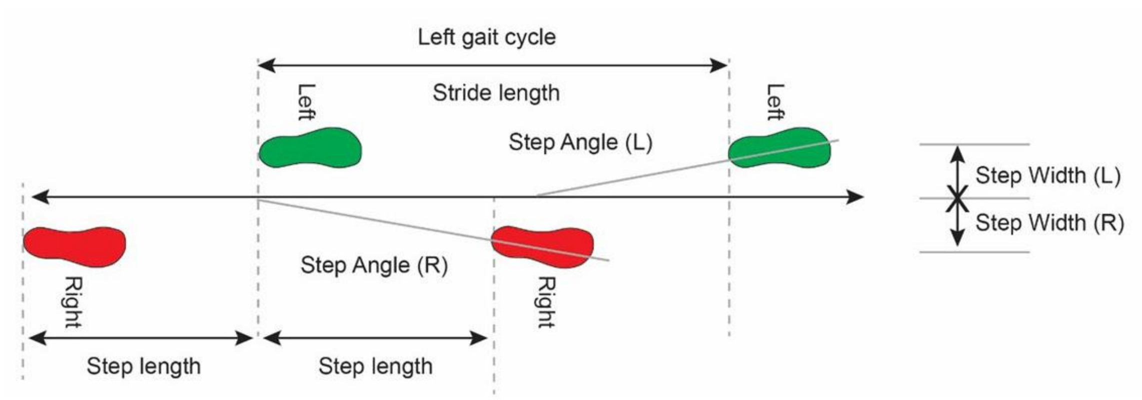


Figure 1: Gait parameters
(The Gait Cycle: Phases, Parameters to Evaluate & Technology | Gait Cycle Analysis | Tekscan)

This project aims to use advanced machine learning (ML) and artificial intelligence (AI) methodologies to identify and characterize subtle alterations in human movement patterns as potential biomarkers for long COVID. The primary research question of this project is: "Can machine learning and AI-driven analysis of human movement patterns reveal reliable markers for the diagnosis and monitoring of long COVID?". We believe that the lasting effects of COVID-19 on the body's neuromuscular and cardiovascular systems may cause small changes in movement, such as shifts in gait and balance. These changes are hard to notice in routine clinical exams but can be detected and analyzed using advanced machine learning techniques.

This project aligns with recent research suggesting that long COVID might involve different neuromuscular and autonomic problems. A study by Mancini and Horak (2010) showed that wearable sensors and gait analysis can be useful for monitoring chronic conditions, and that machine learning algorithms can detect specific movement patterns related to diseases. Also, AI has been successful in finding subtle patterns in complex data, like those from continuous movement tracking, which might be missed otherwise.

In this project, we will use both supervised and unsupervised machine learning methods to study large datasets of movement patterns from people with and without long COVID. The goal is to develop models that can tell these groups apart based on their movement data, identifying possible biomarkers. Successfully finding these markers could improve our understanding of long COVID and lead to new diagnostic tools and treatments, offering hope for better management of this condition.

3. Data used

The data used for this project are that of the COVID trials which were run at the Florida Atlantic university medical school. The data was collected using stepsense, a high performance gait and movement analysis system developed by PD-M in York. This sophisticated system enabled comprehensive gait analysis on a total of 170 patients (including test patients) including both individuals with long COVID and those without the condition. The long COVID patients are referred to as case (PP in ID) and regular people are referred to as control (QC in ID).

The subjects underwent three different types of trials. The first trial was a walking test where they walked towards and away from the camera at three distinct speeds (Slow, normal and fast). The second trial involved the Gans SOP balance assessment which included several components:
(i) Romberg tests (with eyes open and eyes closed)
(ii) Romberg tandem tests (with eyes open and eyes closed)

(iii) CTSIB (with eyes open and closed)
(iv) Fukuda tests (with eyes open)

The third trial was a similar Gans SOP balance assessment but included cognitive distractors where visual detection were used when the eyes were open and auditory distractions were applied when the eyes were closed.
The data used for this research consists of the dataset comprises of three dimensional (X-Y-Z) skeletal joint movements captured during patient gait activities. These movements are recorded to provide an accurate representation of each patient's gait which includes a range of specific variables and parameters critical to understanding their movement patterns. In each frame there is a timestamp and the joint position and rotation information for all the tracked joints for a total of 32 joints.
The data used also consists of a dataset which contains the session information and includes data such as the patient ID, session ID, the test performed, test description and the day the test was performed. Additionally, it records the conditions of each session and any important observations made during the tests. This detailed information is useful for tracking and analyzing how the tests change over time and what might cause any variations.

Stepsense's advanced technology makes it possible to track and analyze multiple skeletal joints with great accuracy during walking activities. The three-dimensional data is key to finding small differences and abnormalities in gait that may not be seen through traditional observation. By capturing movements along the X, Y, and Z axes, the dataset gives a complete picture of how each joint moves during the gait cycle.

Using three-dimensional joint movements is especially helpful for this type of analysis. It allows us to spot subtle gait problems that might not be visible to the naked eye. For example, slight changes in joint angles or movement patterns can be measured and analyzed to pinpoint specific issues.

The use of identifiers and patient IDs ensures that each piece of data is correctly linked to the right person while keeping their privacy safe. This is very important because the medical data is sensitive. Including details like date of birth, age, weight, and height helps normalize the data, making sure differences due to age, body size, and stature are considered in the analysis. This demographic information is essential for making accurate comparisons and drawing meaningful conclusions about how long COVID affects gait patterns.

4. Methods used

The data pipeline for the project is illustrated in the flow diagram below. This diagram outlines the sequential steps involved in processing the gait and patient data.

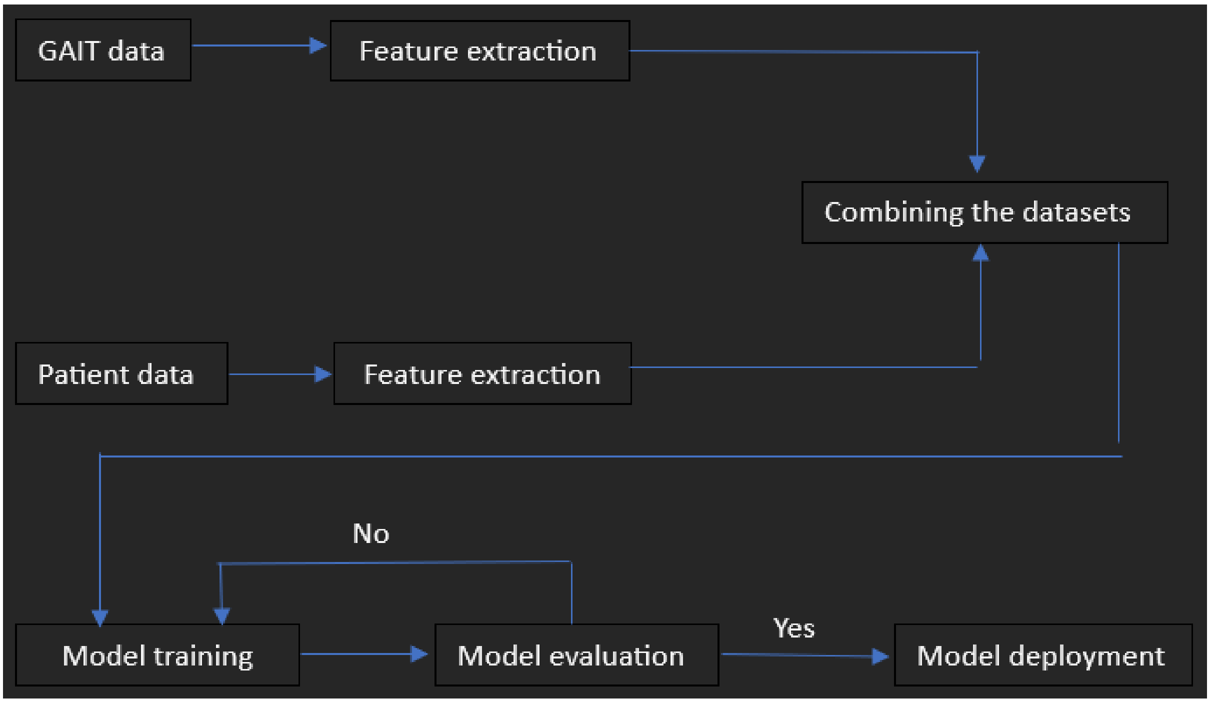


Figure 2: Flow diagram of the project

We shall explore each of the steps mentioned in Figure 2 in detail below.

