Dataset Development for Classifying Kick Types for Martial Arts Athletes Using IMU Devices

Rudy Gunawan, Suhardi, Widyawardana Adiprawita, Tommy Apriantono
School of Electrical Engineering and Informatics, Bandung Institute of Technology (ITB), Bandung 40132, Indonesia

***Abstract*—This study aims to establish a dataset of kicks in Kempo martial arts to categorize athletes' kick types based on their movement patterns. The real problem addressed in this research is the lack of an accurate, efficient, and portable system to automatically recognize and classify kick types in martial arts training, especially outside of controlled laboratory environments. Previous studies often relied on optical motion capture systems, which, while accurate, are expensive and impractical for real-world training settings. To overcome this limitation, this study utilizes wearable motion sensing technology and analyzes the data with an Inertial Measurement Unit (IMU) sensor. During data collection, IMU sensors are attached to athletes to monitor their movements during training or competitions. In this research, an IMU sensor type MPU6050 is used, controlled by an ESP32 microcontroller, and the data is collected via wireless communication to a data collection server. The dataset comprises gyroscope readings for angular velocity and accelerometer measurements for linear acceleration in three axes. The study evaluates three kick types: straight kicks, sidekicks, and roundhouse kicks. It employs machine learning methodologies utilizing three principal classification algorithms: Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Random Forest (RF). These algorithms were selected due to their distinct strengths in processing sensor data derived from the accelerometer and gyroscope embedded within the IMU device. The findings indicate that the SVM algorithm successfully identified and categorized kick types in Shorinji Kempo martial arts athletes with a 96.7 per cent accuracy rate when using a dataset of 70 samples and two sensors. However, when three sensors were used, the accuracy decreased to approximately 92.4 per cent. In contrast, the k-NN algorithm achieved a classification accuracy of 92.4 per cent with a dataset of 70 samples, k = 3, and three sensors. Analyzing the contributions of features to classification provides in-depth insight into the key characteristics of movement patterns for kick type recognition.**

Keywords*—*Classification; dataset; machine learning; inertial sensors; IMU; martial art; sport; motion capture

I. Introduction

Motion pattern recognition plays a crucial role in various fields, spanning across industries and scientific disciplines. There are several reasons why recognizing movement patterns is significant. In healthcare and patient rehabilitation, movement patterns are utilized to monitor progress and offer feedback on their physical movements. In sports, analyzing movement and recognizing patterns can aid in assessing an athlete's performance and subsequently enhancing skills through specific training programs. In the advancement of robotics for healthcare or disability assistance, recognizing movement patterns can enable robots to comprehend and respond to human movements. In scientific and psychological research, identifying movement patterns contributes to understanding human and animal behavior.

Sports research often involves capturing the motion of athletes for analysis. This analysis is crucial for enhancing team strategies and monitoring the technical skills of individual athletes. "Human motion capture", the process of recording human movements, aims to record the whole position of an athlete's body segments. This task might be a tough and time-consuming operation to collect data on the precision and usefulness of the measurement system.

Shorinji Kempo is a Japanese martial art that combines self-defense techniques, philosophy, and character development. Its techniques include punches, kicks, locks, and throws designed to incapacitate an opponent without causing serious injury. In addition to the physical aspects, Shorinji Kempo also emphasizes moral values, discipline, and teamwork, making it more than just a sport, but also a means of personal development.

In some countries, this martial art is competed in various sports events. Shorinji Kempo competitions are divided into two main categories, namely “embu” (technical competition) and “randori” (fighting). To ensure the safety of participants, especially in randori, several strict rules are applied.

Randori is a form of competition in Shorinji Kempo that involves a controlled free fight with strict rules and regulations. Both fighters are required to wear body protectors, helmets, and specialized boxing gloves. In randori, participants use their hands and feet to score points while maintaining safety and adherence to the rules. The head target is only intended for hand punches, while the kick target is only for the body. Safety equipment for randori matches includes head protection, groin protector, gloves (hand gloves) and body protectors.

Motion capture is essential in analyzing kicks in Shorinji Kempo matches because it can provide accurate data on the speed, angle, and efficiency of an athlete’s movements. In these matches, more points are earned from kicks than punches, so detailed analysis of kicking technique is crucial to improve an athlete’s performance and strategy. With motion capture, coaches and athletes can evaluate the accuracy of kick execution, correct deficiencies, and optimize strength and balance to increase the chances of earning maximum points. In addition, this technology also helps to ensure that the technique used is in accordance with competition regulations and minimizes the risk of injury.

In this research, the challenges for accurately classifying martial arts kick types using IMU sensor data due to variations in motion patterns and sensor noise is studied. It will also highlight the specific contributions of this study, which include: 1) constructing a kick motion dataset using accelerometer and gyroscope data from multiple IMU placements; 2) evaluating the performance of machine learning algorithms (SVM, k-NN, and Random Forest) for kick classification; and 3) analyzing how sensor placement and feature extraction affect classification accuracy.

II. Literature Review

A. Motion Capture Technique

Motion capture is an advanced technique employed to accurately record human or object movement, subsequently converting it into digital data for extensive analysis across diverse fields, including animation, sports, gaming, and biomechanics research. This method typically relies on the application of sensors or markers affixed to the subject, capturing essential metrics related to position, orientation, and dynamic behavior of movement. Systems can utilize optical cameras or inertial sensor units (IMUs), or a hybrid of both, to enhance data precision. The resulting information facilitates a comprehensive understanding of movement patterns, enables the replication of actions in virtual settings, and aids in assessing physical performance metrics, such as an athlete's movement efficiency or rehabilitation progress. While motion capture offers highly detailed movement data, challenges remain, including the necessity for costly equipment and the requirement for controlled environments.

Eline Van Der Kruk et al. [1] grouped several motion capture techniques based on the technology and working principles used to capture movement, which are as follows:

1) Optical systems. This method utilizes cameras to track the position of markers or human body features and can be categorized into two types. The marker-based systems use reflective or active markers (such as LEDs) placed on the subject's body, where cameras capture light from the markers to reconstruct movement. Markerless Systems do not require markers; instead, they rely on computer vision algorithms to recognize and track body movements directly from video images. This approach offers high accuracy, particularly in marker-based systems, and is widely used in laboratories and sports applications. However, it requires a controlled environment. The marker-based systems can be time-consuming to set up, while potentially interfering with an athlete's natural movements.

2) Inertial measurement systems. This method utilizes inertial sensors, such as accelerometers, gyroscopes, and magnetometers, attached to the body to measure acceleration, angular velocity, and orientation. It offers the advantage of being usable outside the laboratory, including in sports fields, without the need for specialized cameras or environmental settings. However, its accuracy depends on proper sensor calibration, and it is susceptible to drift errors in orientation data, especially during long-term motion measurements.

3) Electromagnetic systems. This method utilizes electromagnetic fields to track the orientation and position of sensors attached to the body. It provides high accuracy when used in a controlled environment. However, it has certain limitations, including susceptibility to electromagnetic interference from the surrounding environment and a restricted range, making it less suitable for large movements or field sports.

4) Mechanical systems. This method utilizes mechanical circuits attached to the body to directly measure joint angles or positions. It offers high accuracy in joint angle measurement and is not affected by external environmental factors such as light or electromagnetic interference. However, it also has drawbacks, including restricting the athlete's natural movement due to the typically large and rigid mechanical devices. Additionally, it does not provide direct spatial information, as it only measures joint angles.

5) Hybrid systems. Combining two or more motion capture technologies, such as optical and inertial systems, aims to compensate for the weaknesses of each method. This approach offers several advantages, including increased accuracy and flexibility by utilizing the strengths of each system, as well as the suitability for using both in the laboratory and field. However, it also presents challenges, such as the complexity of hardware and software, and higher implementation costs.

The "gold standard" for motion tracking at the moment is optical motion capture (OMC) systems, which use cameras to reconstruct the three-dimensional position of reflecting markers affixed to the body. However, this system also has several weaknesses. The system operates by having the camera detect light emitted by both active and passive markers [2]. Measurements need to be conducted indoors to prevent sunlight from affecting the data. Setting up camera systems necessitates costly laboratory facilities, and ensuring synchronized recording by the cameras is challenging [3]. While these systems boast high accuracy, they require a controlled laboratory setting and often involve a cumbersome setup [4].

