Punch Force Classification using K Means and a Data Logging System

Lilik Anifah
Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
lilikanifah@unesa.ac.id (corresponding author)

Nurhayati
Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
nurhayati@unesa.ac.id

Puput Wanarti Rusimamto
Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
puputwanarti@unesa.ac.id

Muhamad Syarifuddien Zuhrie
Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
zuhrie@unesa.ac.id

Haryanto
Department of Electrical Engineering, Faculty of Engineering, Universitas Trunojoyo Madura, Indonesia
haryanto@trunojoyo.ac.id

Received: 18 October 2024 | Revised: 5 November 2024 | Accepted: 14 November 2024

ABSTRACT

Much research has been conducted worldwide on the recognition and monitoring of punches in martial art sports during the training process. The performance of the punching movements can be accurately analyzed based on the collected data. The current study aims to classify the punches on a punching bag using K Means based on a data logging system. Its stages are hardware design, hardware implementation, hardware testing, learning, and testing. The FSR 402 sensor was used to measure the punching force. GY6500 MPU6500 was also utilized to identify and measure the reaction force of these punches. The data were collected utilizing K Means and were subsequently tested. The results revealed that the system exhibited good performance, proven by its accuracy of 93.6%, precision of 0.933, recall of 0.934, and F1 score of 0.934. Based on these results, it seems that the classification of right and left punch forces can be efficiently carried out. This system can help users analyze the punching bag training process, and thus improve their performance.

Keywords-k means; punching; data logger; classification

I. INTRODUCTION

Punch recognition when utilizing a punching bag is an important study subject in development of martial arts. The punching bag is a useful tool in the athletes' training process, while the employment of technology can help improve their performance. Punching bag technology, including the use of impact sensors and accelerometers that capture the physical characteristics of punches, has been introduced by several researchers [1]. A punching bag system can provide feedback via LEDs, speakers, and screens [2]. Other studies have utilized strain gauges and accelerometers [3], as well as an imaging approach, which deploys Support Vector Machines (SVMs) [4]. Technology can comprehensively analyze force, speed, and punching technique as it recognizes punching bag punches during the training process [5, 6]. Other studies have integrated Machine Learning (ML), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) in punch recognition [7]. However, identifying movements and styles from punches requires data that are the source of the learning and testing process, therefore an investigation of loggers for these data to be collected is much needed. Data loggers are instruments used to monitor and record changes in a system over a period. Most single-channel units are battery-operated, allowing them to record on the go and for long time periods. Data loggers are equipped with sensors, tailored to the data needed, and are then connected to a microprocessor that processes these data. They can be also equipped with internal memory for data storage. Data loggers can be connected to computers and other equipment for viewing and analyzing the obtained data [8].

The data logger can be defined as a device consisting of a microcontroller with an Analog-to-Digital Converter (ADC) that is usually employed to measure and store data in various applications. Data loggers can be stored on SD cards, disks, or internal memory [9], while they are adjusted to the needs of their users [10]. Modern data loggers have high storage capacity, high performance, and are easy to use. Their advantages involve small dimensions, remote-control capabilities, and wireless or cloud services [11-13]. Data loggers have been applied in various fields, e.g. to monitor the outdoor and indoor temperatures of a building [14], for the maintenance and monitoring of vehicle parameters [15], while they can record analog sensor data at 9600 bps with a 10-bit resolution [16]. In the field of power generation, a data logger can be utilized as a wind speed, wind direction, humidity, and temperature meter [12].

The current research was conducted using sensors adjusted to the variables sought and which have been proven useful for developing punching recognition. The originality of this study lies in the sensor utilization to calculate the force of the punch, which is analyzed based on the action and reaction of the latter. The punch action represents the punch carried out by the athlete, which is indicated by the presence of a signal sent from the FSR 402 sensor implanted in the athlete's gloves. These gloves are specifically manufactured for punching on the punching bag. The reaction of the punch is the impact of the punch performed by the athlete on the punching bag. The punching bag will move, and this movement will be detected by the GY6500 MPU6500 sensor implanted in the punching bag. With the data from this sensor, the movement (punch action) will be also detected. The results of the data logger are then used as learning data and the resulting weight is utilized as testing data to categorize the strength of the punch. The contribution of this study is that the data logger is anticipated to capture the athletes' punch data so that training analysis can be carried out based on them.