(i) Data preprocessing

The first step in the project included data preprocessing in which the data underwent rigorous preprocessing to ensure its quality and consistency. For any missing values in the dataset, appropriate methods such as imputation was used to fill in the gaps or remove the records from the dataset if the missing data was too extensive. Along with this, the GAIT data was synchronised to ensure uniform alignment across all recordings ensuring that they could be compared accurately. Outliers (test subjects) which could have distorted our findings were identified and removed. Similarly, the demographic data was thoroughly cleaned making sure that any inconsistencies were eliminated to maintain uniformity throughout the dataset.

(ii) Feature extraction

The next step in the project involved the extraction of key features from the raw GAIT data which played a crucial role in both statistical analysis and machine learning model training. Given the complexity of human GAIT, this step was essential especially when considering conditions like long COVID.

(a) Velocity calculation

Velocity is a fundamental metric in gait analysis as it indicates the overall speed at which a person moves. In this project, we concentrated on two main velocity metrics: average velocity and maximum velocity

The average velocity was calculated by calculating the mean speed across all frames of the gait cycle. The average speed provides an overall assessment of the subject's walking speed. This metric was particularly useful in detecting potential decline in mobility or endurance which may be associated with long COVID (Kowal et al. 2023).

Maximum velocity on the other hand captured the highest speed reached during any part of the gait cycle. The maximum velocity is captured by sorting the absolute velocity values in descending order and then averaging the top 10 values. This approach highlights the peak performance by giving a reliable estimate of the highest speed reached during the session. When considering long COVID, a decrease in maximum velocity might suggest lingering impacts on neuromuscular function or cardiovascular health (Keklicek et al. 2022).

(b) Step length and step width calculation

In addition to velocity, step length and step width are key factors that offer a deeper understanding of gait. These metrics were extracted from the three-dimensional coordinates of the left and right foot, offering insights into the spatial and temporal characteristics of walking.

Step length

The step length was calculated by computing the Euclidean distance between the left and right foot in 3D space. This provides an indication of the stride efficiency and can highlight asymmetries in gait. In context to long COVID, variations in step length could suggest impairments in balance or coordination.

Step width

The step width was calculated by computing the lateral distance that is the absolute difference between the x-coordinates of the left and right foot in three dimensional space. Step width is associated with stability and balance. An increased step width may indicate a person's attempt to compensate for instability or to prevent falls. In people with long COVID, variations in step width might suggest changes in motor control or issues with maintaining balance.

(c) Demographics calculation

BMI (Body mass index) was derived from the demographic dataset as it plays a crucial role in understanding the overall health and mobility of patients. In context to gait analysis, BMI can influence various aspects of movement such as balance, step length and stability.

(iii) Combining the dataset

After the feature extraction process, the next step was to integrate the extracted GAIT parameters with the patient demographic data to create a comprehensive dataset. This step was critical because it allowed for the analysis of gait characteristics within the context of individual patient attributes such as age, height, BMI and gender.

The integration process involved merging multiple dataframes: one containing the gait parameters (average velocity, maximum velocity, step length and step width) and another containing the demographic information. The merging was done based on unique patient identifier to ensure that each subject's gait data was correctly associated with their demographic information.

(iv) Statistical analysis

Statistical analysis played a crucial role in this project acting as an essential step before moving on to training machine learning models. The goal was to confirm that the study groups were similar in their demographic traits, minimizing any factors that could distort the results.

To compare demographic values such as age, height, BMI and gender between the COVID-19 and control group subjects, t-tests were conducted. The t-test is a statistical method used to determine if there is a significant difference in the means of two groups, making it particularly valuable in medical research for assessing difference in characteristics across the subjects (Sullivan and Feinn 2012)

- **Age:** A t-test was conducted to compare the average ages of the COVID-19 and control group. Given that age significantly impacts COVID-19 severity and progression, this analysis ensured that both groups were similar in terms of age distribution, reducing the risk of age related bias (Onder, Rezza, and Brusaferro 2020).

- **Height:** Height was also examined to eliminate potential biases that might arise from differences in stature as physical height can influence gait characteristics. By conducting a t-test, we ensured that any gait differences observed were not due to height variations.

- **BMI:** BMI was a critical factor in this study because it affects balance, stability and movement efficiency. The t-test compared the BMI of the two groups to confirm that any differences in gait patterns were not due to body composition discrepancies (Piirtola and Era 2006).

- **Gender:** Gender was analyzed as well given that male and females both exhibit different gait patterns. A t-test confirmed that the male to female ratios in both groups were comparable, helping to control for gender related gait differences (Whittle 2014).

(v) Machine learning model training

Once the dataset was prepared and validated through statistical analysis, the next step involved the training of machine learning models. These models were designed to classify individuals as COVID-19 patients or control subjects based on their gait features.

(a) Label assignment

Each participant was assigned a binary label to facilitate classification. **COVID-19** patients were labelled as **'0'** and **control** subjects were labelled as **'1'**. This binary labelling system was essential for supervised machine learning where the model learns to predict the class label based on the input features. The labelling process was carefully cross checked to avoid any misclassification which could undermine the model's performance.

(b) Model selection

For this project, we used a variety of machine learning models, each bringing its own strengths to handle the complexity of the dataset. The models chosen included:

- **Logistic regression:**
  This model is a statistical method commonly used for binary classification problems, making it a good fit for distinguishing between COVID-19 and control groups. This model works by estimating the probability that a subject falls into one category or the other based on their input features. It's main advantage lies in simplicity and ease of interpretation as it allows us to clearly understand how each feature influences the likelihood of a particular outcome.

- **Random forest:**
  This model is a powerful ensemble learning technique that builds multiple decision trees during the training process. By combining the predictors from these trees, the model can achieve higher stability and greater accuracy. One of the key strengths of random forest is its ability to handle complex datasets with many features. It does this by randomly selecting subsets of the data for each tree, which helps to

minimize overfitting and improve the model's robustness. This method is especially useful when identifying the most important features that contribute to the prediction.

- **Gradient boosting:**
  It is another ensemble technique which builds models in a sequential manner. With each new model, aiming to correct the mistakes made by the one before it. Gradient boosting is particularly effective for its complex datasets as it can uncover subtle patterns that simpler models might overlook. Gradient boosting is well regarded for its high accuracy and its capacity to handle non linear relationships between features making it a strong choice for tackling challenging and predictive tasks.

- **XGboost**
  It is an advanced take on Gradient Boosting that really shines in terms of speed and efficiency. It's become a favorite in machine learning competitions because of its impressive performance, especially when working with large, messy datasets. This algorithm excels at tackling complex classification problems, making it an excellent choice when dealing with data that might have irrelevant features or noise.

- **Support vector machine**
  Support Vector Machine (SVM) is a robust tool for binary classification that works by finding the best boundary (or hyperplane) to separate different classes with the widest margin. It's especially effective in high-dimensional spaces where the data might not be easily separated by a straight line. SVM shines in scenarios where classes are complex and not linearly separable, making it a great choice for this project, given the multi-dimensional nature of gait parameters (Cortes and Vapnik 1995).

(c) Model training and evaluation

The models were trained using the combined dataset with the gait parameters serving as the input features. The dataset was split into training and testing subsets with 80% of the data used for training and 20% of the data used for testing. This split ensured that the models were trained on a sufficient amount of data.

Each model was trained using its default settings initially. Following this, fine tuning of hyperparameters was done to enhance the model's accuracy and effectiveness. During training, the models were provided with the input features (average and maximum velocity, step length and step width) along with corresponding labels (0 for COVID-19 patients or 1 for control subjects). This process allowed the models to learn how these features relate to the labels, enabling them to detect patterns that distinguish between COVID-19 patients and control subjects.

