Classification of Football Player Actions Using Sensing Data

2024 International Conference on Image Processing and Robotics (ICIPRoB) | 979-8-3503-7476-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIPRoB62548.2024.10544276

Yukito Hirasawa
Department of Computer Science and Engineering
University of Yamanashi
Yamanashi, Japan
g22tk017@yamanashi.ac.jp

Yuichiro Kinoshita
Department of Computer Science and Engineering
University of Yamanashi
Yamanashi, Japan
0000-0002-2206-8976

Latifah Munirah Kamarudin
Faculty of Electronic Engineering and Technology
Universiti Malaysia Perlis
Perlis, Malaysia
0000-0002-2547-3934

Ammar Zakaria
Faculty of Electronic Engineering and Technology
Universiti Malaysia Perlis
Perlis, Malaysia
0000-0002-7108-215X

***Abstract*—This study classifies the actions of football players using sensing data acquired from wearable sensors attached to players and the ball. More than 800 sensing data with the labels of five types of player actions were created as a dataset. The neural networks were trained using 19 input items created by considering time-series variations in player and ball locations. The trained neural network model demonstrated a classification accuracy of 84.0 %. The model successfully obtained sufficient accuracy for all types of actions.These results demonstrate that the sensing data and created input items can be effectively utilized for classifying the actions of football player.**

***Index Terms*—football, wearable sensor, action classification, neural network**

I. Introduction

Football is a globally popular sport with numerous matches hosted every year. Owing to the challenges related to observing every match, the summaries of player evaluations and key events are essential. For generating statistics such as shots, passes, and goals, professional scorekeepers meticulously review match videos; this task consumes approximately 10 hours per game and results in significant workload [1].

To mitigate this burden and streamline match analysis, various methods have been proposed [2]–[4], and notably, video-based techniques [5]–[8] have gained prominence. Sen et al. [5] utilized a pre-trained convolutional neural network to extract features from video clips of player events such as fouls and corner kicks, classifying these events into 10 categories using long short-term memory. Similarly, Muhammad et al. [6] categorized video clips into six events. However, these studies identified challenges, including player occlusion and video brightness variability. An emerging solution involves enhancing video analysis with sensing data from sensors attached to players and the ball, as suggested in [1], [8], [9].

The recent trend of equipping players and balls with sensors, exemplified by the sensor-embedded balls used in the FIFA World Cup Qatar 2022, has gained acceptance in professional football leagues. These trackers extend beyond sensor-embedded balls to wearable player sensors. Thus, the accumulation of extensive sensing data from players and balls during matches is anticipated, offering valuable insights for match analysis. Compared to video-based analyses, studies employing sensing data in football are less prevalent. Most existing research focuses on classifying match events such as fouls and corner kicks. In contrast, few studies prioritize the classification of player actions such as passes and dribbles. The ability to categorize a wide array of player actions would facilitate the creation of more nuanced player evaluations and highlight reels.

This study aimed to classify football player actions using sensing data. Sensing data facilitates the precise acquisition of numerical coordinates for both players and the ball. Consequently, a straightforward classifier based on sensing data could achieve classification accuracy for football player actions comparable to, or even surpassing, that of video-based methods. Two-dimensional sensing data of players and the ball are acquired using a local positioning system, with player actions manually labeled through match video analysis. This labeled sensing dataset serves as the foundation for training a neural network to classify the actions of football players.

The primary contribution of this paper is the introduction of a methodology for classifying football player actions using a neural network with a simplistic architecture, in conjunction with sensing data. Moreover, this research represents an initial step toward demonstrating the efficacy of sensing data within a multimodal learning framework that integrates both video and sensing data.