II. METHODS

The stages of this research are hardware design, hardware implementation, testing, learning, testing, and result analysis. The hardware design of this research is depicted in Figure 1, where it is demonstrated that the FSR 402 sensor is implanted on the right- and left-hand punching glove. This sensor is capable of capturing the punching force signal from both the right and left hands. FSR 402 sensors are connected to Wemos. The GY6500 MPU6500 sensor is implanted on the punching bag to identify movements on the latter. Furthermore, the data can be sent via the Internet and can be accessed by the user.

The Force Sensing Resistor is a Polymer Thick Film (PTF) device that shows a decrease in resistance when there is a force that hits the active surface [17, 18]. This sensor has been widely implemented in various fields [19-21]. GY6500 MPU6500 is a sensor that can identify direction, calculating the degrees of the punch angle. This sensor has been used for various purposes, namely to calculate the imbalance of the step length of the right and left legs when walking [22]. The scenario of this data logger system is depicted in Figure 2. When the user hits the punching bag using a punching glove, the FSR 402 implanted in the punching glove will receive a signal indicating the magnitude of the punching force applied to the punching bag. Then the signal indicating the magnitude of the punching force will be sent via the internet. The GY 6500 MPU6500 sensor will send a signal of the movement that occurs due to the punch. These three signals will be subsequently recorded and can be analyzed to determine the extent of the training performance.

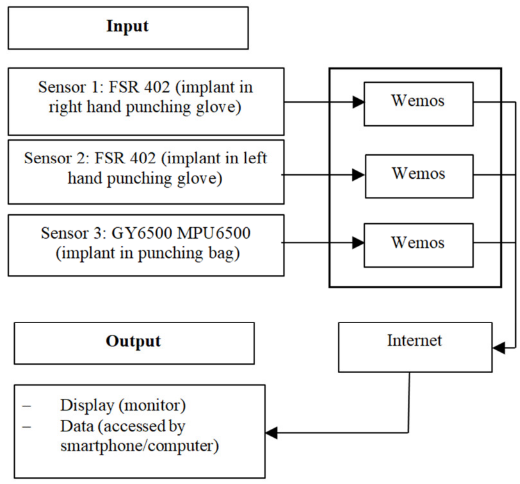

Fig. 1. Electronics design.

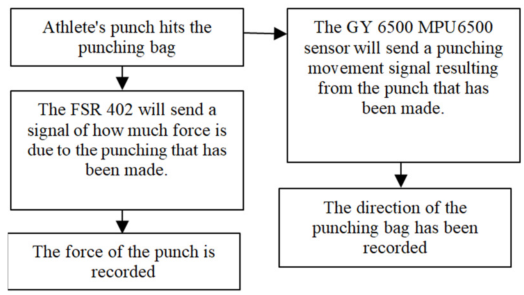

Fig. 2. Scenario of data logger system.

The data generated by the data logger are analyzed and divided into learning and testing data. These data will be processed and formulated into an 1×5 matrix consisting of signal information on the right and left gloves, the force sent by the sensor on the right and left gloves, and the position shift of the punching bag. This matrix is then used as input. The classification carried out comprises 8 clusters:

- Cluster 1: right-hand low force punches.
- Cluster 2: right-hand medium force punches.
- Cluster 3: right-hand high force punches.

- Cluster 4: right-hand punches performed with very high force.
- Cluster 5: left-hand low force punches.
- Cluster 6: left-hand medium force punches.
- Cluster 7: left-hand high force punches.
- Cluster 8: left-hand punches performed with very high force.

The research used K Means, one of the most popular Artificial Intelligence (AI) methods, due to its simplicity, low computational complexity, and good performance [23]. K Means has been modified with a computational geometry approach [24, 25], while it can be applied to large data conditions [25]. K Means has been implemented in several fields, such as the health and traffic sectors [26, 27]. The algorithmic learning steps of the K Means are [28]:

- Determine the number of clusters (k). In this study the number of clusters is 8 (k=8).
- Determine the point of each cluster that becomes the initial centroid (cluster center). Given that in this study there are 8 clusters, there are equivalently 8 initial cluster centers.
- The grouping of all data points is carried out according to the distance of the nearest centroid that has been determined by distance calculations, such as the Euclidean distance formula:

$$d = \arc\min\left(\sum_{x=1}^{x=i}(x_i - c_i)\right) \quad (1)$$

- Repeat the previous step until the specified number of epochs is reached.