The performance of each model was evaluated using several metrics including:

- **Accuracy:** The percentage of correctly classified instances (both COVID-19 patients or control subjects). While accuracy gives an overall sense of the model's performance, it can be deceptive when the classes are imbalanced as the model might perform well on the majority class but poorly on the minority class (Sokolova and Lapalme 2009).

- **Precision:** The percentage of correctly identified positive cases out of all the cases the model predicted as positive. This metric is really important in scenarios when the cost of false positives is high (Sokolova and Lapalme 2009).

- **Recall:** The proportion of true positive predictions out of all actual positives in the dataset. Recall is crucial Fmedical diagnosis where failing to identify a true positive could have serious consequences.

- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of the model's performance

- **Area under ROC curve (AUC-ROC):** AUC-ROC is a performance measurement for classification problems at various threshold settings. The ROC is a probability curve and the AUC represents the degree or measure of separability achieved by the model. An AUC close to 1 indicates excellent model performance.

(d) Voting classifier

After training each model individually, a voting classifier was implemented. This ensemble method combines the predictions from multiple models, making the final decision based on the majority vote. By pooling the strengths of various models, the voting classifier often enhances overall performance (Dietterich 2000). This approach is particularly effective when different models capture different aspects of the data, leading to a more balanced and accurate prediction. In this project, the voting classifier was chosen to capitalize on the diverse strengths of the individual models.

5. Results

The general characteristics of the groups were are given in table 1

Characteristics	COVID group n=22	Control group n=138	P-value
Age, years	49.72	45.74	0.345
Height, m	1.53	1.62	0.00592
BMI, kg/m²	26.385	25.731	0.835
Male/Female	9/14	50/79	

Table 1: General characteristics of the study subjects

The following were observed after the analysis of demographics of the data:

- **Age**: The average age of the COVID group is 49.72 years, while the control group has an average age of 45.74 years. The p-value for age is 0.345, indicating that there is no statistically significant difference in age between the two groups.
- **Height**: The average height of participants in the COVID group is 1.53 meters, compared to 1.62 meters in the control group. The p-value for height is 0.00592, which is statistically significant. This suggests that height is a significant distinguishing factor between the two groups, with the COVID group being shorter on average.
- **BMI**: The average Body Mass Index (BMI) for the COVID group is 26.385 kg/m², while the control group has an average BMI of 25.731 kg/m². The p-value for BMI is 0.835, indicating that there is no significant difference in BMI between the two groups.
- **Gender Distribution**: The gender breakdown shows 9 males and 14 females in the COVID group, and 50 males and 79 females in the control group.

The p-value for height (0.00592) is significant, suggesting that height is a meaningful factor in differentiating between the COVID and control groups. This finding highlights the importance of considering height in further analyses related to long COVID.

	COVID group n=22	Control group n=138
Average velocity, mm/s	791.021955	790.262202
Maximum velocity, mm/s	1266.423302	1191.995895
Step length, mm	213.321045	214.606298
Step width, mm	188.384514	194.692208

Table 2: Comparison of gait parameters

The gait analysis parameters which were evaluated for case (COVID) and control subjects are presented in table 2.

For average velocity, the COVID group exhibited an average of 791.02, slightly higher than the control group's average of 790.26. This minimal difference suggests that the overall walking speed of individuals in both groups is fairly comparable. However, the higher mean velocity in the COVID group could indicate a subtle alteration in gait dynamics.

The maximum velocity, which measures the highest speed attained during the gait cycle was notable higher in the COVID group, with an average of 1266.42 compared to 1191.91 in the control group. This significant difference might reflect an increased variability in gait speed among COVID affected individuals, potentially pointing to disruptions in neuromuscular control. The increased maximum velocity in the COVID group might also reflect moments of faster movement, possibly triggered by instability or uneven pacing, issues less common in the control group.

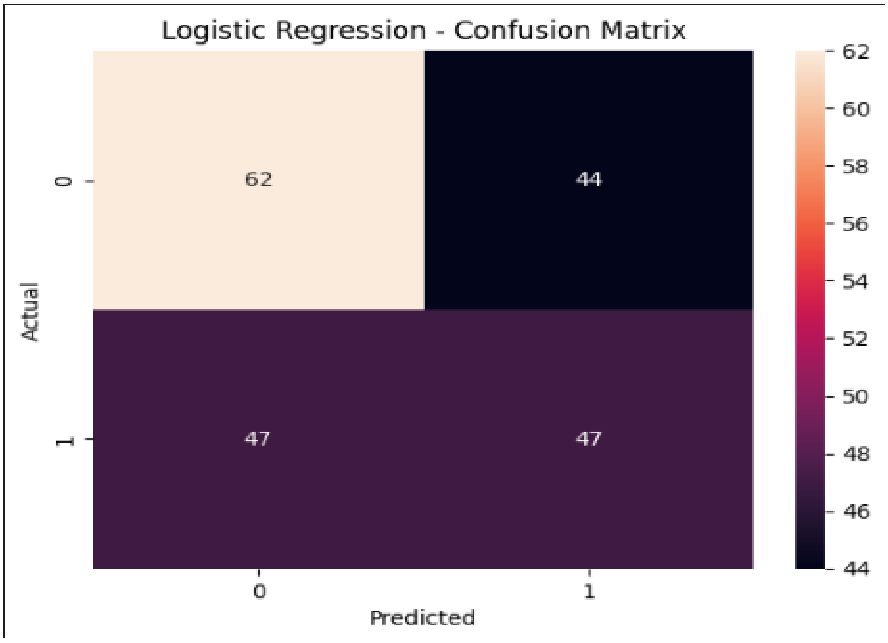
In terms of step length, the COVID-19 group showed a slightly shorter average of 213.32 compared to 214.61 in the control group. While this difference is minimal, it could point to subtle alterations in stride patterns among those affected by COVID-19. A reduced step length may be linked to symptoms such as fatigue, muscle weakness, or balance issues, which are commonly reported by individuals experiencing long COVID.

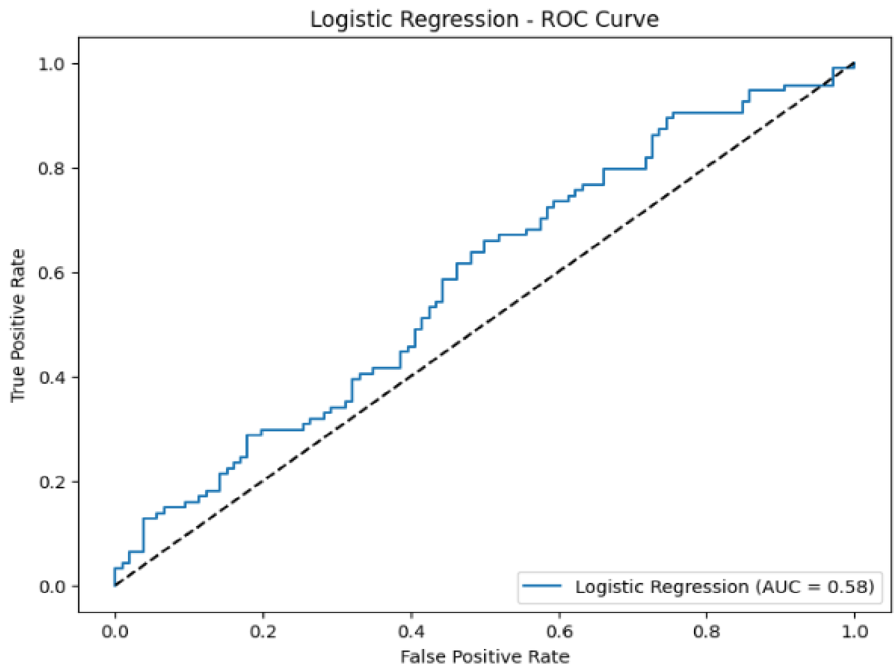
Lastly, the average step width was found to be narrower in the COVID-19 group, with a mean of 188.38 compared to 194.69 in the control group. This narrower step width may suggest balance issues, as individuals might reduce lateral movement to stabilize their gait. This adjustment could be a compensatory strategy to maintain balance and prevent falls, especially among those recovering from COVID-19.