The use of an exoskeleton-based mechanical system to measure and adjust ankle joint stiffness during rehabilitation therapy. The exoskeleton is equipped with mechanical sensors such as rotary encoders and actuators to provide real-time measurements and feedback. The implementation of the mechanical system is used for biomechanical analysis, rehabilitation, and sports applications. Mechanical systems are very relevant for high-precision measurements in this field, although they have challenges such as user comfort and limited flexibility.

To get around the drawbacks of traditional motion capture systems, researchers propose using recently developed inertial measurement units and magnetic angular rate gravity (MARG) sensor technologies. The use of IMU devices is seen as a more feasible option, particularly for data collection in non-laboratory settings.

B. Application in Sports

The movement measurement system is categorized into sports based on the nature, area, or place, and the number of athletes being observed. Individual sports and team sports (involving more than three players) are distinguished. Team sports involve significant measurement and occlusion, unlike individual sports that focus more on technical aspects. Precision is less crucial in team sports as the emphasis is on tracking team movements. Individual sports are further divided into larger and smaller volume categories, in addition to indoor and outdoor classifications. A highly accurate optical motion technology is being used to cover smaller quantities, while larger volume individual sports are best suited for kinematics measurements. The image processing system and IMU are the most suitable options for this purpose [1].

In the realm of sports, each technique serves a specific purpose and has its own set of performance indicators. For instance, the indicators of leg muscle strength utilized for running in athletics will differ from those used for kicking in combat sports. In running, the focus is on achieving maximum speed and sustaining movement over an extended period. On the other hand, kicking techniques in martial arts showcase performance indicators such as the speed and strength of a kick. Therefore, evaluating the performance of a sports technique must consider the way the technique is executed, as each technique engages different sets of supporting muscles.

Analyzing an athlete's movement patterns is crucial in assessing the strength and effectiveness of their abilities. The strength of a martial artist's leg muscles during a kick should be complemented by reaction speed and kicking velocity. Consequently, evaluating an athlete's strength should be based on their performance of techniques aimed at scoring points in a match.

The IMU is a micro-electromechanical systems (MEMS) technology-based system. It is an inexpensive, portable, lightweight, and simple-to-implement system made up of tiny parts that doesn't require fixed infrastructure [5]. These sensors typically include accelerometers and gyroscopes, sometimes magnetometers also, to monitor changes in position and orientation. IMUs often engage in a process called "sensor fusion", in which data from several types of sensors are combined or "fused" to produce more complete or accurate information about a system or environment.

In various studies on motion capture, the quantity of sensors utilized in sensor fusion IMU measurements may vary. Some researchers choose to use all three available sensors, while others may choose to use only a subset of them. This decision is based on specific requirements, and there are several reasons why a sensor may not be used for a measurement.

Numerous research employs accelerometers, designated as GF-IMU (gyroscope-free IMU) [6], to ascertain relative motion solely by accelerometer data. This method utilizes three triads of accelerometers to measure both transverse and rotational accelerations, together with angular velocity, eliminating the necessity for gyroscopes. By eliminating gyroscopes and embracing flexible accelerometer placement, this GF-IMU design lowers cost and enhances deployability in real-world scenarios. It paves the way for lightweight, adaptable inertial sensing. The work examines the GF-IMU as a nonlinear control system and employs a nonlinear Kalman filter to analyze the state. It demonstrates that three triads suffice for precise motion estimation, while more triads merely provide supplementary information.

The research focuses on estimating lower extremity motion trajectories with a Graph Convolution Network, utilizing five IMUs placed on the lower extremities, each equipped with a 3-axis accelerometer and 3-axis gyroscope sensor [7]. Luigi Truppa et al. [8] conducted research on innovative sensor fusion algorithms for motion tracking with online bias compensation, utilizing only inertial sensors (e.g., accelerometers and gyroscopes) due to the susceptibility of magnetic sensors to magnetic interference. Another study focused on reducing the number of IMUs for motion sensors, as seen in the research by Wicent et al. [10], which involved estimating lower body kinematics using a Kalman filter with Lie Group restrictions, employing only 2 sensors located at the hip and feet positions.

By integrating motion pattern recognition technology with artificial intelligence and advanced sensors, deeper insights can be obtained, leading to the development of smarter and more responsive systems tailored to different contexts and requirements. [9].

Research on motion recognition using machine learning has been extensively conducted. For instance, evaluating machine learning algorithms for workers involves utilizing seventeen motion sensors and employing five machine learning algorithms to categorize movements [11]. The five algorithms consist of support vector machine, logistic regression, k-NN, random forest, and multilayer perceptron. Akhavian and Behzadan [12] analyzed construction worker productivity by applying machine learning classification methods using data from accelerometers and gyroscopes integrated into smartphones. Another study focuses on creating a system for whole-body motion estimation using IMU to represent the contact conditions between the product and the user based on user-defined contact boundaries [13]. Cecilia Monoli et al. [14] have developed and validated a method for monitoring human gait kinematics underwater using an IMU wearable sensor.

C. Machine Learning Classification

The choice of machine learning algorithms will affect the results of the study. Several factors, including training speed, prediction speed, minimum data sets (MDS), performance, accuracy, ease of use, ease of implementation, and the availability of resources required to train the model, can be considered when selecting an appropriate classification algorithm.

The Support Vector Machine algorithm approach in machine learning has proven to be very effective for regression and classification tasks. SVM utilizes an optimal hyperplane that maximizes the margin between classes, enabling precise data classification. By applying the maximum margin principle, SVM enhances model robustness against variations in new data, relying on the distance between the hyperplane and the closest data points from each class. Its effectiveness is further amplified, especially with large datasets, as it only utilizes support vectors, a small subset of the training data, resulting in efficient computation and improved generalization.

The kernel in the SVM algorithm is a mathematical function used to transform the data feature space into higher dimensions. Kernels are crucial in SVM as they enable the model to address non-linear classification problems. This kernel function empowers SVM to identify intricate decision boundaries in the high-dimensional feature space.

The k-Nearest Neighbors algorithm is the simplest yet most effective machine learning method in classification tasks. The fundamental principle of k-NN is that similar objects tend to be close to each other in feature space. In classification, objects are labelled based on the majority labels of their k nearest neighbors. Selecting the appropriate k value is crucial when implementing k-NN. A small k value can make the model sensitive to noise, while a large k value can lead to an overly general model. Some advantages of this algorithm include its simplicity, ease of implementation, lack of assumptions about data distribution, and suitability for small datasets with simple structures. The k-NN is a suitable choice for straightforward tasks and small datasets, but careful consideration is needed for parameter selection and noise management to achieve optimal results.

Utilizing decision trees, the random forest algorithm is a supervised machine learning technique used for regression, classification, and other applications. Random forests excel in handling large and complex datasets, high-dimensional feature spaces, and analyzing feature importance. This technique is widely adopted across various sectors due to its ability to minimize overfitting and maintain high prediction accuracy.

Random forest classifiers build a collection of decision trees using a randomly chosen subset of the training data. These decision trees (DT) are aggregated from a randomly selected subset of the training data to produce the final prediction. The results of each decision tree are then combined.

SVM and k-NN are two widely used algorithms in the domain of supervised machine learning, where models learn from labeled datasets to make predictions on unseen data. These algorithms excel at handling both classification and regression problems, making them versatile tools in various real-world applications.

III. Proposed System

The goal of this study is to establish a dataset of athletes' leg kicks to develop models for categorizing martial arts kick techniques. To achieve this, a system must be developed to collect data on an athlete's movement, while executing a kicking technique.

In this study, we presents a system for tracking human movements in real-time using IMU sensor fusion. This sensor will record data on an athlete's kicking movements to create a kick dataset for developing models to classify kicking techniques for martial arts athletes.

The tool to be developed is a wearable device that captures the kicking movement patterns of a martial arts athlete. This device will gather data using a gyroscope and accelerometer sensor combined with an IMU sensor. It is designed to function wirelessly, making it convenient for athletes to perform their movements. Therefore, including a battery in each device is essential. The device will be placed on the part of the athlete's body, where the movements will be observed.

The system proposed utilizes the MPU6050 IMU sensor positioned on various sections of the athlete's leg to gather gyroscope and accelerometer data. An ESP32 microcontroller manages the task of acquiring and transmitting data from each device. The data will be wirelessly transmitted for collection on a server computer, where it will be stored and subsequently processed using specific algorithms. The system proposed is depicted in Fig. 1.