The testing process is carried out by analyzing the performance of each sensor, namely the FSR 402 sensors implanted in the right and left gloves. Data extraction is carried out, and the data are marked. The results of this process are subsequently utilized as data for the learning process employing K Means until the weights used for the testing process are obtained. The analysis is carried out by calculating the accuracy, precision, recall, and F1 score from the testing results. Accuracy, precision, recall, and F1 score are given by [29]:

$$acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1\ score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (5)$$

where TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

III. RESULTS AND DISCUSSION

The implementation results are depicted in Figures 3 and 4. The FSR 402 is implanted in a glove, as can be seen in Figure 3. When punching, the FSR will send a signal that can indicate the frequency of force hitting the punching bag.

Fig. 3. Glove with FSR 402 implanted inside.

The punching bag used in this study is displayed in Figure 4. Its top is glued with the GY 6500 sensor. The MPU6500 will send a signal of the movement that occurs due to the blow imposed on the punching bag. This sensor will send a position on the x and y axes (the movement in the z axis is not allowed, causing the punching bag to only move in the x and y directions). After obtaining the x and y positions, the movement is calculated by:

$$r = \sqrt{x_i^2 - y_i^2} \quad (6)$$

Fig. 4. Punching bag with GY 6500 MPU6500 implanted on its top.

An analysis of the FSR 402 sensor utilization has been carried out and the results obtained are detailed in Figure 5. When compared with the reference sample, a difference with a standard deviation of 0.432 N is observed. This difference is shown in Figure 5, where the blue line is the result of the system, while the orange line is the result of the reference sample.

The extraction feature process begins with the data collection. The classification consists of 8 clusters. Data are extracted with the purpose of forming a matrix which will be used in the training process by deploying K Means. After data acquisition, an 1×5 matrix is obtained, consisting of signal information on the right and left gloves, the force sent by the sensor on the right and left gloves, and the position shift of the punching bag. The data are divided into 2 parts, namely the data used for the learning process and those employed for the testing process. The learning process utilizes 25 data for each cluster, so the total number of the learning samples is 200. The testing data samples are 1200, comprising 150 samples for each cluster. The learning process is carried out using K Means with a learning rate of 0.1, and for 1000 epochs. The former produces weights for each cluster, which are utilized for the testing process. The latter is carried out twice, as it comprises the testing process using learning data and the testing process utilizing testing data. The obtained confusion matrix, as described in Table I, represents an accuracy of 95%. Clusters 1-4 show the punches performed with the right hand, and clusters 5-6 display the punches performed with the left hand. In Table I, it is explained that for clusters 1 and 5, the system had been able to classify data correctly. However, for cluster 2, there is a data misclassification, where 1 data sample is classified as belonging to cluster 1. This also occurs in clusters 3 and 4, where a misclassification is, respectively, observed.

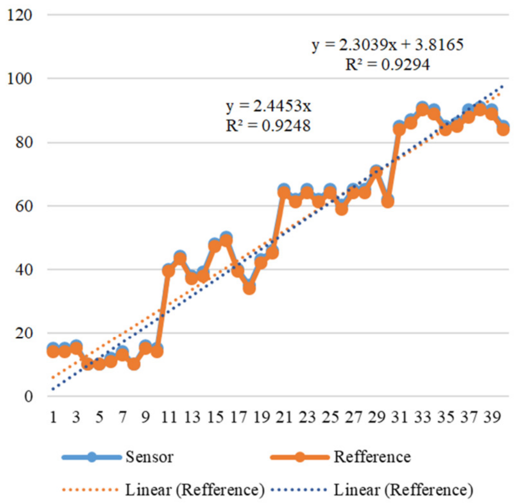

Fig. 5. Results of FSR 402 sensor analysis.

TABLE I. CONFUSION MATRIX OF THE FIRST TESTING PROCESS (LEARNING DATA)

	C1	C2	C3	C4	C5	C6	C7	C8
C1	25	0	0	0	0	0	0	0
C2	1	24	0	0	0	0	0	0
C3	0	0	24	1	0	0	0	0
C4	0	0	1	24	6	0	0	0
C5	0	0	0	0	25	0	0	0
C6	0	0	0	0	0	24	1	0
C7	0	0	0	0	0	1	22	2
C8	0	0	0	0	0	0	3	22

Clusters 5-8 exhibit the classification of punches performed with the left hand. In cluster 5, the system had been able to classify punches very well. In cluster 6, however, there is an error, because the results of the distance calculation are close to the ones in cluster 7. Regarding clusters 7 and 8, there are 3 classification errors. Such errors occur due to the punching force produced between the two classes, so, the results of the Euclidean distance calculation allow it to slip in the cluster before and after classification. The performance results of the testing process using learning data are outlined in Table II. This first testing process is useful to find out to what extent the data being learned can be recognized. According to the results, this process obtained a precision value of 0.927, a recall value of 0.926, and an F1 score of 0.927.