The results of all 6 models used in the study are mentioned below:

1. Logistic regression

	Accuracy	Precision	Recall	F1-score
Case (0)	0.55	0.57	0.58	0.58
Control (1)		0.52	0.50	0.51



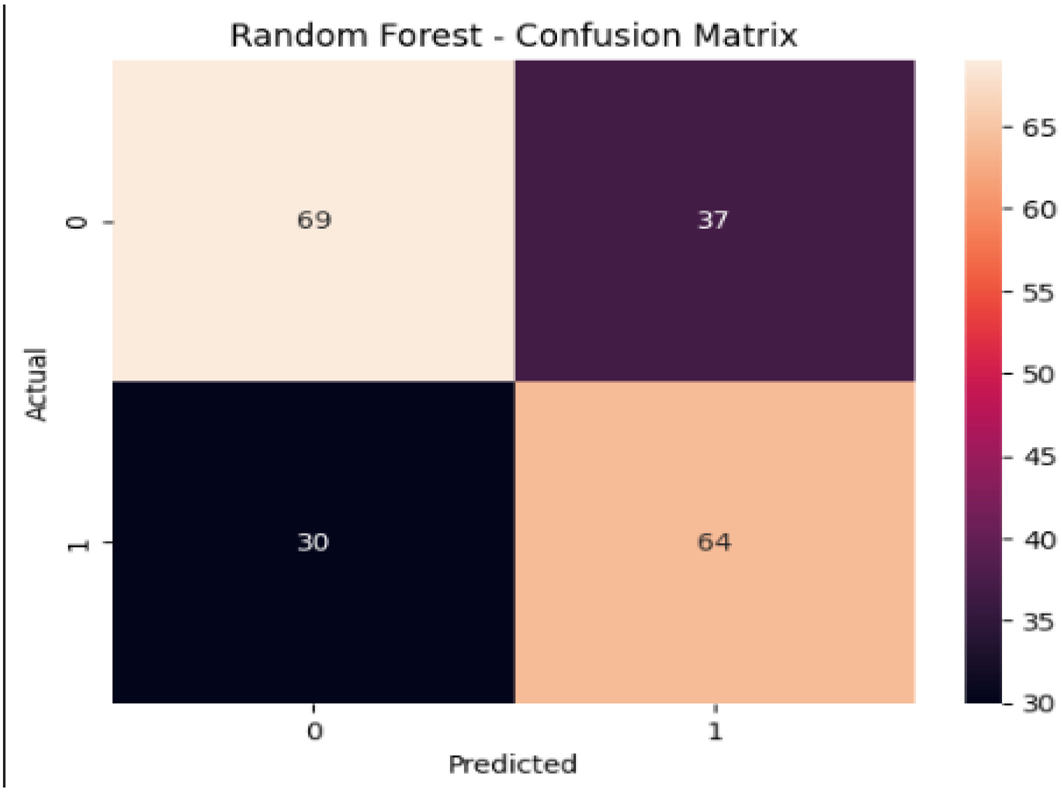


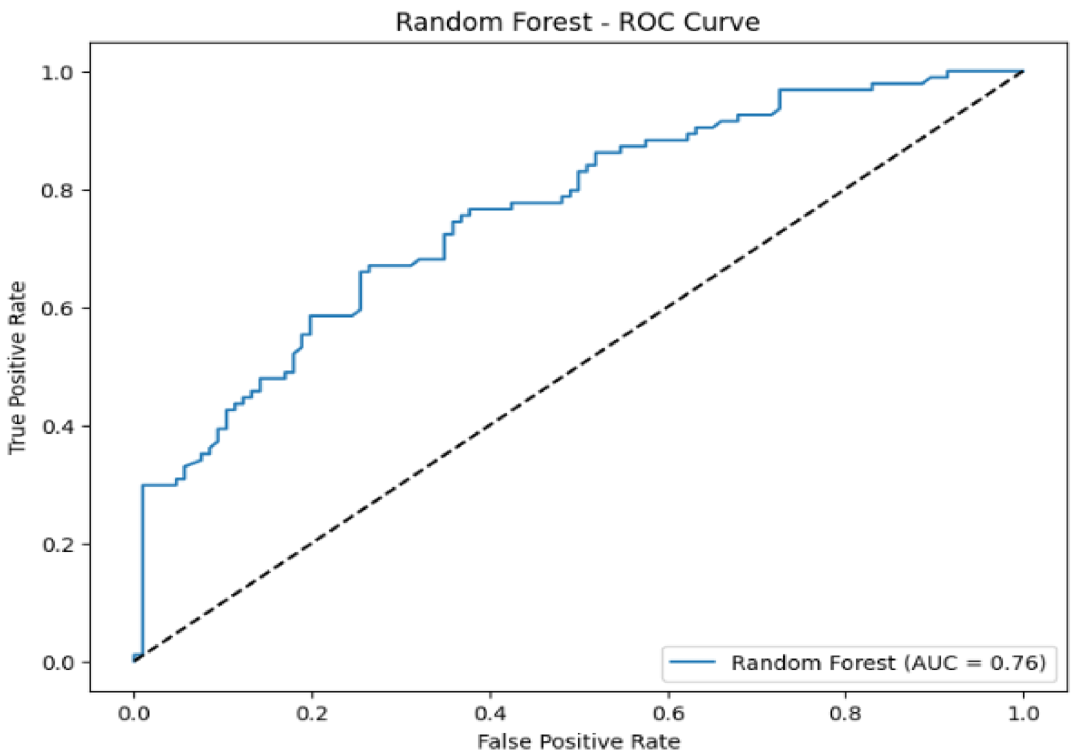
The logistic regression model which was applied to classify control and case subjects in this study achieved an accuracy of 55% indicating that it correctly classified 55% of the total cases. However, both precision and recall for the individual classes suggest moderate performance. The precision for class 0 (case) was 0.57 indicating that 57% of the predicted cases were actually cases. Similarly, the precision for class 1 (control) was 0.52, meaning that 52% of the predicted cases were true cases. The recall values for both classes were 0.58 and 0.50, respectively, indicating that the model identified 58% of the actual cases and 50% of the actual controls.

The confusion matrix further illustrates this performance, with 62 true positives for class 0 and 47 for class 1, but also a considerable number of misclassifications (44 false negatives for class 0 and 47 false positives for class 1). The ROC curve, with an AUC of 0.58, reflects the model's limited ability to distinguish between classes, as it is only slightly better than random guessing. Overall, the model demonstrates modest classification ability, with room for improvement.

2. Random forest

	Accuracy	Precision	Recall	F1-score
Case (0)	0.67	0.70	0.65	0.67
Control (1)		0.63	0.63	0.66