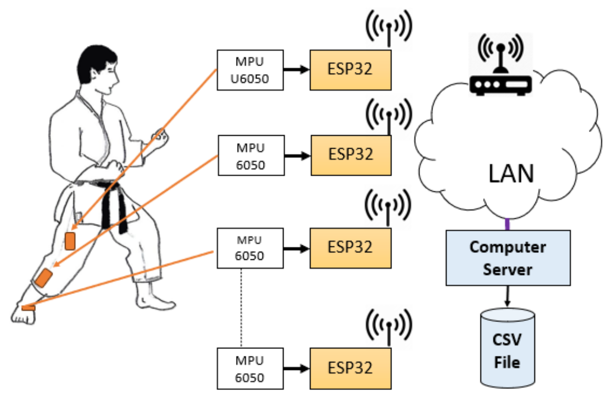

Fig. 1. Proposed motion capture system using IMU 6050.

This sensor position point was selected to accurately represent an athlete's kicking movement. There are three key points for sensor placement: the thigh, shank, and foot. Each sensor, used to capture movement on an athlete's body, follows the device construction depicted in Fig. 2. Each sensor includes an MPU6050, ESP32 as a controller, and a 3.7V lithium-ion polymer (LiPo) battery for power supply and charging control. The battery is essential for enabling the sensor to operate independently without requiring an external power source.

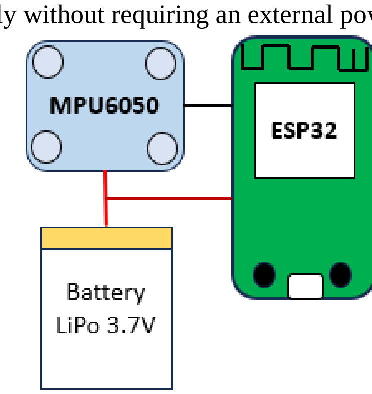

Fig. 2. Configuration of the components of IMU device.

This study focuses on analyzing leg kick techniques in a martial arts context. The research will develop a tool to classify the types of kicks performed by athletes based solely on the output of the kick itself, excluding factors like muscle strength and fatigue. The study involves testing a dataset with machine learning algorithms to determine the minimum data required for accurate classification.

The research progress through three stages: first, developing a motion capture tool using inertial sensors; second, using this tool to record athletes' kicking movements and create a dataset of martial arts kicks; and third, testing the dataset to predict the classification of each kick.

A. Hardware Specification

The MPU 6050, manufactured by InvenSense Inc., comprises three 16-bit analog-to-digital converters (ADCs) for converting the gyroscope output and three additional 16-bit ADCs for converting the accelerometer output. To accurately track both fast and slow movements, the components offer a user-programmable gyroscope full-scale range of ±250, ±500, ±1000, and ±2000°/sec (dps) and a user-programmable accelerometer full-scale range of ±2g, ±4g, ±8g, and ±16g. The MPU6050 supports a sample rate of 1 KHz. Communication with device registers can be achieved through either I2C at 400 kHz or SPI at 1 MHz.

ESP32 is a cost-effective, energy-efficient System on a Chip (SoC) microcontroller created by Espressif, featuring Wi-Fi and Bluetooth capabilities along with a dual-core processor. It is suitable for use in industrial settings, with an operational temperature range of -40°C to +125°C. The ESP32 boasts a Tensilica Xtensa Dual-Core 32-bit LX6 microprocessor running at either 160 or 240 MHz. It is equipped with 448 KB of ROM for booting and core functions, as well as 520 KB of SRAM for data and instructions.

The study confirms that the ESP32 can reliably sample sensors at high rates, making it suitable for real-time applications. However, its built-in ADC lacks precision, emphasizing the need for an external ADC to ensure accurate sensor data acquisition [15].

User Datagram Protocol (UDP) is utilized to manage the transmission of data produced by multiple sensors from a client device to a server computer for storage and subsequent processing. UDP is part of the Internet Protocol (IP) suite and functions at the transport layer within the OSI (Open Systems Interconnection) reference model. With the inclusion of a WiFi module, UDP is readily accessible in the ESP32 microcontroller chip, simplifying the wireless data transfer process.

For this application, the User Datagram Protocol was selected mainly because of its low latency and lightweight nature, which makes it appropriate for real-time data transfer. The connection-oriented strategy and error-checking methods of Transmission Control Protocol (TCP) result in overhead, whereas UDP functions as a connectionless protocol. As a result, data delivery is accelerated with less latency since acknowledgment packets, retransmissions, and congestion control are no longer necessary.

In applications like motion capture or real-time monitoring, where data from IMUs or other sensors need to be transmitted continuously, the slight loss of packets is often acceptable compared to the delays caused by TCP's reliability mechanisms. UDP's simplicity allows for high throughput, especially when working with high-frequency data streams, making it ideal for scenarios requiring time-critical communication, such as tracking human motion or sending sensor data for immediate processing.

This block diagram illustrates the flow of the data capture and processing system from the IMU (Inertial Measurement Unit) sensor to the stage of forming a dataset for machine learning needs. Body movement data captured by the IMU device is sent wirelessly via the UDP (User Datagram Protocol) protocol to the WiFi Access Point, then forwarded to the computer, to be recorded in the form of log data. After the data is received, a pre-processing process is carried out to clean and prepare the data so that it is ready for use, such as normalization, segmentation, or noise removal. The results of this pre-processing are then stored as a dataset, which is then used in the training or testing process of the model at the machine learning stage to recognize movement patterns or activity classification. As shown in Fig. 3.

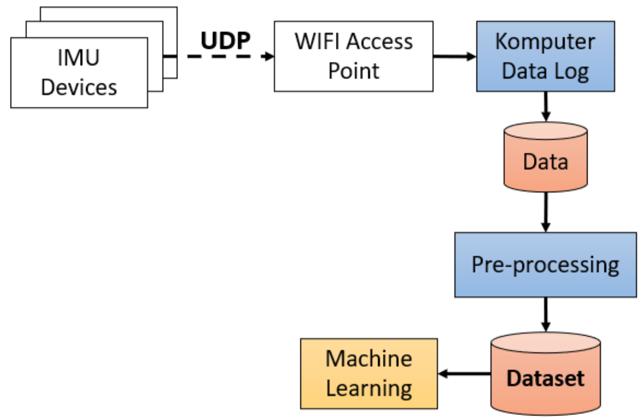

Fig. 3. Block diagram of data processing.

The sensor device starts by initializing and then searching for a connection to the server. Once the connection is established, it continuously sends sensor data to the server in a specific format. The protocol is encapsulated in datagrams using the UDP protocol and designated ports for each sensor device. The device follows a sequential process upon startup, searching for a connection to a UDP server with a distinct port number assigned to each sensor.

B. Data Transmission and Collection

The data transmission medium will be chosen wirelessly to reduce the number of cables connected to the participants. A local network will be established with a wireless access point capturing connections from each client, specifically through an IMU sensor on the volunteer subject. All sensor data will be transmitted through this connection. A computer will serve as the central server to gather data from all IMU sensors.

Multiple server applications will run on this computer to receive connections from each IMU sensor. The computer will directly connect to the wireless router access point using a UTP (Unshielded Twisted Pair) cable. Each IMU device, functioning as a UDP client, will connect to its corresponding server application set up in pairs. The separation of receiving servers will be achieved by assigning a different port to each client, allowing each server to store data independently for each sensor. The configuration of the server applications is illustrated in Fig. 4.

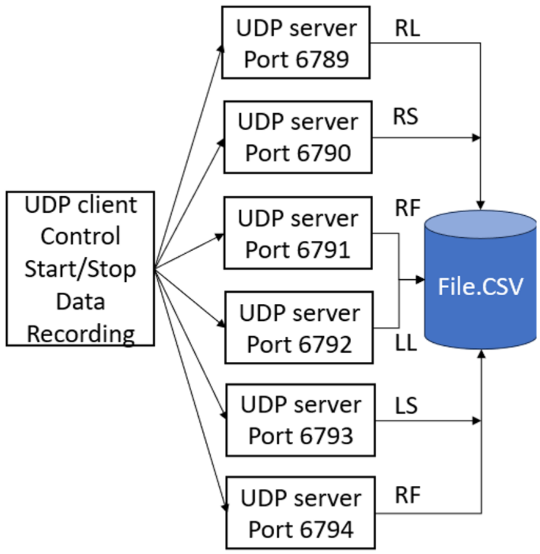

Fig. 4. Configuration of the UDP system protocol for wireless data collection from IMU devices to the server.

