Distinguish between Obese and Normal Body Types through Gait Analysis using Classification Models

Jae-sik Yi
Dept. of Computer Science and Engineering
Incheon National University
Incheon, Korea
iics4u@inu.ac.kr

Ji Hun Han
Dept. of Computer Science and Engineering
Incheon National University
Incheon, Korea
poly9010@gmail.com

Min Jeong Kim
Dept. of Computer Science and Engineering
Incheon National University
Incheon, Korea
wlgnsdl12334@gmail.com

Youn-Sik Hong
Dept. of Computer Science and Engineering
Incheon National University
Incheon, Korea
yshong@inu.ac.kr

Abstract—For gait analysis, an IMU sensor was mounted on the knee and gait related data was collected. Various gait parameters such as gait time, stance swing ratio, heel strike, and toe off can be extracted from the dataset. To explore the relationship between gait parameters and individual gait characteristics, we analyzed the gait patterns of normal and obese people were analyzed based on BMI (Body Mass Index). To apply it to a classification model of machine learning, different gait cycles between subjects were normalized. Gait data was collected from eight subjects in their 20s. Using this dataset, we applied a logistic regression model, and obtained the classification accuracy of 92%. We also investigated the correlation between BMI and gait parameters and found that, the correlation between BMI and cadence was -0.66.

Keywords—Gait Analysis, IMU (Inertial Measurement Unit), BMI (Body Mass Index), Cadence, Classification Model

I. Introduction

There is a continuing trend of applying machine learning, the core of artificial intelligence, which has developed rapidly in recent years, to solve long-standing big data processing and high-performance computing problems in various academic and applied fields such as biotechnology and health care.

In response to this trend, we produced and used a wearable module with a built-in inertial measurement unit (IMU) sensor to analyze walking on flat ground. IMU sensors are small, easy to attach to the body, and are relatively inexpensive. IMU sensors can measure gravity and acceleration using magnetometers and accelerometers, and can detect changes in Euler angles (roll, pitch, yaw) using balance meters. Existing research mainly focused on using IMU sensors to extract walking parameters such as walking speed, gait cycle, number of steps, stance phase time, and swing phase time.

A person's gait parameters are variable depending on physical condition, gender, and age. By repeating the gait cycle for a certain period of time, meaningful parameters for gait analysis can be extracted. For this purpose, a self-produced wearable module was mounted on the subject's knee, and feature points were derived based on the change in the angle of movement of the foot.

We extracted gait parameters such as gait cycle, gait time, and number of steps per minute from each feature point, and analyzed gait by training a classification model of machine learning with a dataset composed of gait parameters. As a result, we propose to derive the person's body mass index (BMI), classify them as obese and normal people, and calculate the Pearson correlation coefficient through correlation analysis between body mass index and main characteristics to find the walking patterns of obese people compared to normal people.

To prove this, subjects of various ages, body types, and gender conditions were equipped with inertial measurement equipment and made to walk, and the collected gait dataset was input into a classification model of machine learning. As a result, gait analysis was used to distinguish obese people from normal people. The accuracy of the classification model was 0.92, which was equivalent to the accuracy of the statistical analysis technique.

II. Related Works

Weijun Tao et al. demonstrated that through gait analysis, kinematics and movement parameters of human walking events can be determined and musculoskeletal function can be quantitatively evaluated, and the current state of gait analysis technology based on wearable sensors placed on the patient's feet or waist [1].

For gait analysis, Zabir Mohammad et al. proposed an event class-based ensemble architecture approach to improve the precision of elderly fall detection. Using wearable sensors on the waist to capture various information, the proposed method was evaluated on a dataset and showed its feasibility for deployment in elderly care facilities and homes [2].

Woo Cheol Shin collected and analyzed the instantaneous current consumption used when a person walks on a treadmill in an unrestrained state, and statistically analyzed the data sets collected at specific points during the experimenters' gait cycle to determine whether differences in walking occurred. And as a result, we tried to find out the standard value of general walking [3].

Wilfried Elmenreich explains that the ability of animals to fuse data from multiple sensors is highly evolved in many animal species, and that today the application of the fusion concept in the technological domain has become a new discipline across many fields of science and is focused on the principles, architectures and methods of sensor fusion. The purpose was to give an overview of the topic [4].

De Jong Yeong et al. used multiple vision cameras, radar sensors, LiDAR sensors, and ultrasonic sensors in self-driving cars to recognize the environment, and while radar sensors can provide information such as the speed of moving obstacles and perform distance mapping, To compensate for limitations that are generally not suitable for object recognition applications due to low resolution, radar information was fused with other sensor data such as camera and LiDAR [5].