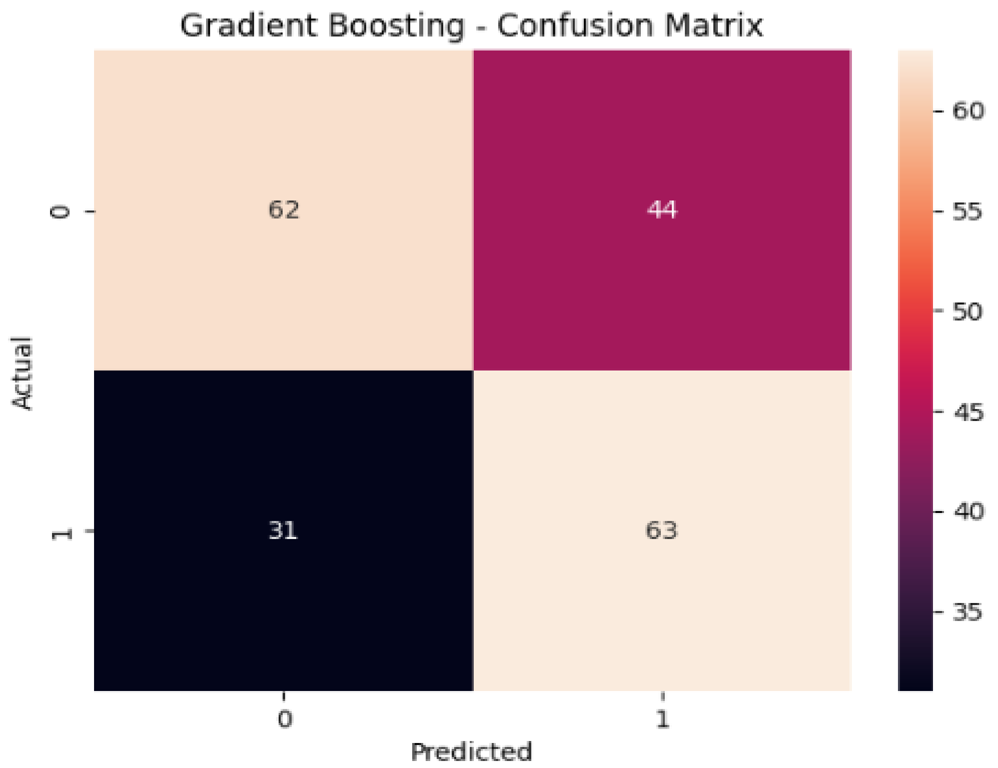


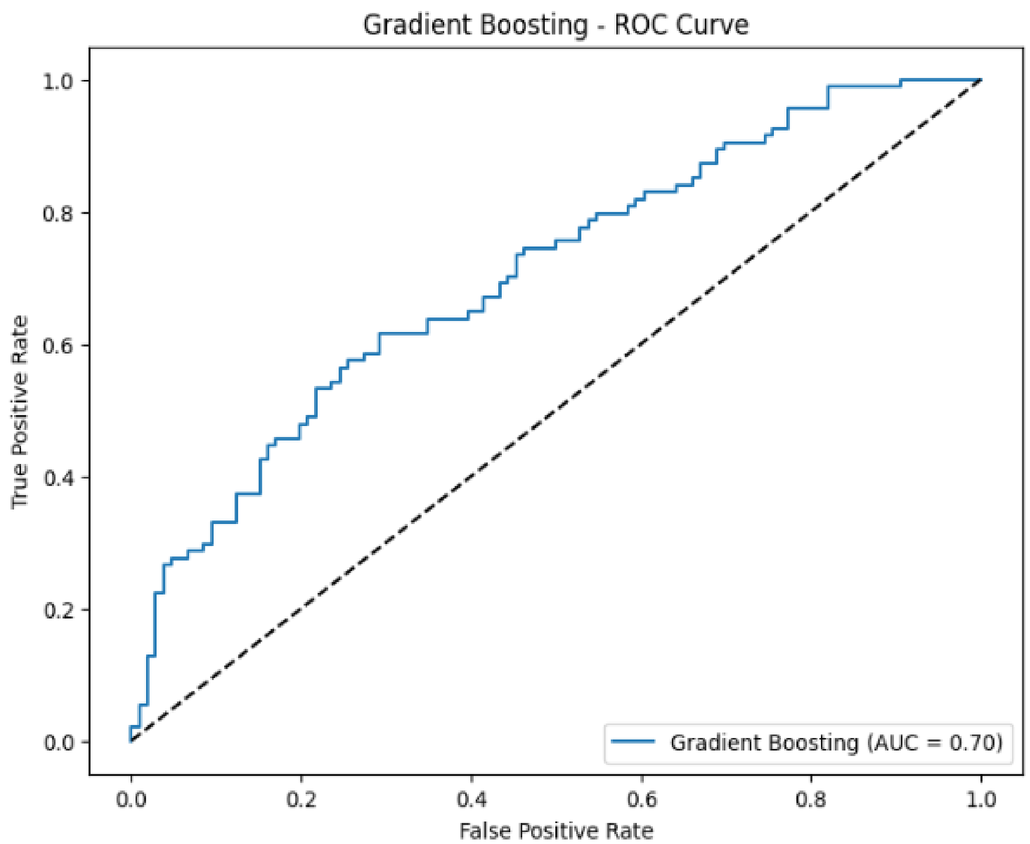
The random forest model model used to classify yielded an accuracy of 67%. The model demonstrated balanced performance across both classes, with a precision of 0.70 for controls and 0.63 for cases. This indicates that 70% of the predicted controls and 63% of the predicted cases were correctly identified. The recall values were 0.65 for controls and 0.68 for cases, meaning the model successfully identified 65% of actual controls and 68% of actual cases.

The confusion matrix shows that 69 true positives were identified for class 0, while 64 were correctly classified for class 1. However, there were also notable misclassifications, with 37 false negatives for class 0 and 30 false positives for class 1. The ROC curve, with an AUC of 0.76, indicates that the random forest model has a good ability to distinguish between control and case subjects, substantially better than random guessing.

3. Gradient boosting

	Accuracy	Precision	Recall	F1-score
Case (0)	0.62	0.67	0.58	0.62
Control (1)		0.59	0.67	0.63


Gradient Boosting - Confusion Matrix
62
44
31
63
Actual
0
1
Predicted
0
1
60
55
50
45
40
35

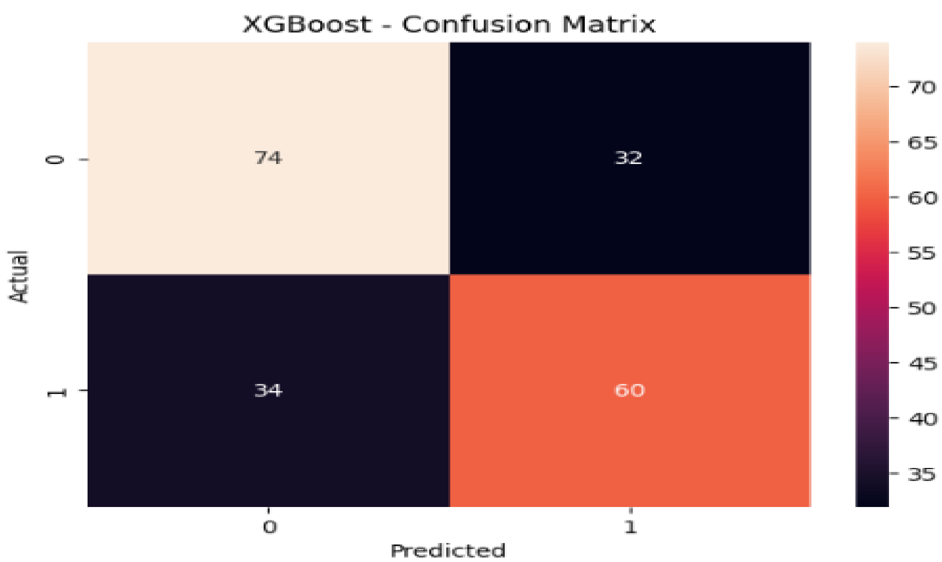

Gradient Boosting - ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Gradient Boosting (AUC = 0.70)