When the device is powered on, data from the IMU sensor will be consistently transmitted to the server. The transmitted data will be shown on the console to monitor the functionality of each IMU device. To prevent unnecessary data storage, a system was developed to determine when to begin recording data and when to cease storing it. This system prevents the storage of irrelevant data during the setup or device testing phase.

An additional program application is used to determine the start and stop times for recording the data stream. This application functions as a UDP/IP network-based client program on the server computer, sending commands to each capture server on different ports to initiate or terminate data storage. Being a network program, it can be placed on a separate computer if needed, facilitating operators in data retrieval tasks.

Data from each IMU sensor will be stored in separate comma separated value (CSV) files. Each data sent from the sensor will be received with additional time stamp data inserted before being stored in the file. All data sent by the IMU device will be stored without any prior selection process. Therefore, several pre-processing stages are necessary for formatting the data appropriately for subsequent processes. The data acquisition process on the server side is illustrated in the flow diagram in Fig. 5.

C. Data Format

Before processing the data using different classification algorithms, the data sent from each sensor device will initially be saved in a CSV file format. The sensor data will be stored in a specific order and format as shown in Fig. 5. Prior to being transmitted to the storage server, the sensor data will be supplemented with the device identity (ID) data from each sensor device. This ID represents the unique identity of each IMU sensor to distinguish one device from another.

The name of the device ID is hardcoded on each individual device. A timestamp (TM) will be added by the server application to record the arrival time of data. This timestamp will be placed at the end of each data record. Subsequently, the data along with its timestamp will be saved in a data file. The format utilizes a Unix timestamp, representing the number of seconds since January 1, 1970, 12:00 am GMT (Greenwich Mean Time).

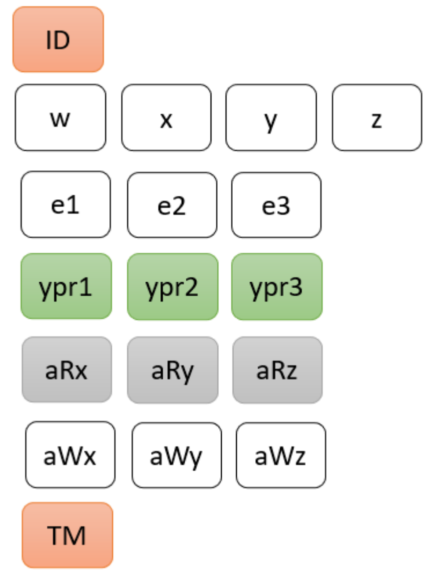

Fig. 5. Data format.

where,

- ID: device id
- w, x, y, z: quaternion
- e0, e1, e2: Euler distance
- ypr0, ypr1, ypr2: gyroscope
- aRx, aRy, aRz: real accelerometer
- aWx, aWy, aWz: world accelerometer
- TM is a timestamp

Accelerometer “raw” data are raw values directly generated by the MPU6050 accelerometer sensor without additional processing. This value is usually in the form of 16-bit data (as an integer) taken from the sensor's accelerometer register. This value represents the acceleration measured by the sensor, which is then converted to more meaningful units such as G-force or meters per second squared (m/s^2) using a certain scale.

Accelerometer "world" data is acceleration data that has been converted from a sensor frame of reference to a world frame of reference. This usually involves a transformation process to account for the orientation of the sensor so that we can understand the acceleration relative to the earth or the real-world environment.

Data obtained from the gyroscope on the MPU6050 sensor is measured in degrees per second (°/s). Gyroscopes measure angular velocity or the rate of change of angles on three axes: X, Y, and Z. In this research, the data used is only gyroscope data and raw accelerometer data.

IV. Research Method

This research follows a systematic approach to develop a classification system for identifying kick types in martial arts using IMU sensor data and machine learning algorithms. The research methodology is divided into several main stages: 1) data acquisition, 2) data preprocessing and segmentation, 3) feature extraction, 4) model training and classification, and 5) evaluation and validation.

The sensors will be placed on several parts of the athlete's feet. Each sensor will be given an identifier to distinguish where the data comes from. The position and name of the sensor location are shown in Fig. 6. For the right leg, the device labeled RL to be placed in the right thigh position, RS is for the right shank position, and the RF label is for the right sole position. Similarly, the left side displays the label LL for the left thigh position, LS for the left shank position, and LF for the left foot position.

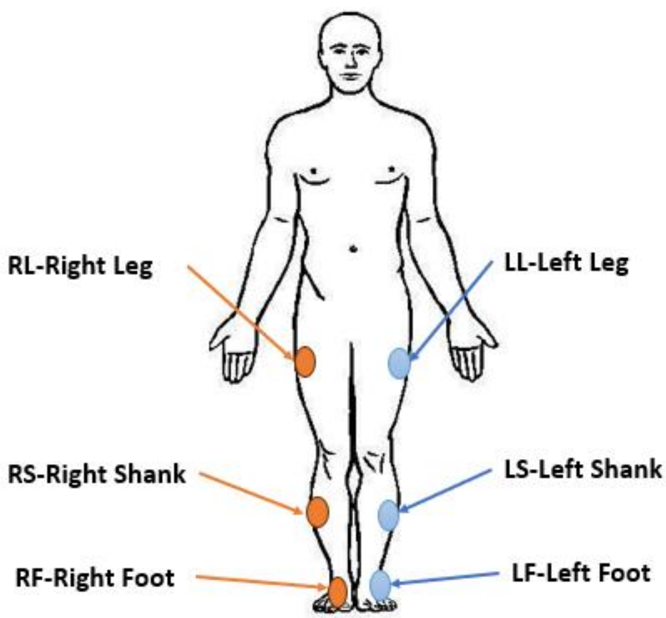

Fig. 6. Labeling of sensor positions on the lower limbs.

In biomechanical analysis, the placement of motion sensors can significantly impact the quality of data obtained, with options including placement at joints, inter-joint segments, or a combination thereof. While sensor placement at joints effectively captures local motion, such as rotation and joint angles. Interjoint placement, particularly at strategic locations like the thigh, calf, and foot, enhances the analysis of complex activities such as kicking. This method facilitates the measurement of comprehensive kinematic parameters, including translation, rotation, and acceleration across entire body segments, thus providing a more holistic view of movement dynamics. Given that, energy and momentum in activities such as kicking derived from intricate interactions within the kinematic chain focusing solely on localized joint motion, may lead to incomplete analyses and lack of insight into critical biomechanical interactions.

In the context of kicking biomechanics, inter-joint placement provides benefits in analyzing the energy relationship and momentum transfer between joints. Thus, this placement is better for capturing kicking motion patterns compared to direct joint sensor placement, which only provides local data without considering the interrelationships between body segments.

To secure the sensor to the athlete's body, a rubber band is utilized. The sensor is inserted with a plastic clip connected to the back of the device box, as illustrated in Fig. 7. This approach effectively maintains the sensor device in place during multiple leg kick movements. Periodically, the sensor device's position needs readjustment as it may shift from its original placement.

Fig. 7. Place the IMU sensor device on the respondent's leg using a rubber band.

In the data collection process, experienced martial arts practitioners were requested to demonstrate kicks. Each respondent would warm up for ten minutes before executing the specified kick movements. They would perform one kick type at a time, following the operator's start and stop commands. The kick would commence immediately upon the start signal and cease upon the stop signal. Respondents were instructed to repeat the same kick ten times. Following a five-minute break, they were then asked to perform another kick type ten times using the same procedure.

This study will create a dataset using kicks performed by martial arts experts. The dataset will serve as a reference for evaluating kick classification, focusing on three types of kicks: mawashi (roundhouse kick), maegeri (front kick), and shokuto (side kick) (see Fig. 8, 9, and 10). The dataset will specifically cover these three types of kicks. The gathered data will undergo processing to ensure the validity of the kick dataset.

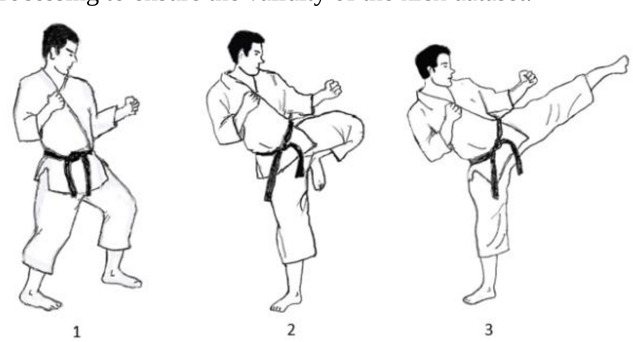

Fig. 8. Type of kick mawashi (round house kick).