PETER S. MA YBECK is an optimal linear estimator, the Kalman filter is a recursive data processing algorithm. One view of optimality is that the Kalman filter contains all the information that can be provided, and depends on the precision to estimate the current values of the variables of interest. Process all possible measurements regardless [6].

BYJU compared the advantages and disadvantages of mean, median, and mode among statistical analysis techniques. The advantage of an average is that it accounts for every value to calculate the average, and the disadvantage is that very small or large values can affect the average. The advantage of the median is that it is not affected by very small or large values, and the disadvantage is that it is the average of locations, so when there are many observations, it takes time to sort the data in ascending or descending order. The advantage of modes is that if the dataset is not numeric, only the average can be used, and the disadvantage is that there may or may not be more than one [7].

According to D. Hazry et al., an inertial measurement unit (IMU) is a major component of inertial guidance systems used in guided missiles and ships and operates by detecting the current rate of acceleration as well as changes in rotational properties including pitch, roll and yaw. Major products such as the HG1700 and Crista MEMS of measuring devices were compared in terms of various errors [9].

III. GAIT CYCLE AND EXTRACTION OF GAIT PARAMETERS

A. *Understanding of IMU*

An inertial measurement unit (IMU) is an electronic device that measures and reports a body's specific force, angular rate, and sometimes the orientation of the body, using a combination of accelerometers, gyroscopes, and sometimes magnetometers. IMUs are typically used to maneuver modern vehicles including motorcycles, missiles, aircraft (an attitude and heading reference system), including unmanned aerial vehicles (UAVs), among many others, and spacecraft, including satellites and landers [8].

Three-axis MEMS-based gyroscopes are also being used in portable electronic devices such as tablets, smartphones, and smartwatches. This adds to the 3-axis acceleration sensing ability available on previous generations of devices. Together these sensors provide 6 component motion sensing; accelerometers for X,Y, and Z movement, and gyroscopes for measuring the extent and rate of rotation in space (roll, pitch and yaw) [8]. Fig. 1 shows the measurement of the inertial measurement device detector.

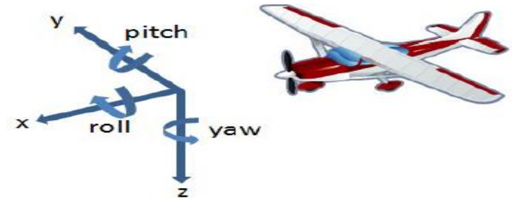

Fig. 1. An IMU sensor measures [8].

Roll refers to left and right rotation in the X-axis direction, Pitch refers to rotation in the Y-axis direction, and Yaw refers to rotation in the Z-axis direction, which is the same as the direction of gravity [8, 9].

B. *Gait Cycle Definition*

The gait cycle is divided into stance phase and swing phase as shown in Fig. 2. The stance phase, which accounts for 60% of the entire gait, is the section in which the soles of the feet contact the ground and support the body weight. The swing phase, which accounts for the remaining 40% of the entire gait, is the section where the feet are in the air. The gait cycle can be further subdivided into eight patterned gait stages. The Stance phase can be divided into 5 stages from Heel-Strike (HS) to Pre-Swing (PS), and the Swing phase can be divided into 3 stages from Toe-Off (TO) to Terminal Swing (TS).

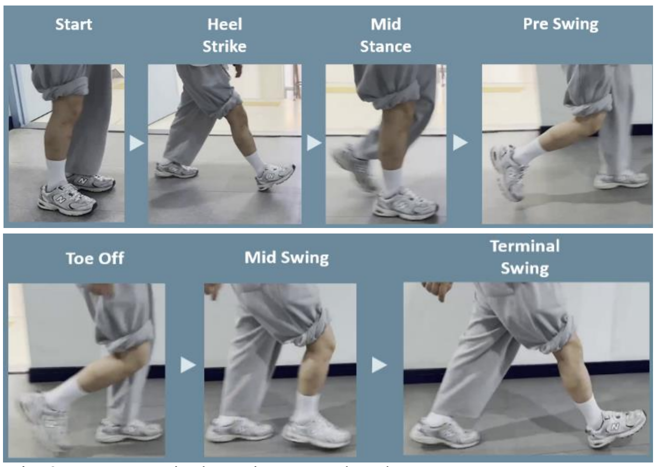

Fig. 2. Human's gait phases in a normal cycle.

C. Gait Parameter Extraction

In Table 1, ax, ay, and az represent the acceleration values measured by the accelerometer, gx, gy, and gz represent the angular velocity values measured by the gyroscope, and mx, my, and mz represent the measured values of the direction of the magnetic field affecting the magnetometer sensor.