TABLE II. PERFORMANCE OF FIRST TESTING PROCESS (LEARNING DATA)

Performance	Value	Performance	Value
Precision C1	1.000	Recall/sensitivity 1	0.962
Precision C2	0.960	Recall/sensitivity 2	1.000
Precision C3	0.960	Recall/sensitivity 3	0.960
Precision C4	0.774	Recall/sensitivity 4	0.960
Precision C5	1.000	Recall/sensitivity 5	0.806
Precision C6	0.960	Recall/sensitivity 6	0.960
Precision C7	0.880	Recall/sensitivity 7	0.846
Precision C8	0.880	Recall/sensitivity 8	0.917
Precision	0.927	Recall/sensitivity	0.926
Accuracy	0.950	F-1 Score	0.927

The confusion matrix from the testing process using testing data is shown in Table III. In this process, the system accuracy is 93.6%. It can be seen that the inter-class feature exhibits a type of transition, e.g. in clusters 3 and 4.

TABLE III. CONFUSION MATRIX OF THE SECOND TESTING PROCESS (TESTING DATA)

	C1	C2	C3	C4	C5	C6	C7	C8
C1	142	8	0	0	0	0	0	0
C2	7	139	4	0	0	0	0	0
C3	0	2	144	4	0	0	0	0
C4	0	0	10	140	0	0	0	0
C5	0	0	0	0	146	4	0	0
C6	0	0	0	0	3	147	0	0
C7	0	0	0	0	0	0	130	20
C8	0	0	0	0	0	0	15	135

TABLE IV. PERFORMANCE OF SECOND TESTING PROCESS (TESTING DATA)

Performance	Value	Performance	Value
Precision C1	0.947	Recall/sensitivity 1	0.953
Precision C2	0.927	Recall/sensitivity 2	0.933
Precision C3	0.960	Recall/sensitivity 3	0.911
Precision C4	0.933	Recall/sensitivity 4	0.972
Precision C5	0.973	Recall/sensitivity 5	0.980
Precision C6	0.961	Recall/sensitivity 6	0.974
Precision C7	0.867	Recall/sensitivity 7	0.878
Precision C8	0.900	Recall/sensitivity 8	0.871
Precision	0.933	Recall/sensitivity	0.934
Accuracy	0.936	F-1 Score	0.934

The system performance in the testing process is described in Table IV. The acquired precision is 0.933. The system precision result is obtained from the average of the precision values of each cluster. The highest precision value is obtained from cluster 6, and the lowest from cluster 7. The average system recall value is 0.934. The F1 score is 0.934, exhibiting that the system had been able to classify the punching force very well.

The observed failure in classification emerges because the features of clusters 3 and 4 are close and partially overlap. This allows some data that should be in cluster 3 to be classified as being in cluster 4, and vice versa, as portrayed in Figure 6. The distance between the enter points of clusters 3 and 4 is close and some data overlap laying the foundation for misclassifications.

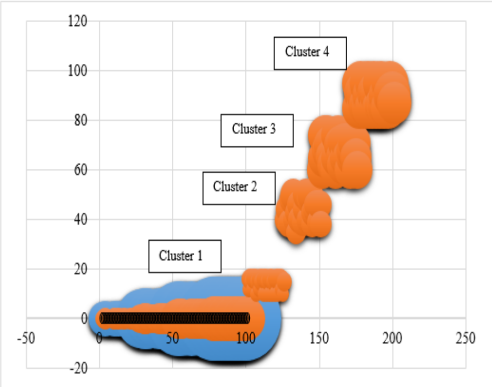

Fig. 6. Illustration of the analysis of the results of the right-hand punch test (clusters 1-4).