II. Related works

Research in football utilizing sensing data from players and the ball typically involves investigators independently defining features and executing the detection and classification of football events and player actions. Vidal-Codina et al. [10] developed a method for detecting and classifying events and

TABLE I
DEFINED ACTIONS

Name	Definition
Short pass	A player aims to pass the ball to a teammate within a distance under 30 meters, excluding cross, goalkeeper throw, goalkeeper punt kick, and throw in.
Long pass	A player aims to pass the ball to a teammate over a distance of 30 meters or more, also excluding cross, goalkeeper throw, goalkeeper punt kick, and throw in.
Dribble	A player maneuvers the ball with the intent of attacking the opponent's goal.
Cut	A player intercepts an opponent's pass or takes the ball during a dribble, aiming to regain possession for the team.
Trap	A player halts an incoming ball and controls it to facilitate the subsequent action.

actions in football matches using two-dimensional sensing data of players and the ball, employing a rule-based framework. In this approach, events in football matches were identified when a player was in close proximity to the ball. The detected actions were categorized into four labels: pass, shoot, cross, and none, using classifiers based on predefined rules. Despite the classification precision ranging from 86–90 %, the recall rates for shoot and cross were only 52 % and 39 %, respectively, with a predominant classification of actions as passes. These results indicate the complexity of simultaneously performing event and action classification using a singular classifier, suggesting the potential for enhancing classification precision through specialized systems dedicated to classifying football player actions.

Imai et al. [1] detected football player actions by analyzing changes in ball trajectory and derived features from sensing data for action classification using a random forest. The action classification included 10 labels such as pass and shoot. The F-scores for pass and trap were 86.0 % and 85.9 %, respectively, whereas the scores for shoot and dribble were lower at 44.4 % and 30.7 %, respectively, highlighting a variance in classification efficacy based on action type.

Previous studies have predominantly relied on data from brief instances (∼0.5 s before and after a player contacts the ball), without considering time-series changes in player and ball locations. To enhance the classification accuracy for a broader range of football player actions, this study expands the data scope to longer time periods and incorporates the dynamics of time-series changes in player and ball locations.

III. DATA COLLECTION

A. Equipment

This study aims to facilitate multimodal learning through the integration of match video and sensing data from players and a ball. Consequently, video data was acquired along with sensing data. The sensing data were recorded using a local positioning system kit (Gengee, Insait Pro K1), comprising wearable sensors affixed to each player's left arm, a sensor embedded within the ball, and six anchors strategically positioned around the pitch. Four anchors were placed 2 m from each corner of the pitch, while the remaining two aligned with the extension of the centerline. The collected sensing data included two-dimensional coordinate information, with the horizontal and vertical directions of the pitch represented by x and y axes, respectively. These data were recorded with a precision of 1 cm at a frequency of 20 Hz.

TABLE II
NUMBER OF LABELED ACTIONS

Name	Match1	Match2	Total
Short pass	584	749	1333
Long pass	31	73	104
Dribble	80	135	215
Cut	46	78	124
Trap	461	619	1080

The video was captured using a combination of a video camera (Sony, FDR-AX43A) and an automatic tracking camera (Move'n See, PIX4TEAM), which tracks the estimated ball location based on player positions. Owing to occasional failures of the automatic tracking camera in capturing the ball, the sensing data from the embedded ball sensor was utilized to ascertain the location of the ball during matches.

The sensing data and video were collected from two matches in Malaysia, with recording durations of 80 and 100 min, respectively.

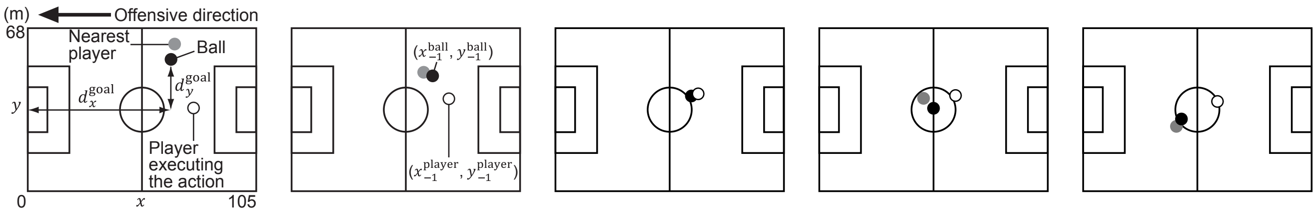
B. Labeling of actions using match videos