TABLE I
EXPLANATION FOR AXISE EACH SENSOR

accelerometer	gyroscope	magnetometer [a]
ax	gx	mx
ay	gy	my
az	gz	mz

Roll, Pitch, and Yaw values, which are the rotation radii of the sensor, were derived to measure joint movement using the nine extracted values. The formulas for calculating Roll(Φ) and Pitch(θ) are as follows.

$$\varphi = atan(\frac{A_Y}{A_X^2+A_Z^2}) \tag{1}$$

$$\theta = atan(\frac{A_X}{A_Y^2+A_Z^2}) \tag{2}$$

The Yaw value, which is the radius of rotation in the Z-axis direction, cannot be obtained with an acceleration sensor because the direction is the same as the Z-axis where gravity acts. The formula for calculating Yaw(ψ) using the angular velocity sensor and geomagnetic sensor is as follows.

$$X_H = mxcos\theta + mysin\phi sin\theta + \ mzcos\phi sin\theta \tag{3}$$

$$Y_H = mycos\phi - \ mzsin\phi \tag{4}$$

$$\Psi = tan^{-1}\left(\frac{Y_H}{X_H}\right) + D \tag{5}$$

IV. GAIT ANALYSIS AND CLASSIFICATION OF OBESE AND NORMAL PERSON

A. Gait Analysis

The independent variable for extracting gait feature points was centered around physical information, and the dependent variable was centered around gait parameters acquired from roll data. The body information is the body mass index (BMI), which is the square of the height divided by the weight, and peak-to-peak value is the difference between the angle when the knees are furthest forward and the angle when the knees are furthest back. Stride time is the total walking time, stance time is the stance phase, swing time is the swing phase, and cadence is the number of steps per minute.

Gait cycles were extracted from data obtained from subjects, and gait parameters were extracted from a graph overlapping the gait cycles. The extracted gait parameters are stance time, swing time, peak-to-peak value, and cadence. The gait parameter value is the average value when gait cycles are overlapped.

Stance time and swing time are the time elapsed in the stance phase and swing phase, respectively. Gait time is the time taken to complete walking in a 15m straight line, which was set as the experimental condition in this paper. Cadence refers to the number of steps per minute. i.e. This is the number of steps taken in 1 minute, and since the IMU was worn only on the right leg in the experiment, it is the number of times the right foot touched the ground in 1 minute. Peak-to-peak value is the difference between the maximum and minimum values in the gait cycle shown in the graph in Fig. 5.

B. Classification of Obese and Normal Person

We have mostly adopted statistical analysis techniques such as mean, median, and mode for gait analysis in previous studies, and found that they have advantages but also disadvantages [3, 7]. For more accurate classification, a classification model of machine learning was adopted, avoiding statistical analysis techniques.

When the data of a total of 8 subjects was displayed as a scatter distribution chart of peak-to-peak value data according to BMI, it could be divided into a total of 3 clusters (showing the distribution chart). When the clusters were divided into two, it was confirmed that they were divided according to BMI, which represents the x-axis. A BMI of 26 or more is obese, and a BMI of less than 26 is normal or underweight. The distribution of cadence and BMI also shows a similar pattern. Subjects with higher BMI were classified as having significantly lower steps per minute.

The sample Pearson correlation coefficient is the covariance of two variables divided by the product of their respective standard deviations in interval scale (interval scale) or proportional scale (ratio scale) data. The Pearson correlation coefficient between BMI (label) and main characteristics shown in Fig. 3 is as follows.

- Gait Time (= 0.62) : A positive correlation coefficient with a large value means that the larger the BMI, the longer the Gait Time, meaning that obese people have a slower walking speed than normal people.
- Peak-to-peak value (= -0.45) : It is a negative correlation coefficient with a moderately large value, meaning that the larger the BMI, the smaller the PTP, meaning that obese people have shorter strides than normal people.
- Gait Number (=0.49): It is a positive correlation coefficient with a moderately large value. The larger the BMI, the more steps per minute (cadence), meaning that obese people take more steps than normal people even when walking the same distance.

i.e. When walking, obese people generally move their knees at a smaller angle, resulting in shorter strides, so even when walking the same distance, they take more steps and walk at a slower speed than normal people.