The left-hand stroke, based on the confusion matrix in Table III, shows that there is a misclassification in clusters 5 and 6. There are 4 data classified as being in cluster 6 even though they should be classified as being in cluster 5. After a further analysis, it turns out that this is because some data from clusters 5 and 6 overlap, as depicted in Figure 7. This overlap position is responsible for the wrong results of the Euclidean distance calculation and must be separated from the neighboring center.

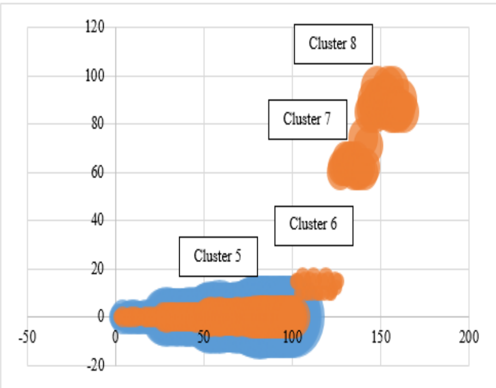

Fig. 7. Illustration of the analysis of the results of the left-hand punch test (clusters 5-8).

Cluster 6 and 7 data positions do not overlap. Thus, the classification error condition between the two classes does not occur. For clusters 7 and 8 there is a higher overlap, where there are 20 data samples classified as being in cluster 8 (and should be classified as being in cluster 7) and 15 samples classified as being in cluster 7 (and should be classified as being in cluster 8). Figure 7 displays that clusters 7 and 8 do have some data proximity. This is what causes the 20 samples from the Euclidean distance calculation to be closer to the center point of cluster 8 than the center point of cluster 7. This phenomenon also occurs in 15 other samples. This issue can be found in several other studies, possibly due to the final calculation results of the distance using the Euclidean distance [30]. So, one of the challenges of using K Means is to produce optimal centroids in each cluster [31]. Another challenge concerns how to label the initial data so that they can improve system performance [32].

The current research was conducted to facilitate punching bag users, who have so far been limited to punching exercises without knowing their progress in these exercises. The use of the proposed technology is expected to meet their needs, offering an alternative solution. In previous research, punching bag users could only hit at a certain point of the punching bag [1-3]. In this study, though, users have been able to hit dynamically anywhere on the punching bag. However, this study still has some weaknesses, as it is not camera-based, and the punching force produced must not be more than 100 N. Another difference with previous studies is that they employed SVMs [4], CNNs, and LSTM [7], while this research applies K Means based on a data logger. The advantage of this system is that it has been equipped with a data logger, so, users can see their progress by accessing it.

IV. CONCLUSIONS

The purpose of this study is to classify the punches on a punching bag using K Means based on a data logging system. The data were recorded properly in the data logger system, and were subjected to feature extraction. The results were subsequently subjected to a learning process, obtaining the weights which were later used in the testing process. The testing process results are 93.6% accuracy, 0.933 precision, 0.934 recall, and 0.934 F1 score. Based on these findings, it seems that the classification of the right and left punch forces can be efficiently performed. For now, the maximum force that can be captured is 100 N, meaning that the current research cannot meet the needs of professional athletes. It is, therefore, necessary to conduct experiments on the proposed sensors to accommodate higher punching forces. In addition, other Artificial Intelligence (AI) methods can be deployed to identify and classify punches. The contribution of this research is that the utilized data logger is able to capture the force data of the punches performed by the athletes. Thus, training analysis process can be carried out based on these data. If the magnitude of the punching force data is known, users can analyze it and improve their performance. The novelty of this research is the implementation of K Means in analyzing the force of the athlete's punch. Another novelty is that the punch can be dynamically analyzed because the sensor is implanted in the glove.

ACKNOWLEDGMENT

The authors would like to thank the Directorate of Research and Community Service of the Ministry of Education and Culture of the Republic of Indonesia for providing funding, and the Universitas Negeri Surabaya for fully supporting this research.

REFERENCES

[1] A. Yadav and N. Sakle, "Development of low-cost data logger system for capturing transmission parameters of two-wheeler using Arduino," *Materials Today: Proceedings*, vol. 72, pp. 1697–1703, Jan. 2023, https://doi.org/10.1016/j.matpr.2022.09.471.

[2] V. Essien, C. A. Bolu, J. Azeta, I. P. Okokpujie, O. Kilanko, and S. A. Afolalu, "Application of Data Logger for Monitoring Indoor and Outdoor Temperature of Buildings: A Review," *IOP Conference Series: Materials Science and Engineering*, vol. 1107, no. 1, Dec. 2021, Art. no. 012181, https://doi.org/10.1088/1757-899X/1107/1/012181.