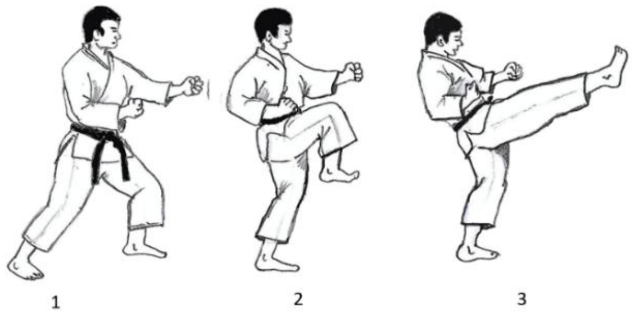

Fig. 9. Types of kick maegeri (front kick).

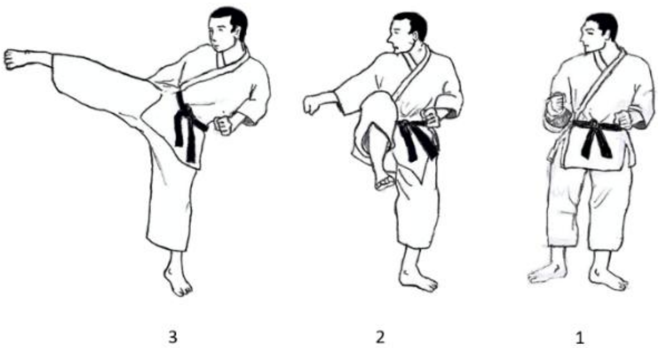

Fig. 10. Types of kick shokuto (side kick).

V. Data Processing

The process of constructing a dataset from sensor data is illustrated in Fig. 11. The data undergoes pre-processing to form a usable dataset. Pre-processing involves segmentation and re-sampling to ensure uniform length and format, specifically converting the data into a single line for each instance.

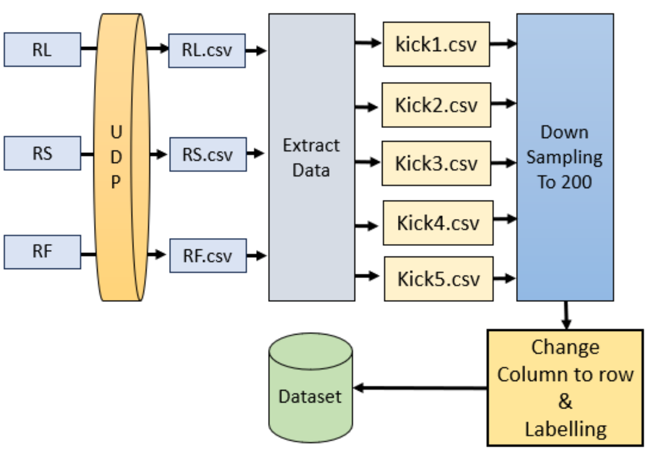

Fig. 11. Block diagram of pre-processing raw data.

Prior to being stored as a dataset, the data length for each sensor is standardized. The re-sampling procedure (down sampling) is crucial to ensure that the data length for each sensor is consistent at 200 samples. Consequently, only data with more than 200 samples is deemed suitable for further processing. Sensor data with fewer than 200 samples will be excluded due to potential imperfections in data collection.

To process sampling data with the next algorithm, the data length must be consistent. Therefore, data from sensor devices must be adjusted to match the data length. The MPU6050 sensor operates at a sampling frequency of 1 KHz, resulting in approximately 1000 samples per second. A martial arts athlete's kick lasts less than 1 second, so the average data length collected for each kick ranges from 600 to 800 samples. It is important that the adjustments made to equalize the data length do not remove significant information.

The down-sampling process is utilized to standardize the data length and reduce the number of data samples. It is crucial that this process does not result in the loss of essential information. This is done to ensure that the data interpretation remains unchanged, allowing the retained data to still capture the characteristics of a kick. Fig. 12 to Fig. 17 display the data after re-sampling. These figures illustrate that the graph's characteristics remain consistent before and after the re-sampling process.

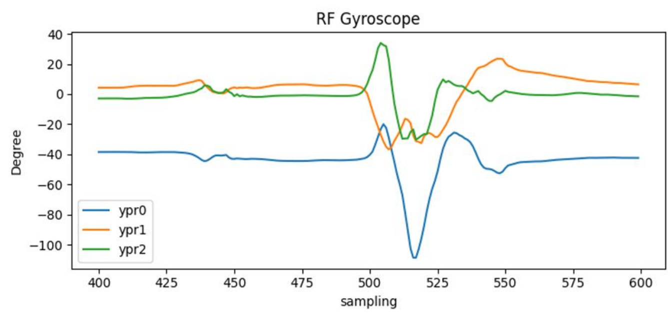

Fig. 12. Right lower limb mawashi kick data from RF gyroscope sensor data after down sampling to 600.

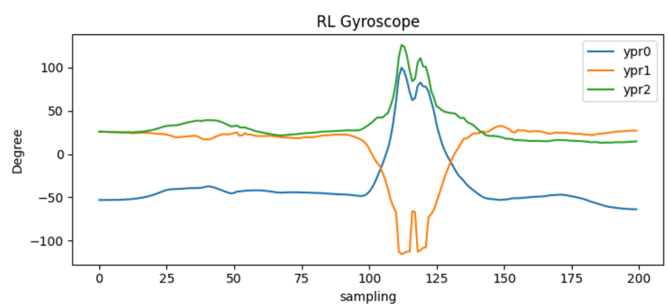

Fig. 13. Right lower limb mawashi kick data from the RL gyroscope sensor data after down sampling to 200.

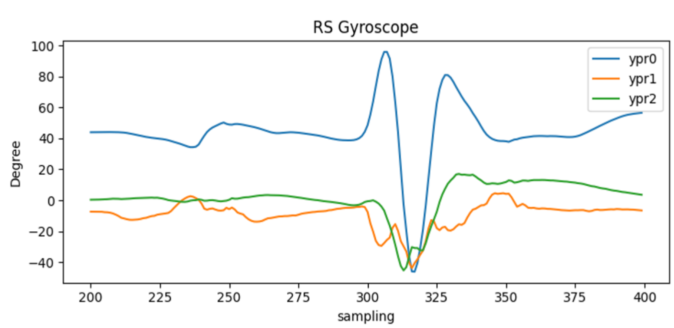

Fig. 14. Right lower limb mawashi kick data from the RS gyroscope sensor data after down sampling to 400.

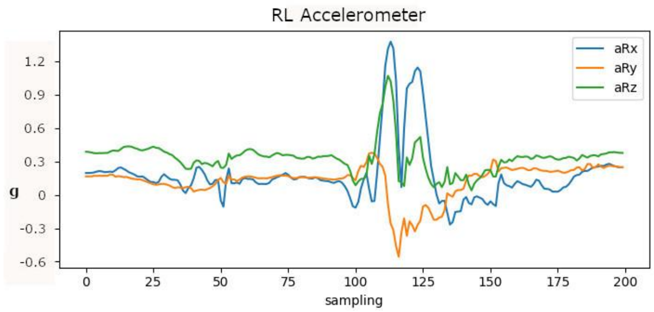

Fig. 15. Right foot mawashi kick data from RL accelerometer sensor data after down sampling to 200.

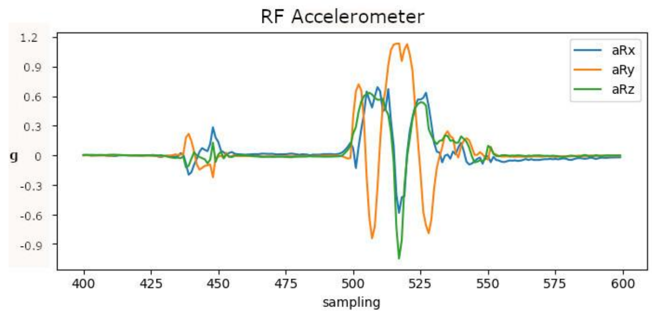

Fig. 16. Right foot mawashi kick data from the RF accelerometer sensor data after down sampling to 600.