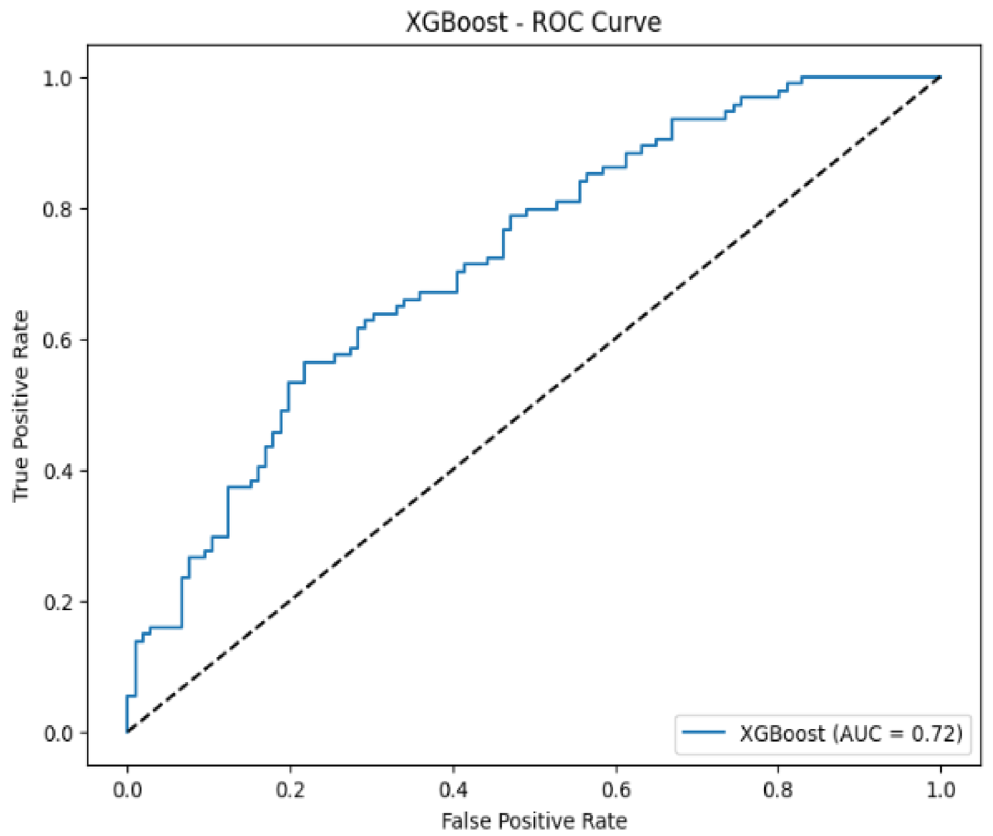
The gradient boosting model used to predict yielded a moderate performance. The overall accuracy of the model is 62%, indicating that the model's predictions are moderately reliable. The classification report shows that the precision for the case class (0) is 0.67, while for the control class (1) it is 0.59. Recall rates are 0.58 and 0.67 for case and control classes, respectively, with F1-scores close to the accuracy, suggesting a balanced performance across precision and recall.

The confusion matrix reflects that the model correctly identified 62 case subjects but misclassified 44 as control. On the other hand, it correctly classified 63 control while misidentifying 31 as case. The ROC curve for the Gradient Boosting model yielded an AUC (Area Under the Curve) of 0.70. This indicates that the model has a fair ability to discriminate between control (1) and case (0) subjects. An AUC of 0.70 suggests that there is a 70% chance that the model will correctly differentiate between a randomly chosen control subject and a randomly chosen case subject.

4. XGBoost

	Accuracy	Precision	Recall	F1-score
Case (0)	0.67	0.69	0.70	0.69
Control(1)		0.65	0.64	0.65



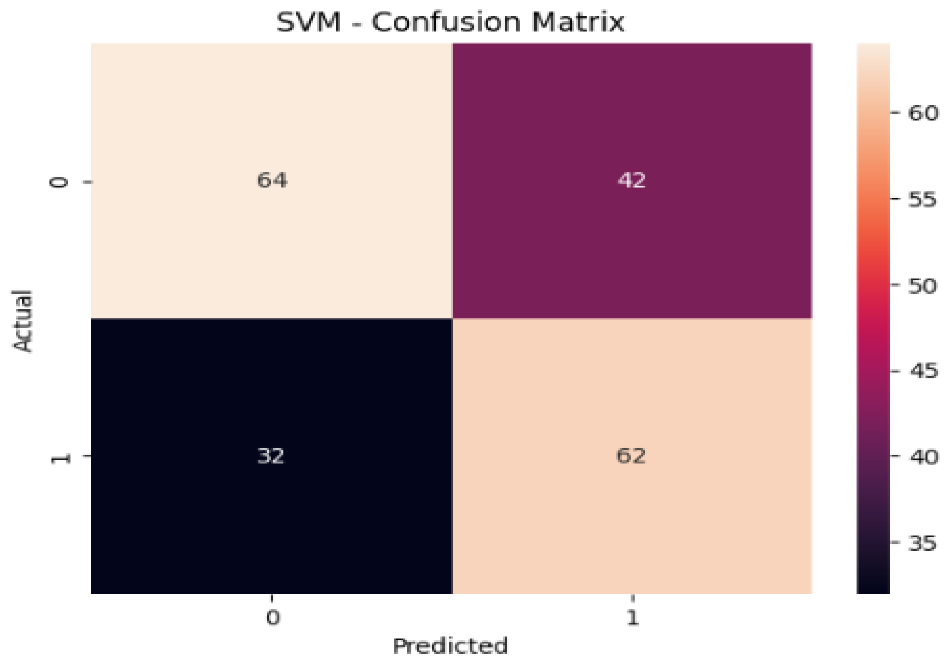


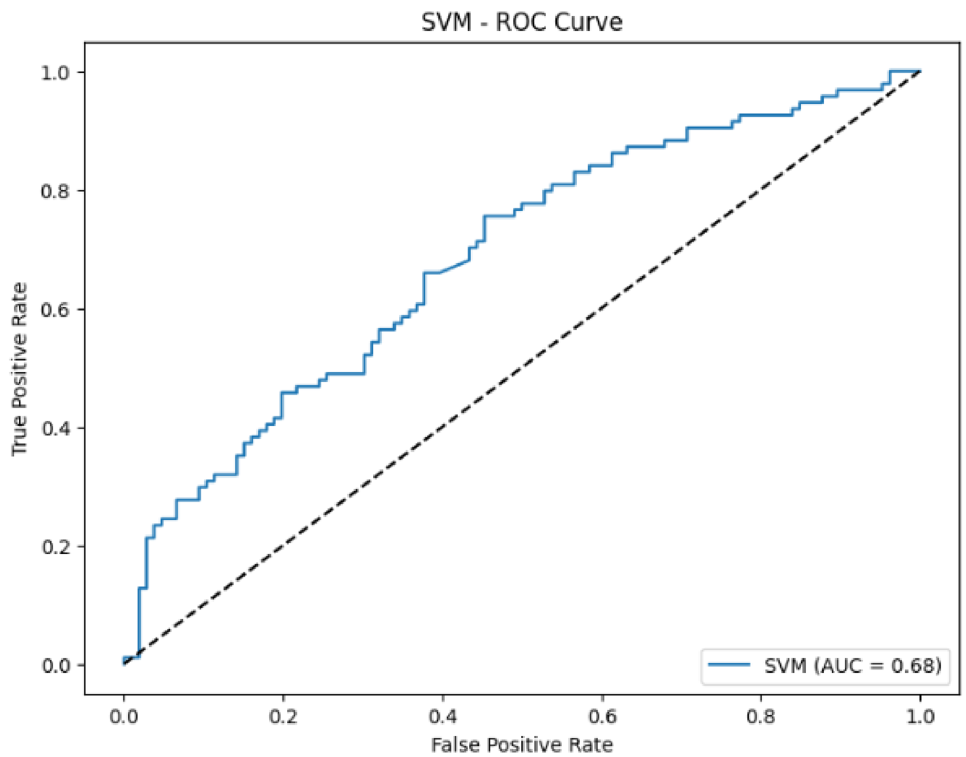
The application of the XGBoost model to predict control (1) and case (0) subjects resulted in an accuracy of 67%, indicating a moderate level of performance. The classification report reveals a precision of 0.69 for the control class and 0.65 for the case class, with corresponding recall values of 0.70 and 0.64. The F1-scores for both classes are close to the overall accuracy, which suggests balanced performance across precision and recall.

The confusion matrix shows that the model correctly classified 74 case subjects but misclassified 32 as controls. Similarly, it correctly identified 60 controls while incorrectly predicting 34 as cases. Additionally, the ROC curve for the XGBoost model yielded an AUC of 0.72, indicating that the model has a fairly good ability to distinguish between control and case subjects. An AUC of 0.72 suggests a 72% probability that the model will correctly differentiate between a randomly selected control and case subject.