[3] P. C. Wang, C. M. Uang, and Z. S. Ho, "A Low Cost Data Logger for Industry Applications," *Applied Mechanics and Materials*, vol. 389, pp. 513–518, 2013, https://doi.org/10.4028/www.scientific.net/AMM.389.513.

[4] J. A. V. P. Raja, J. A. B. Raj, M. Bharath, P. C. K. Kumar, and V. Rajendran, "Multichannel Data Logger for Combat Vehicle," *International Journal of Engineering & Technology*, vol. 7, no. 3.6, pp. 410–414, Jul. 2018, https://doi.org/10.14419/ijet.v7i3.6.17482.

[5] S. Engelberg, T. Kaminsky, and M. Horesh, "instrumentation notes - A USB-Enabled, FLASH-Disk-Based Data Logger," *IEEE Instrumentation & Measurement Magazine*, vol. 10, no. 2, pp. 63–66, Apr. 2007, https://doi.org/10.1109/MIM.2007.364976.

[6] D. Ringgau, M. Anyi, and K. Hong Ping, "Design of a Data Logging System for a Small Scale Hydrokinetic Turbine - ASM Science Journal," *ASM Science Journal*, vol. 14, pp. 1–8, Jun. 2021, https://doi.org/10.32802/asmscj.2020.649.

[7] Z. W. Siew, C. H. Wong, S. E. Tan, H. P. Yoong, and K. T. Kin Teo, "Design and development of a tablet based real time wireless data logger," in *2012 IEEE Global High Tech Congress on Electronics*, Aug. 2012, pp. 111–116, https://doi.org/10.1109/GHTCE.2012.6490135.

[8] S. Agrawal, V. Kumar, N. Anand, V. K. Agarwal, and A. Islam, "Development of data acquisition system and data analysis technique for automotive applications," in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, Oct. 2016, pp. 1–4, https://doi.org/10.1109/STARTUP.2016.7583954.

[9] G. Sanasie, Z. I. A. Radzak, and M. A. Zakaria, "Preventive Maintenance Data Logger Monitoring System," *Mekatronika: Journal of Intelligent Manufacturing and Mechatronics*, vol. 3, no. 1, pp. 68–78, Jun. 2021, https://doi.org/10.15282/mekatronika.v3i1.7154.

[10] A. T. Andreola, M. Jinbo, D. Senter, E. L. Todero, and G. Cardoso, "Low cost data acquisition system for wind prospecting," in *2016 12th IEEE International Conference on Industry Applications (INDUSCON)*, Aug. 2016, pp. 1–6, https://doi.org/10.1109/INDUSCON.2016.7874463.

[11] M. Petre *et al.*, "The interactive punching bag," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, Feb. 2012, pp. 1469–1470, https://doi.org/10.1145/2212776.2212490.

[12] S. A. Zaidi and V. Chouvatut, "Mae Mai Muay Thai Layered Classification Using CNN and LSTM Models," in *2022 26th International Computer Science and Engineering Conference (ICSEC)*, Sep. 2022, pp. 351–356, https://doi.org/10.1109/ICSEC56337.2022.10049339.

[13] D. Dinu, B. Millot, J. Slawinski, and J. Louis, "An Examination of the Biomechanics of the Cross, Hook and Uppercut between Two Elite Boxing Groups," *Proceedings*, vol. 49, no. 1, 2020, Art. no. 61, https://doi.org/10.3390/proceedings2020049061.

[14] S. Kasiri, C. Fookes, S. Sridharan, and S. Morgan, "Fine-grained action recognition of boxing punches from depth imagery," *Computer Vision and Image Understanding*, vol. 159, pp. 143–153, Jun. 2017, https://doi.org/10.1016/j.cviu.2017.04.007.

[15] E. Arseneau, S. Mekary, and L. A. Léger, "V̇o2 Requirements of Boxing Exercises," *The Journal of Strength & Conditioning Research*, vol. 25, no. 2, Feb. 2011, Art. no. 348, https://doi.org/10.1519/JSC.0b013e3181ef64cb.

[16] D. Dinu, B. Millot, J. Slawinski, and J. Louis, "An Examination of the Biomechanics of the Cross, Hook and Uppercut between Two Elite Boxing Groups," *Proceedings*, vol. 49, no. 1, 2020, Art. no. 61, https://doi.org/10.3390/proceedings2020049061.