Utilizing the acquired match videos, the sensing data corresponding to each action observed in the two matches were labeled. For these matches, five distinct actions were identified for labeling, as listed in Table I. These actions were defined based on criteria from JSTATS [11] and Opta [12]. Contrary to previous studies [1], [10], the actions in this study include nuanced distinctions, such as differentiating between Short pass and Long pass, and between cut and trap, which are typically challenging to classify.

The labeling details included the time of action, the uniform color of the team executing the action, and the name of the action. The actions occurring while the video camera failed to capture the ball were excluded from labeling. Moreover, actions not listed in Table I and those with incomplete sensing data were also omitted from the labeling process. Table II presents the count of labeled actions, illustrating a notably higher frequency of short passes and traps compared to other actions.

C. Making dataset

With these labeled actions, a dataset was prepared for training a neural network. Table III lists the data count for



(a) $t = -2$ (b) $t = -1$ (c) $t = 0$ (d) $t = 1$ (e) $t = 2$

Fig. 1. Example extracted sensing data of players and the ball (short pass).

TABLE III
NUMBER OF DATA FOR TRAINING A NEURAL NETWORK

Name	Training	Validation	Total
Short pass	160	40	200
Long pass	128	40	168
Dribble	160	40	200
Cut	160	40	200
Trap	160	40	200

each action. The actions of short pass, dribble, and trap were selected to maintain a dataset size of 200 for each action. The occurrences of long pass and cut actions were less than 200. To avoid overfitting in the neural network due to nonuniform action frequencies, the dataset was balanced by adjusting the number of data points for each action. Consequently, data augmentation was employed for actions with fewer instances, involving duplication of existing data, thereby training the neural network twice with this augmented data. The aggregate dataset comprised 968 entries, segmented into 768 for training and 200 for validation, as outlined in Table III.

IV. MODEL TRAINING

A. Composition of input items

In this study, the classification of actions is presumed to be complex when the coordinates of players and the ball are directly fed into the neural network. To address this, we developed input items deemed crucial for effective action classification.

The input data items were derived from sensing data over a 3-s duration. As depicted in Fig. 1, the sensing data of players and the ball were extracted at five distinct time points with 0.75-s intervals. The primary reference time point was designated at the moment a player interacted with the ball. The other four time points were set at 0.75 and 1.50 s before and after this reference point. These time points are represented as $t = \{-2, -1, 0, 1, 2\}$, ordered from earliest to most recent. To standardize the offensive direction across all actions in the dataset, the x coordinates of the team attacking from left to right were inverted, simulating an offense from right to left. At each specific time point t, the x and y coordinates of the player executing the action were denoted as x_t^{player} and y_t^{player}, respectively. Similarly, the coordinates of the ball were defined as x_t^{ball} and y_t^{ball}.

The initial set of input items incorporated the distance of the ball from the opponent's goal at the first time point, in both x and y coordinates. Given that the pitch dimensions were 105 m in length and 68 m in width, the distance in the x coordinate is defined as follows:

$$d_x^{\text{goal}} = x_{-2}^{\text{ball}} \tag{1}$$

while the distance in the y coordinate is defined as follows:

$$d_y^{\text{goal}} = \mid y_{-2}^{\text{ball}} - 34 \mid . \tag{2}$$

The velocity of the ball between each pair of successive timepoints was selected as the second set of input items. The speed between timepoints t and $t+1$ ($t = -2, -1, 0, +1$) in the x coordinate system is calculated as follows:

$$v_x^{t,t+1} = \mid x_t^{\text{ball}} - x_{t+1}^{\text{ball}} \mid /0.75 \tag{3}$$

while the speed in the y-coordinate is defined as follows:

$$v_y^{t,t+1} = \mid y_t^{\text{ball}} - y_{t+1}^{\text{ball}} \mid /0.75. \tag{4}$$