5. Support vector machine (SVM)

	Accuracy	Precision	Recall	F1-score
Case (0)	0.63	0.67	0.60	0.63
Control (1)		0.60	0.66	0.63



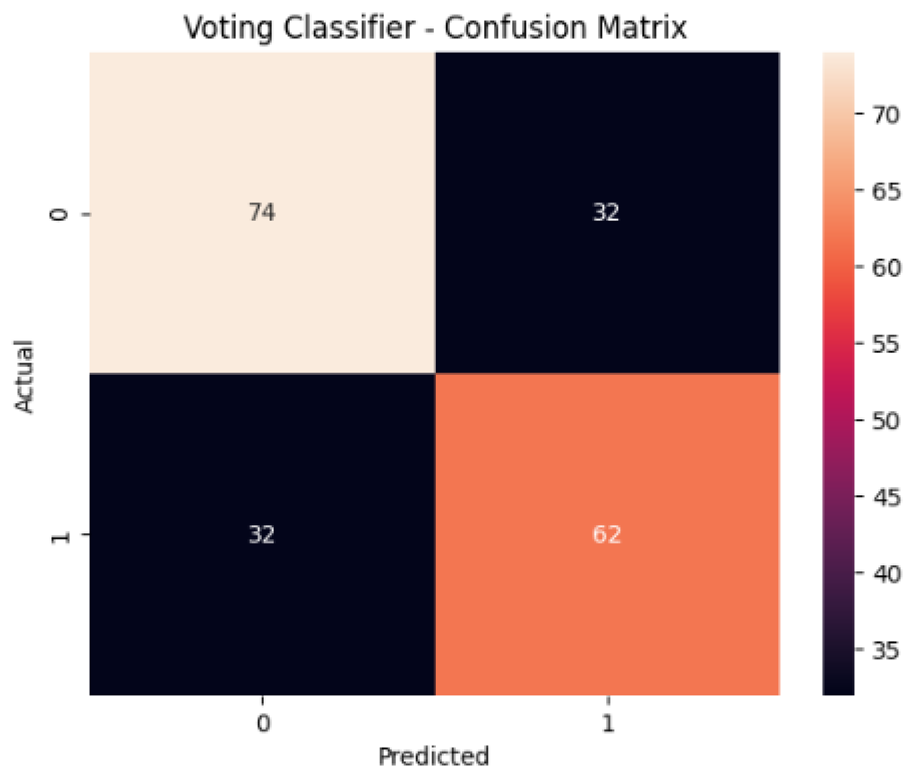


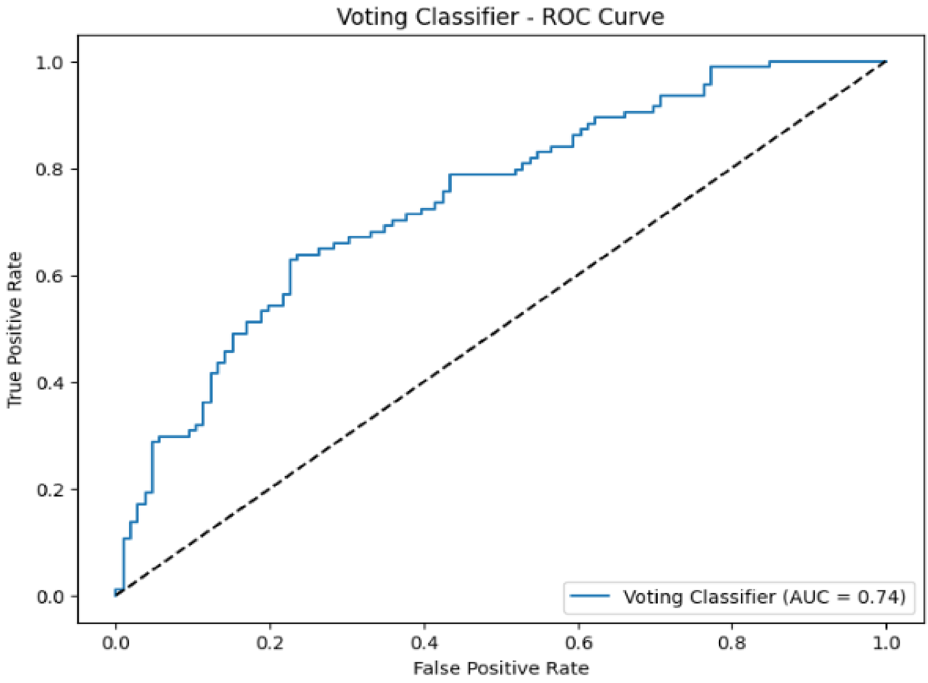
The Support Vector Machine (SVM) model was applied to predict control (1) and case (0) subjects, resulting in an overall accuracy of 63%. The confusion matrix indicates that the model correctly classified 64 case subjects, while 42 were incorrectly predicted as controls. On the other hand, the model accurately identified 62 controls but misclassified 32 as cases. These results are reflected in the classification report, where the precision for the control class (0) is 0.67 and for the case class (1) is 0.60. Both classes have a recall value of around 0.63, resulting in F1-scores that align with the model's overall accuracy.

Furthermore, the ROC curve for the SVM model yielded an AUC of 0.68, indicating a modest ability to differentiate between control and case subjects. An AUC of 0.68 suggests that the model has a 68% probability of correctly distinguishing between a randomly chosen control subject and a randomly chosen case subject.

6. Voting classifier

	Accuracy	**Precision**	**Recall**	**F1-score**
Case (0)	0.68	0.70	0.70	0.70
Control (1)		0.66	0.66	0.66





The application of the Voting Classifier model to predict control (1) and case (0) subjects achieved an accuracy of 68%, indicating a moderate level of predictive performance. The confusion matrix reveals that the model correctly identified 74 case subjects, while 32 were misclassified as controls. Conversely, the model correctly predicted 62 control subjects but incorrectly classified 32 as cases. The classification report shows a precision of 0.70 for the control class and 0.66 for the case class, with both classes having recall values that align closely with their precision. The F1-scores for both classes are consistent with the overall accuracy, suggesting balanced performance in terms of precision and recall.

The ROC curve for the Voting Classifier model yielded an AUC of 0.74, indicating a fair ability to differentiate between control and case subjects. An AUC of 0.74 implies that the model has a 74% probability of correctly distinguishing between a randomly selected control subject and a randomly selected case subject.

7. Discussion

The present study aimed to investigate the differences in gait parameters between individuals affected by long COVID and those in a control group and predict whether a subject is suffering from long COVID or not using various machine learning models. The primary findings reveal subtle but noteworthy differences in gait characteristics, as well as varying degrees of accuracy in the classification models used to predict COVID and control cases based on these gait features.

(a) Differences in gait parameters

The analysis of gait parameters revealed subtle yet meaningful differences in walking patterns between individuals who had recovered from COVID-19 and those in the control group. Interestingly, while the average walking speed was almost the same for both groups, the COVID group showed a noticeably higher maximum velocity. This variation in peak speed may indicate that people recovering from COVID-19 experience greater fluctuations in their gait, possibly due to lingering neuromuscular challenges or compensatory mechanisms developed in response to persistent symptoms like fatigue or muscle weakness (Moreno-Pérez et al. 2021). These findings are consistent with existing research that highlights the lasting impact of COVID-19 on physical and neurological function (Keklicek et al. 2022; Kowal et al. 2023).

Research suggests that individuals recovering from COVID-19 may exhibit subtle alterations in gait, such as shorter step length and narrower step width, which could indicate underlying balance issues or adjustments in walking patterns. These changes might be a compensatory mechanism to maintain stability, especially in those experiencing post-COVID-19 symptoms like fatigue or muscle weakness. Studies have identified that these alterations in gait dynamics are not uncommon and could reflect the lingering effects of the virus on the neuromuscular system, potentially impacting balance and overall mobility (Franco et al. 2024).

These findings underscore the importance of gait analysis as a non-invasive method to monitor and understand the long-term impacts of COVID-19 on physical function, particularly in individuals who report ongoing symptoms even after the acute phase of the illness has passed.

(b) Model performance

Six machine learning models were used to classify the case and control subjects based on the extracted gait parameters. The performance of these models varied, with the Voting Classifier model demonstrating the highest accuracy at 68%, followed by the Random Forest and XGBoost models, both achieving 67% accuracy. The Support Vector Machine (SVM) and Gradient Boosting models showed moderate performance with accuracies of 63% and 62%, respectively, while the Logistic Regression model had the lowest accuracy at 55%.

The Voting Classifier and Random Forest models exhibited relatively balanced performance across both classes, as reflected in their precision, recall, and F1-scores. The higher AUC values (0.74 and 0.76, respectively) for these models indicate a fair ability to distinguish between COVID and control subjects. These results suggest that ensemble methods, which combine the predictions of multiple models, might be particularly effective in handling the variability present in the gait data of COVID-affected individuals.