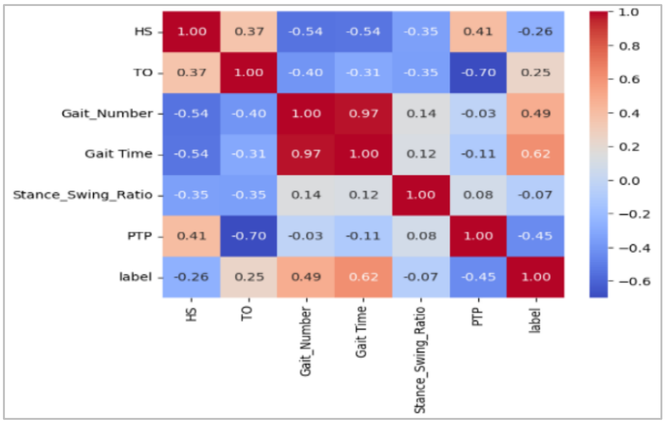

Fig. 3. Pearson Correlation Coefficient Collection between Gait Parameters

V. EXPERIMENT

A. Building an Experimental Environment

As an experimental environment for analyzing gait patterns by BMI group, we secured 8 subjects with various physical conditions, aged from 20s to 50s, and wore IMUs on their right knees, as shown in Fig. 4. As you can see, the subjects were asked to walk a distance of 15 meters, repeated 10 times. The data detected by the IMU was transmitted to a high-performance laptop PC via wireless communication and then stored in the Firebase database. For reference, it is possible to experiment with the IMU worn on both legs, but in preliminary experiments, no significant differences were found in the gait cycle between the left and right legs, so it was worn only on the right knee.

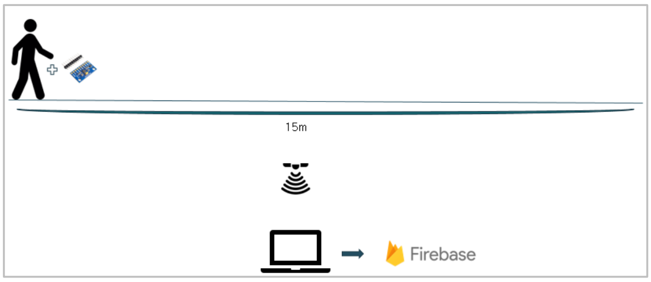

Fig. 4. Experiment Environment

B. Experimental Process

Fig. 5 uses the Python Pandas library for gait pattern analysis to convert the subjects' gait sensor raw data into a csv format that is easy to handle in an analysis program, and then reads the data into a data frame, including heel strike, toe off, gait number, and gait. It shows 6 columns of time, stance swing ratio, and peak-to-peak value, and a total of 841 rows.

To analyze the sensor data collected in the experimental environment, we used Python's Pandas library to convert the sensor data into csv format, performed data preprocessing to remove unnecessary columns, and programmed using Logistic Regression as a classification model of machine learning. Obesity was classified, and the results were visualized using the matplot library, and the results of correlation analysis of parameters using the Pearson correlation coefficient were visualized as a heatmap.

	HS	TO	Gait_Number	Gait Time	Stance_Swing_Ratio	PTP
0	0.075659	0.426537	0.285714	0.357631	0.193890	0.319674
1	0.399393	0.416467	0.285714	0.357631	0.417693	0.530924
2	0.305106	0.412104	0.285714	0.357631	0.329058	0.474930
3	0.335456	0.423516	0.285714	0.357631	0.335315	0.485365
4	0.329386	0.392636	0.285714	0.357631	0.353563	0.504963
..	...	...	...	...	...	...
836	0.132110	0.776455	0.285714	0.369021	0.317066	0.089845
837	0.130896	0.757994	0.285714	0.369021	0.335315	0.103080
838	0.108235	0.765379	0.285714	0.369021	0.353563	0.083227
839	0.139394	0.772427	0.285714	0.369021	0.371812	0.097480
840	0.258771	1.000000	0.285714	0.369021	0.353563	0.000000

[841 rows x 6 columns]

Fig. 5. Readed and Scaled Data

Fig. 6 shows that the classification accuracy of the Logistic Regression model is 0.92.

Accuracy: 0.92
Classification Report:

	precision	recall	f1-score	support
0	0.95	0.86	0.91	73
1	0.90	0.97	0.93	96
accuracy			0.92	169
macro avg	0.93	0.92	0.92	169
weighted avg	0.93	0.92	0.92	169

Fig. 6. Accuracy of Logistic Regression Model

Fig. 7 shows the performance of the classification model as an ROC curve, and the total area under the curve (AUC) was 0.85 to 0.93, indicating that the sensitivity and specialness of binary classification were highly evaluated.