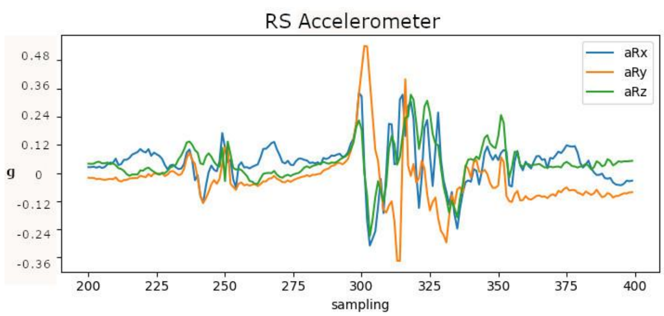

Fig. 17. Right foot mawashi kick data from the RS accelerometer sensor data after down sampling to 400.

Each sensor transmits data to the server with varying sample lengths, ranging from 500 to 800 samples per kick. The length of the data is determined by the duration of the kick. Before being input into the classification model algorithm, the data must undergo preprocessing to ensure uniformity in length. This initial process involves extracting data from files, segmenting it, and resampling to standardize the data length.

Data resampled to a length of 200 will be merged into a single file per kick, summarizing all sensor data. Sensor data with less than 200 samples will be discarded as it may indicate sensor or data transmission errors. Consequently, a significant amount of data remains unused due to sensors not meeting the minimum length requirements.

The subsequent step involves converting the data format from column-oriented to a single row for each kick. This alteration is essential to ensure that the resulting dataset contains only one row per kick. Each kick will have a column length of 3600 parameters derived from three sensors, including two gyroscopes and accelerometers, each with three outputs and a sampling length of 200.

The following step involves labeling all stored data based on the type of kick. This labeling process occurs concurrently with the input of new kick data that will be added to the dataset. To facilitate this task, a dedicated program was utilized to incorporate labels into the kick dataset.

Testing and classification of the dataset will be done using machine learning algorithms, specifically k-NN and SVM. The main concept of machine learning is to build a model that can recognize patterns in past data and apply those patterns to similar data in the future. k-NN and SVM are supervised machine learning algorithms commonly used for classification tasks.

The categorization model will subsequently undergo training and testing. The dataset will be partitioned into two segments, with 70 per cent designated for training and 30 per cent for testing. The correctness of the model will be evaluated using the findings from the test data.

The initial classification algorithm utilized was k-NN. This choice was made due to its simplicity, effectiveness for small to medium datasets, and capability to handle multi-Class. The selection of the value k in the K-Nearest Neighbors (KNN) algorithm is a crucial step that can significantly impact model performance. Typically, an odd number is chosen for k. In this case, the value of k is selected to be between 3 and 5. The evaluation of the k-NN algorithm's accuracy will be conducted using multiple datasets and varying k values. The comparison results are presented in Table II. Subsequently, the optimal number of selected k values for each dataset size will be determined.

The dataset to be tested comprises seventy data points resulting from re-sampling, pre-processing, and labeling. The model's accuracy will be evaluated to determine if it meets the criteria for a reliable prediction. Table I displays the breakdown of kick types in the dataset.

TABLE I. NUMBER AND TYPE COMPOSITION OF DATASET

Label	Number
Maegeri	16
Mawashi	22
Shokuto	32
Total	**70**

VI. RESULT

Modeling for classifying kick types will utilize seventy datasets that have been processed successfully in the prior stage. The model's accuracy will be assessed to determine if it meets the criteria for effective prediction.

The results of classification prediction accuracy using the k-NN algorithm are shown in Table II. For k=3 and dataset seventy using all sensor data (RL, RS. RF), the accuracy reaches 0.924, while if we use only two sensors the highest accuracy is achieved at 0.881 using sensors (RL, RS) with seventy datasets and k=3. For those using one sensor, the accuracy is 0.762 using an RS sensor with k=3 and a dataset of seventy.

TABLE II. ACCURACY OF THE K-NN ALGORITHM

No	Sensor	K	Dataset	Accuracy
1	RL, RS, RF	3	70	0.924
2	RL, RS, RF	5	70	0.843
3	RS, RF	3	70	0.838
4	RS, RF	5	70	0.800
5	RL, RS	3	70	0.881
6	RL, RS	5	70	0.833
7	RS	3	70	0.762
8	RS	5	70	0.714

Prediction accuracy results using the SVM algorithm with the same number of data sets are shown in Table II. The lowest accuracy value of this algorithm is 0.833, and the highest accuracy value is 0.967. When compared with the k-NN algorithm, the prediction accuracy of SVM is better.

The accuracy results with the model algorithm using random forest are shown in Table III. The highest accuracy is shown with a value of 0.900 when using 3 sensors (RL, RS, RF) and the lowest when using 1 RS sensor with an accuracy value of 0.795. For calculations using only two sensors, the accuracy value reaches 0.890 using sensors (RL, RS) and 0.905 using sensors (RS, RF).

The findings demonstrate that SVM's ability to handle small, high-dimensional datasets, combined with its use of kernel functions, allowed it to effectively capture complex patterns in the IMU data, resulting in superior classification accuracy. In contrast, k-NN and Random Forest faced challenges due to their sensitivity to feature scaling, noise, and correlated features, making SVM the best-suited algorithm for this specific task. These results highlight the importance of matching the algorithm to the dataset characteristics to achieve optimal performance.

TABLE III. ACCURACY OF SVM AND RANDOM FOREST ALGORITHM

No	Sensor	Dataset	SVM Accuracy	Random Forest Accuracy
1	RL, RS, RF	70	0.924	0.900
2	RS, RF	70	0.871	0.905
3	RSL RS	70	0.967	0.890
4	RS	70	0.833	0.795

Performance of k-NN's classification depends heavily on the distance metric (e.g., Euclidean distance). In the dataset, variations in sensor data due to noise or irrelevant features could disproportionately influence the distance calculations. The performance was contingent upon the selection of k, where smaller values resulted in overfitting and higher values induced underfitting. The k-NN performed better when kick types had distinct and isolated clusters in the feature space, but its performance degraded for overlapping or closely spaced data points, as seen in this study.

Random Forest relies on an ensemble of decision trees, which partition the feature space based on thresholds. While this approach is robust for large datasets, it can struggle with small datasets, as individual trees may overfit the training data. In the dataset, many features derived from IMU sensors (e.g., accelerations and angular velocities) were highly correlated. Random Forest frequently overvalues the significance of associated features, resulting in potential bias in feature selection and diminished generalization performance. From observation, although Random Forest showed decent performance, its accuracy was lower than SVM due to its inability to effectively capture the complex relationships between features in this dataset.

Analysis of the feature space indicated that the kick types (e.g. maegeri, shokuto, and mawashi) were not perfectly linearly separable, but there was some structured separation among the classes in the feature space. Analysis revealed that angular velocity and acceleration data were critical in distinguishing between kick types. For example, Maegeri involved a rapid forward acceleration of the shank. Shokuto displayed lateral angular velocity and sideways acceleration. Mawashi exhibited curved motion patterns, visible in both acceleration and gyroscope data.

The classification model's performance is assessed by comparing its predictions to the actual values (ground truth) using a confusion matrix. This matrix clearly indicates the number of correct and incorrect predictions for each class. Fig. 18 to Fig. 21 present the confusion matrices for each model in this study. These tables summarize how well the models classify data, highlighting both accurate and inaccurate predictions. Additional metrics such as precision, recall, and F1-score provide deeper insights into the model's strengths and weaknesses across various application scenarios.

F1-score is a classification model evaluation metric that combines precision and recall into one value. This metric is useful when there is class imbalance in the dataset and when we want to find a balance between False Positives (FP) and False Negatives (FN). The higher the F1-score, the better the model is at handling the balance between precision and recall.