On the other hand, the Logistic Regression model's modest performance, with an accuracy of 55%, suggests that linear models might struggle to capture the complex patterns in gait data that differentiate COVID cases from controls. The ROC curve for this model, with an AUC of 0.58, further supports its limited predictive power.

(c) Implications for long covid diagnosis

The findings of this study have significant implications for the diagnosis and management of long COVID. The observed differences in gait parameters between COVID and control groups suggest that gait analysis could serve as a useful tool for identifying individuals who may be experiencing prolonged effects of COVID-19. While the changes in gait characteristics are subtle, they could be indicative of underlying issues related to balance, neuromuscular control, or fatigue, which are common in long COVID patients .

This study provides a foundation for future research into using gait analysis as a potential biomarker for long COVID. By offering a non-invasive and accessible method for early detection, gait analysis could become an important tool for early detection and ongoing monitoring of the condition.

(d) Limitations and future plans

The sample size, particularly for the COVID group, was relatively small, which may limit the generalizability of the findings. Additionally, the study focused on a limited set of gait parameters; incorporating additional biomechanical or physiological data could provide a more comprehensive understanding of the impact of COVID-19 on gait.

Future research should aim to validate these findings in larger and more diverse populations. Moreover, exploring the use of more advanced machine learning techniques, such as deep learning, could potentially improve the predictive accuracy and reveal deeper and more detailed patterns in gait data.

8. References

- Abdullahi, A, S. A Candan, M. A Abba, A. H Bello, M. A Alshehri, E. Afamefuna Victor, N. A Umar, and B. Kundakci. 2020. “Neurological and Musculoskeletal Features of COVID-19: A Systematic Review and

Meta-Analysis." *Frontiers in Neurology* 11. doi: 10.3389/fneur.2020.00687.

- Balcom, Erin F., Avindra Nath, and Christopher Power. 2021. "Acute and Chronic Neurological Disorders in COVID-19: Potential Mechanisms of Disease." *Brain* 144(12):3576–88. doi: 10.1093/brain/awab302.
- Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20(3):273–97. doi: 10.1007/BF00994018.
- Dietterich, Thomas G. 2000. "Ensemble Methods in Machine Learning." Pp. 1–15 in *Multiple Classifier Systems*. Berlin, Heidelberg: Springer.
- Franco, Juan Victor Ariel, Luis Ignacio Garegnani, Maria-Inti Metzendorf, Katharina Heldt, Rebekka Mumm, and Christa Scheidt-Nave. 2024. "Post-Covid-19 Conditions in Adults: Systematic Review and Meta-Analysis of Health Outcomes in Controlled Studies." *BMJ Medicine* 3(1). doi: 10.1136/bmjmed-2023-000723.
- Godaert, Lidvine, Emeline Proye, David Demoustier-Tampere, Pecory Souleymane Coulibaly, Fanny Hequet, and Moustapha Dramé. 2020. "Clinical Characteristics of Older Patients: The Experience of a Geriatric Short-Stay Unit Dedicated to Patients with COVID-19 in France." *Journal of Infection* 81(1):e93–94. doi: 10.1016/j.jinf.2020.04.009.
- Hausdorff, J. M, and N B. Alexander. 2005. *Gait Disorders Evaluation and Management*.
- Helms, Julie, Stéphane Kremer, Hamid Merdji, Raphaël Clere-Jehl, Malika Schenck, Christine Kummerlen, Olivier Collange, Clotilde Boulay, Samira Fafi-Kremer, Mickaël Ohana, Mathieu Anheim, and Ferhat Meziani. 2020. "Neurologic Features in Severe SARS-CoV-2

Infection." *New England Journal of Medicine* 382(23):2268–70. doi: 10.1056/NEJMc2008597.

- Jamari, J., Muhammad Imam Ammarullah, Gatot Santoso, S. Sugiharto, Toto Supriyono, Muki Satya Permana, Tri Indah Winarni, and Emile van der Heide. 2022. "Adopted Walking Condition for Computational Simulation Approach on Bearing of Hip Joint Prosthesis: Review over the Past 30 Years." *Heliyon* 8(12):e12050. doi: 10.1016/j.heliyon.2022.e12050.
- Karaarslan, Fatih, Fulya Demircioğlu Güneri, and Sinan Kardeş. 2022. "Long COVID: Rheumatologic/Musculoskeletal Symptoms in Hospitalized COVID-19 Survivors at 3 and 6 Months." *Clinical Rheumatology* 41(1):289–96. doi: 10.1007/s10067-021-05942-x.
- Keklicek, Hilal, Halit Selçuk, İlke Kurt, Sezer Ulukaya, and Gülnur Öztürk. 2022. "Individuals with a COVID-19 History Exhibit Asymmetric Gait Patterns despite Full Recovery." *Journal of Biomechanics* 137:111098. doi: 10.1016/j.jbiomech.2022.111098.
- Kowal, Mateusz, Ewa Morgiel, Sławomir Winiarski, Ewa Gieysztor, Marta Madej, Agata Sebastian, Marcin Madziarski, Nicole Wedel, Krzysztof Proc, Katarzyna Madziarska, Piotr Wiland, and Małgorzata Paprocka-Borowicz. 2023. "Effect of COVID-19 on Musculoskeletal Performance in Gait and the Timed-Up and Go Test." *Journal of Clinical Medicine* 12(13):4184. doi: 10.3390/jcm12134184.
- Lechner-Scott, J., M. Levy, C. Hawkes, A. Yeh, and G. Giovannoni. 2021. "Long COVID or Post COVID-19 Syndrome." *Multiple Sclerosis and Related Disorders* 55:103268. doi: 10.1016/j.msard.2021.103268.
- Mancini, Martina, and Fay B. Horak. 2010. "The Relevance of Clinical Balance Assessment Tools to Differentiate Balance Deficits."

European Journal of Physical and Rehabilitation Medicine 46(2):239–48.

- Moreno-Pérez, Oscar, Esperanza Merino, Jose-Manuel Leon-Ramirez, Mariano Andres, Jose Manuel Ramos, Juan Arenas-Jiménez, Santos Asensio, Rosa Sanchez, Paloma Ruiz-Torregrosa, Irene Galan, Alexander Scholz, Antonio Amo, Pilar González-delaAleja, Vicente Boix, Joan Gil, and COVID19-ALC research group. 2021. "Post-Acute COVID-19 Syndrome. Incidence and Risk Factors: A Mediterranean Cohort Study." *The Journal of Infection* 82(3):378–83. doi: 10.1016/j.jinf.2021.01.004.
- Onder, Graziano, Giovanni Rezza, and Silvio Brusaferro. 2020. "Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy." *JAMA* 323(18):1775–76. doi: 10.1001/jama.2020.4683.
- Phillips, Steven, and Michelle A. Williams. 2021. "Confronting Our Next National Health Disaster — Long-Haul Covid." *New England Journal of Medicine* 385(7):577–79. doi: 10.1056/NEJMp2109285.
- Piirtola, Maarit, and Pertti Era. 2006. "Force Platform Measurements as Predictors of Falls among Older People - a Review." *Gerontology* 52(1):1–16. doi: 10.1159/000089820.
- Pistoia, Francesca, Raffaele Ornello, Patrizia Sucapane, Carmine Marini, and Simona Sacco. 2021. "Symptoms of Gait and Coordination Impairment in a Patient with COVID-19 Interstitial Pneumonia." *Neurological Sciences* 42(8):3083–86. doi: 10.1007/s10072-021-05341-9.
- Sokolova, Marina, and Guy Lapalme. 2009. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing & Management* 45(4):427–37. doi: 10.1016/j.ipm.2009.03.002.

- Sullivan, Gail M, and Richard Feinn. 2012. “Using Effect Size—or Why the P Value Is Not Enough.” *Journal of Graduate Medical Education* 4(3):279–82. doi: 10.4300/JGME-D-12-00156.1.
- Whittle, Michael W. 2014. *Gait Analysis: An Introduction*. Butterworth-Heinemann.