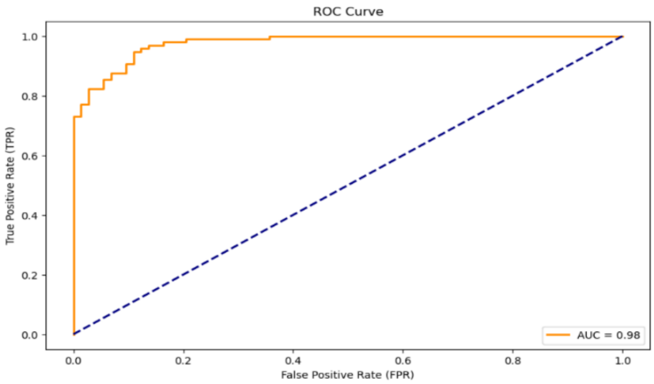

Fig. 7. ROC Curve

The receiver operating characteristic (ROC) curve is a graphical representation that describes the performance of a binary classification model at varying threshold values and is a representation of the true positive rate (TPR) relative to the false positive rate (FPR) at each threshold setting [8].

C. Experimental Result

We analyzed the correlation between body BMI and the two parameters extracted during the subjects' walking, peak-to-peak value and cadence. The correlation index between BMI and Peak-to-peak value was -0.84, and the correlation index between BMI and Cadence was -0.66. .

It was concluded that in the case of the obese group with a high BMI, the cadence was small, the walking speed was slow, and the stride length was also short.

VI. CONCLUSION

We were able to determine the gait cycle from a dataset that preprocessed the roll raw data measured by the accelerometer sensor built into the IMU while the subjects walked while wearing the IMU. To analyze gait patterns, gait-related parameters were extracted by overlapping gait cycles, and among them, peak-to-peak and cadence were used to distinguish BMI groups. In addition, in order to minimize the impact of distinct physical conditions on the experimental results and to infer the characteristics of individual walking patterns, the subjects were selected using a control method. In the experiment targeting the control group, significant differences were confirmed between the subjects, and based on this, Gait feature points were extracted from the stance phase, which accounts for 60% of the gait cycle.

Although there was a limited number of subjects, it was confirmed that gait patterns could be meaningfully determined using only gait feature points. The accuracy of the classification model of machine learning(Logistic Regression) trained with the preprocessed dataset for walking pattern analysis was very high at 0.92, proving that feature point-based walking pattern analysis is applicable.

In this paper, we proposed applying the computer engineering field of gait analysis using a classification model of machine learningto the health and medical field to distinguish between obesity and normal body type using human BMI, and the accuracy of classification of experimental results was also evaluated using statistical analysis techniques. Equivalence was confirmed.

In the near future, based on this study, we plan to continue research on an abnormal gait discrimination model by analyzing gait posture by applying a machine learning ensemble model to healthcare in the health care field. Additionally, I think that linear regression model is one of nice classification model of machine learning. I will adopt other 1-2 classification models including SVM in experiment at my additional study.

REFERENCES

[1] Weijun Tao, Tao Liu, Rencheng Zheng and Hutian Feng, Gait Analysis Using Wearable Sensors, Sensors, December 2012, pp. 2255–2283, 16 February 2012.

[2] Zabir Mohammad, Arif Reza Anwary, Muhammad Firoz Mridha, Md Sakib Hossain Shovon and Michael Vassallo, An Enhanced Ensemble Deep Neural Network Approach for Elderly Fall Detection System based on Wearable Sensors, MDPI, January 2022, pp.1–17., April 26, 2023.

[3] Woo Chul Shin, Gait analysis on the treadmill in an unconstrained state, Incheon National University Graduate School, Master Grade, pp. 1–47, July 2023.

[4] Wilfried Elmenreich, An Introduction to Sensor Fusion, Institut fur Technische Informatik, Research Report 47, pp.1-29, November 19, 2002.

[5] De Jong Yeong, Gustavo Velasco- Hernandez, John Barry and Joseph Walsh, Sensors and sensor technology in autonomous vehicles: A review, Sensors, 2021, 21, 2140, pp.1-37, 18 March 2021.

[6] PETER S. MA YBECK, The Kalman Filter: An Introduction to Concepts, Autonomous Robot Vehicles, pp. 194–204, 1990.

[7] BYJU'S, What are the advantages and disadvantages of mean, median and mode?, https://byjus.com/question-answer.

[8] wikipedia, raw, pitch, yaw, https://en.wikipedia.org/wiki/Euler_angles

[9] D. Hazry, M. Sofian, and A. Zul Azfar, Study of Inertial Measurement Unit Sensor, ICoMMS, October 2009, pp. 5A8-1-5A8-4, 11 – 13 October 2009.