[17] A. Barleanu and M. Hulea, "Neuromorphic Sensor Based on Force-Sensing Resistors," *Biomimetics*, vol. 9, no. 6, Jun. 2024, Art. no. 326, https://doi.org/10.3390/biomimetics9060326.

[18] A. K. Stavrakis, M. Simić, M. Damnjanović, and G. M. Stojanović, "A fully textile-based, tunable, force and strain sensing resistor for e-textile applications," *Heliyon*, vol. 10, no. 4, Feb. 2024, Art. no. e26069, https://doi.org/10.1016/j.heliyon.2024.e26069.

[19] O. Sherif, M. M. Bassuoni, and O. Mehrez, "A survey on the state of the art of force myography technique (FMG): analysis and assessment," *Medical & Biological Engineering & Computing*, vol. 62, no. 5, pp. 1313–1332, May 2024, https://doi.org/10.1007/s11517-024-03019-w.

[20] H. U. R. Siddiqui *et al.*, "Footwear-integrated force sensing resistor sensors: A machine learning approach for categorizing lower limb disorders," *Engineering Applications of Artificial Intelligence*, vol. 127, Jan. 2024, Art. no. 107205, https://doi.org/10.1016/j.engappai.2023.107205.

[21] L. Anifah, M. S. Zuhrie, Muhammad, and Haryanto, "Integrated Smart Real Time Scoring Pencak Silat Based on Internet of Things (IOT)," *International Journal on "Technical and Physical Problems of Engineering" (IJTPE)*, vol. 15, no. 54, pp. 155–163, Mar. 2023.

[22] I. A. E. Zaeni, W. Primadi, M. K. Osman, D. R. Anzani, D. Lestari, and A. N. Handayani, "Detection of the Imbalance Step Length using the Decision Tree," in *2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Sep. 2022, pp. 157–162, https://doi.org/10.1109/ICVEE57061.2022.9930456.

[23] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023, https://doi.org/10.1016/j.ins.2022.11.139.

[24] T. Kr. Biswas, K. Giri, and S. Roy, "ECKM: An improved K-means clustering based on computational geometry," *Expert Systems with Applications*, vol. 212, Feb. 2023, Art. no. 118862, https://doi.org/10.1016/j.eswa.2022.118862.

[25] M. Belhor, A. El-Amraoui, A. Jemai, and F. Delmotte, "Multi-objective evolutionary approach based on K-means clustering for home health care routing and scheduling problem," *Expert Systems with Applications*, vol. 213, Mar. 2023, Art. no. 119035, https://doi.org/10.1016/j.eswa.2022.119035.

[26] S. F. Hussain and M. Haris, "A *k*-means based co-clustering (*k*CC) algorithm for sparse, high dimensional data," *Expert Systems with Applications*, vol. 118, pp. 20–34, Mar. 2019, https://doi.org/10.1016/j.eswa.2018.09.006.

[27] W.-C. Liu and C.-H. Lin, "A hierarchical license plate recognition system using supervised K-means and Support Vector Machine," in *2017 International Conference on Applied System Innovation (ICASI)*, Feb. 2017, pp. 1622–1625, https://doi.org/10.1109/ICASI.2017.7988244.

[28] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, Aug. 2020, Art. no. 1295, https://doi.org/10.3390/electronics9081295.

[29] D. J. Hand, P. Christen, and N. Kirielle, "F*: an interpretable transformation of the F-measure," *Machine Learning*, vol. 110, no. 3, pp. 451–456, Mar. 2021, https://doi.org/10.1007/s10994-021-05964-1.

[30] M. Meilă, "Good (K-means) clusterings are unique (up to small perturbations)," *Journal of Multivariate Analysis*, vol. 173, pp. 1–17, Sep. 2019, https://doi.org/10.1016/j.jmva.2018.12.008.

[31] Md. Zubair, MD. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Annals of Data Science*, vol. 11, no. 5, pp. 1525–1544, Oct. 2024, https://doi.org/10.1007/s40745-022-00428-2.

[32] A. Vouros and E. Vasilaki, "A semi-supervised sparse K-Means algorithm," *Pattern Recognition Letters*, vol. 142, pp. 65–71, Feb. 2021, https://doi.org/10.1016/j.patrec.2020.11.015.