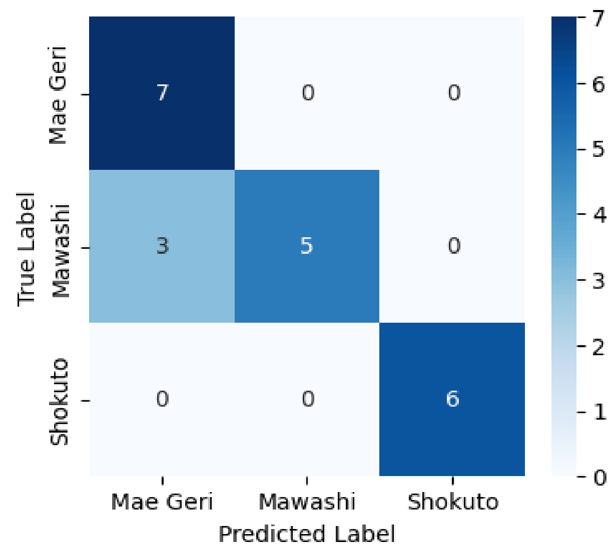

Fig. 18. Confusion matrix k-NN for k=3.

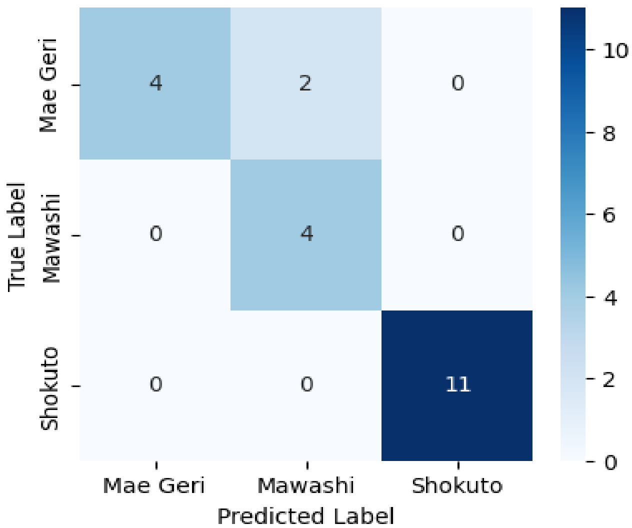

Fig. 19. Confusion matrix k-NN for k=5.

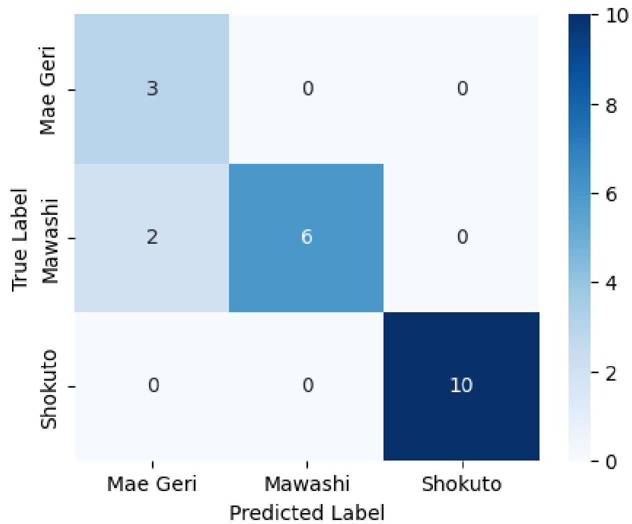

Fig. 20. Confusion matrix random forest.

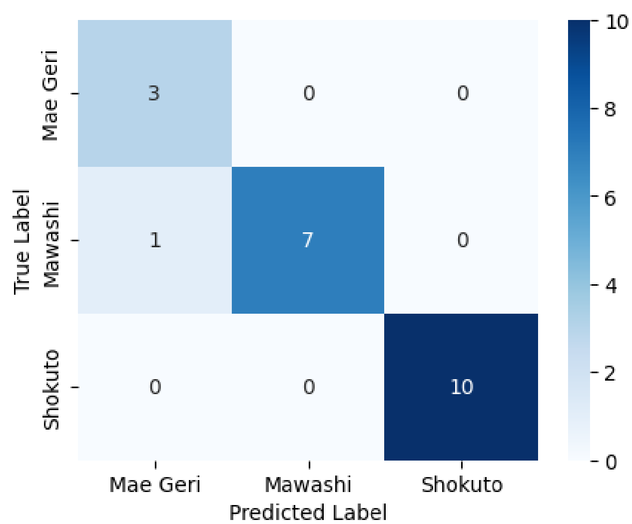

Fig. 21. Confusion matrix SVM.

In the k-NN model with a value of k = 3 (Fig. 18), the F1-score value for this model is Mae Geri: 0.82, Mawashi: 0.77, and Shokuto: 1.00. The average will be F1-score 0.86, and if using weighting according to the number of datasets in its class, it is 0.85. While for the k-NN model with k = 5 (Fig. 19), the F1-Score values are Mae Geri: 0.8, Mawashi: 0.8, and Shokuto: 1.00. The average is 0.87, and if using weighting according to the number of datasets in its class, it is 0.90. The Shokuto class has a perfect F1-score, while Mae Geri and Mawashi still have some errors. This shows that the model has a fairly good performance overall, especially since the Shokuto class has a larger number of samples and high scores, thus increasing the average.

For the random forest model with the confusion matrix shown in Fig. 20, the F1-score value per class is Mawashi = 0.86, Shokuto = 1.00. For the average value of the macro F1-score = 0.87 and the weighted F1-score = 0.90. The model has good performance with a weighted F1-score of 90 per cent, indicating a fairly accurate classification. Shokuto is perfectly classified, without errors. Maegeri has a lower recall (0.60), so it needs to be improved to reduce classification errors, and Mawashi has a good balance between precision and recall.

For the SVM model shown in the confusion matrix in Fig. 21, the F1-score values per class are Mae Geri = 0.86, Mawashi = 0.93, and Shokuto = 1.00. The average F1-score value = 0.93 and the weighted F1-score value = 0.95. This model has very good performance, with a weighted F1 score of 95 per cent, which means the model is able to classify all three types of kicks quite accurately. Shokuto is perfectly classified (F1-score = 1.00), Mae Geri has a small error in recall, so the improvement can focus on reducing the classification error for Mae Geri, while Mawashi has a good balance between precision and recall.

VII. Discussion

This discussion provides an in-depth overview of the implementation of kick classification using IMU and k-NN and SVM machine learning algorithms. The success of both in this context confirms the great potential of this technology in the understanding and analysis of movements, with practical applications in the training of martial arts athletes and the development of more sophisticated movement recognition systems.

Segmenting kicking movements using gyroscope and accelerometer data can be achieved through the following steps: Conduct an initial analysis of the data to recognize patterns in kicking movements. Utilize significant changes in angular velocity (from the gyroscope) and acceleration (from the accelerometer) as cues for the beginning and end of the movement. Establish thresholds for angular velocity and acceleration that indicate the initiation and conclusion of the kicking action. Choose a threshold value that distinguishes between motion and rest or other irrelevant movements. Apply the defined threshold to pinpoint segments of kicking motion. When the angular velocity or acceleration surpasses a threshold, it signifies that movement is underway. Carry out manual validation or human supervision on the identified motion segments to ensure their accuracy in representing the kicking motion. Conduct parameter optimization on machine learning models to enhance classification accuracy, particularly post-implementation of the segmentation process.

To increase the accuracy value of the k-NN algorithm, it is necessary to strive for things such as choosing the optimal k value, balancing the number of samples per very different class so that it is more balanced, wisely selecting features that can help reduce data dimensions and improve accuracy, and it is necessary to carry out exploratory analysis of the data to understand the characteristics of the dataset. Knowing whether there are important trends, patterns, or relationships can guide the k-NN model setup more effectively. By considering and combining several of these strategies, it can improve the performance of the k-NN algorithm and increase the accuracy value on classification tasks.

To achieve more accurate classification in human motion classification tasks, here are some suggestions regarding the efforts to be made and the algorithms that can be used. Some efforts that must be made are to ensure that the data collected from sensors (e.g., accelerometer, gyroscope) is of high quality and free of noise as much as possible. Use appropriate sensor devices and sensor calibration to ensure data accuracy.

At the data preprocessing stage, more effort needs to be made in data cleaning to remove noise and artifacts from sensor data, data normalization to ensure consistent scale across all features, segmentation and windowing into appropriate time windows to capture movement well, extraction of appropriate features relevant features such as time statistics, frequency features, and time-frequency domain-based features.

Improvement efforts can also be made by trying other available techniques, such as using feature selection techniques (for example, PCA, LDA) to reduce the dimensions of the data and retain the most significant features. To overcome limited datasets, data augmentation techniques can be used to generate more training examples, such as adding small noise, rotation, or scaling to the movement data. Cross-validation methods to prevent overfitting guarantee that the model performs effectively when applied to unknown data.

Several algorithms can be tried for classification techniques, such as Convolutional Neural Networks (CNNs), which have excellent advantages in capturing spatial and temporal features from movement sensor data. They can automatically extract features from raw data. Applications: CNNs are often used for human activity classification with time-series data and sensor signals. Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNNs) algorithms are designed to handle sequential and time-series data, so they are very suitable for temporal human movement classification. LSTMs are used to capture long-term dependencies in human motion data. By following the steps outlined and selecting a suitable algorithm, the accuracy of classifying an athlete's kick can be greatly enhanced.

VIII. Conclusion

This section provides an in-depth overview of the implementation of kick classification using IMU, k-NN, Random Forest and SVM machine learning algorithms. The success of these algorithms in this context confirms the great potential of this technology in the understanding and analysis of movements, with practical applications in the training of martial arts athletes and the development of more sophisticated movement recognition systems.

SVM, with its ability to handle non-linear relationships via kernel tricks, provides solid performance in kick classification. Selection of an appropriate kernel, such as the RBF (Radial Basis Function) kernel, can affect SVM performance. The advantages of SVM are especially visible when the kick motion pattern has high complexity and cannot be clearly separated by a straight line.

Implementation of the k-NN algorithm on a kick dataset that utilizes IMU data demonstrates success in identifying movement patterns with an accuracy of 0.924 for k=3 and 0.843 for k=5. This performance is slightly lower compared to the SVM algorithm, with an accuracy of 0.924. All the results mentioned above are based on the use of 3 sensors (RL, RS, RF). However, when only two sensors are utilized, the k-NN accuracy drops to 0.838 (RS, RF) and 0.881 (RL, RS), while the SVM accuracy decreases to 0.871 (RS, RF) and 0.967 (RL, RS). For the random forest model, the accuracy is 0.900 with three sensors (RL, RS, RF), while with two sensors, the accuracy is 0.905 (RS, RF) and 0.890 (RL, RS), which are still lower than the k-NN and SVM algorithms.

The primary advantage of k-Nearest Neighbors (k-NN) lies in their simplicity and effectiveness in handling classification problems. As a non-parametric, instance-based learning algorithm, k-NN does not require an explicit training phase, making it easy to implement and adaptable to various types of data. However, despite its straightforward nature, choosing an optimal value for k is crucial to ensuring model performance. A small k value can lead to overfitting, where the model becomes too sensitive to noise and fails to generalize well to new data. In contrast, a high k value may lead to underfitting, which results in the model oversimplifying decision boundaries and misclassifying complex patterns. To mitigate these risks, it is essential to use cross-validation techniques to determine the best k value for a given dataset. Additionally, data preprocessing steps such as feature scaling and noise reduction can significantly enhance the accuracy and reliability of k-NN in practical applications.

IX. Future Work

The current study indicates that it is possible to use IMU-based wearable systems and machine learning algorithms together to accurately categorize different forms of martial arts kicks. But there are several ways that the system could be improved in the future to make it more useful, practical, and easy to use for daily athletic training.

Future work is to reduce the number of sensors used in the system while maintaining comparable levels of classification accuracy. Reducing the number of sensors would also lower the cost and complexity, enabling broader adoption among athletes for everyday training without interfering with movement. If possible, reducing sensors to a compact form factor such as a single wearable smart device (e.g., a smartwatch or ankle-worn sensor) could significantly improve system convenience, ease of use, and real-world applicability.

Subsequent research will concentrate on enhancing machine learning and feature extraction methodologies to mitigate the limitations posed by reduced sensor inputs. Identifying optimal sensor sites (such as the most effective body parts), integrating data from several sensors, or utilizing sophisticated algorithms like deep learning that can discern intricate patterns from limited data.

Acknowledgment

The authors assert that there are no conflicts of interest pertaining to this research. This study was conducted independently without any financial, institutional, or commercial influence from any organization. All data collection, analysis, and interpretations were performed objectively to ensure the integrity and validity of the findings.

This research has received ethical clearance from the Ethics Commission of the Bandung Institute of Technology (ITB) under the approval KEP/II/2024/X/M281024RG-PDPG. The commission has reviewed and confirmed that this study complies with ethical standards and is deemed ethically feasible for implementation.

References

[1] Eline Van Der Kruk & Marco M. Reijne. (2018),"Accuracy of human motion capture systems for sport applications; state-of-the-art review", European Journal of Sport Science.

[2] Ferryanto, F., Mahyuddin, A. I., & Nakashima, M. (2022). Development of a markerless optical motion capture system by an action sports camera for running motion. ASEAN Engineering Journal, 12(2), 37–44. https://doi.org/10.11113/aej.v12.16760

[3] Vitali, Andrea & Regazzoni, Daniele & Rizzi, Caterina & Lupi, Giorgio. (2020). Low Cost Markerless Motion Capture Systems: A Comparison Between RGB Cameras and RGB-D Sensors. ASME 2020 International Mechanical Engineering Congress and Exposition 10.1115/IMECE2020-24083.

[4] Samet Ciklacandir, Sultan Ozkan, Yalcin Isler, (2022),"A Comparison of the Performances of Video-Based and IMU Sensor-Based Motion Capture Systems on Joint Angles", 2022 Innovations in Intelligent Systems and Applications Conference (ASYU).

[5] Colyer, S. L., Evans, M., Cosker, D., & Salo, A. I. T. (2018). A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. Sports Medicine - Open, 4(1), 1–15. https://doi.org/10.1186/S40798-018-0139-Y

[6] Yang, P., Chen, Y., Chen, Y., & Dang, F. (2021). Gyro-Free Inertial Measurement Unit With Unfettered Accelerometer Array Distribution and for the Object With Position Change in Center of Gravity. IEEE Sensors Journal, 21(7), 9423–9435.

[7] Yi-Lian Chen, I-Ju Yang, Li-Chen Fu, Jin-Shin Lai, Huey-Wen Liang, (2021) "IMU-Based Estimation of Lower Limb Motion Trajectory with Graph Convolution Network", IEEE Sensors Journal, Vol. 21, No. 21, November 1, 2021.

[8] Luigi Truppa, Elena Bergamini, Pietro Garofalo, Marta Costantini, Carmelo Fiorentino, Giuseppe Vannozzi, Angelo Maria Sabatini, and Andrea Mannini, (2021). "An Innovative Sensor Fusion Algorithm for Motion Tracking with On-Line Bias Compensation: Application to Joint Angles Estimation in Yoga", IEEE Sensors Journal, Vol. 21, No. 19, October 1, 2021.

[9] N. Dawar and N. Kehtarnavaz, (2018) “Action detection and recognition in continuous action streams by deep learning-based sensing fusion,” IEEE Sensors Journal, vol. 18, no. 23, pp. 9660–9668.

[10] Luke Wicent Sy, Nigel H. Lovell, and Stephen J. Redmond.,(2021) "Estimating Lower Body Kinematics Using a Lie Group Constrained Extended Kalman Filter and Reduced IMU Count". IEEE Sensors Journal, Vol. 21, No. 18, September 15, 2021.

[11] Kim, K.N., Chen, J., and Cho, Y. (2019). "Evaluation of Machine Learning Algorithms for Worker's Motion Recognition using Motion Sensors." Proceedings of the ASCE 2019 International Conference on Computing in Civil Engineering (i3CE), Atlanta, GA, USA, June 17-19.

[12] Akhavian, R., and Behzadan, A. H. (2016). “Productivity Analysis of Construction Worker Activities Using Smartphone Sensors.” Proceedings of 16th International Conference on Computing in Civil and Building Engineering, 1067–1074.

[13] Tsubasa Maruyama, Mitsunori Tada, Akira Sawatome, Yui Endo (2018). "Constraint-based Real-Time Full-Body Motion-Capture using Inertial Measurement Units". IEEE International Conference on Systems, Man, and Cybernetics.

[14] Cecilia Monoli, Juan Francisco Fuentez-Pérez, Nicola Cau, Paolo Capodaglio, Manuela Galli, and Jeffrey A. Tuhtan, (2021) "Land and Underwater Gait Analysis Using Wearable IMU",IEEE sensors journal, vol. 21, no. 9, may 1, 2021.

[15] J. Lim, S. Lee, and H. G. Jung, “Action Recognition of Taekwondo Unit Actions Using Action Images Constructed with Time-Warped Motion Profiles,” Apr. 2024, doi: 10.3390/s24082595.