The third set of input items involves the change in distance between the ball and the player executing the action across two consecutive timepoints. The deviation between the timepoints t and $t+1$ ($t = -2, -1, 0, +1$) in the x coordinate system is defined as follows:

$$c_x^{t,t+1} = \mid\mid x_t^{\text{ball}} - x_t^{\text{player}} \mid - \mid x_{t+1}^{\text{ball}} - x_{t+1}^{\text{player}} \mid\mid \tag{5}$$

while the deviation in the y coordinate is defined as follows:

$$c_y^{t,t+1} = \mid\mid y_t^{\text{ball}} - y_t^{\text{player}} \mid - \mid y_{t+1}^{\text{ball}} - y_{t+1}^{\text{player}} \mid\mid . \tag{6}$$

As all input items are defined separately for the x and y coordinates, the total number of input items amounts to $2 + 2 \times 4 + 2 \times 4 = 18$.

Additionally, an extra input item was defined, i.e., the ball possession rate. To calculate the ball possession rate p, the positions of players at the two timepoints preceding the action ($t = -2, -1$) are utilized. If the player nearest to the ball belonged to the same team as the action-taking player at either $t = -2$ or $t = -1$, a value of 0.5 is added to p. Thus, p can presume values of 0, 0.5, or 1.0. To assess the impact of the ball possession rate, the classification accuracies were compared between scenarios with and without the inclusion of the ball possession rate. When the ball possession rate is incorporated, the total number of input items increases to 19.

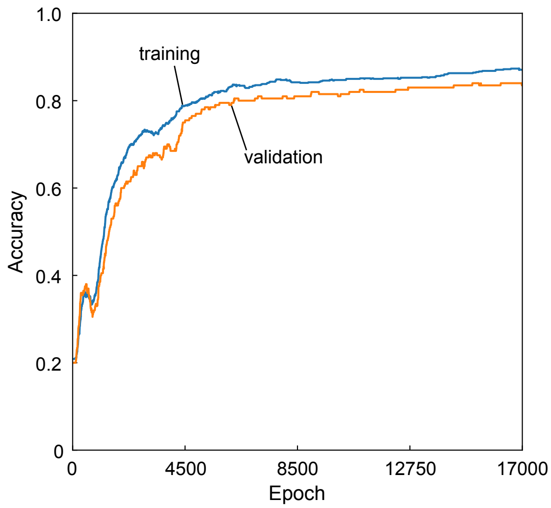


Fig. 2. Training curve when the number of input items is 19.

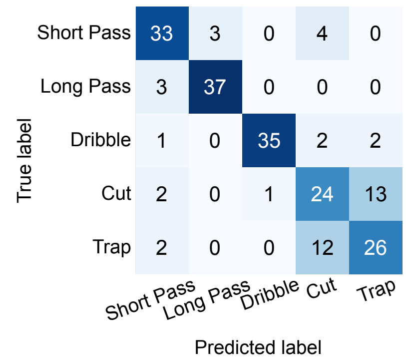
B. Network structure

Two distinct neural networks, one with 18 input items and the other with 19, were trained using 768 training data samples for classifying the actions of the football players. All input items were normalized using the Min–Max normalization. The network architecture comprised a four-layer structure, including an input layer, two hidden layers, and an output layer. The number of units in each layer were configured as 18 or 19, 15, 10, and 5, respectively. The rectified linear unit function was employed as the activation function, whereas the Softmax function was used in the output layer. The cross-entropy loss was selected as the loss function, and the stochastic gradient descent algorithm was utilized for optimization, with a learning rate set at 0.01. Both networks underwent training until the epoch count reached 17,000.

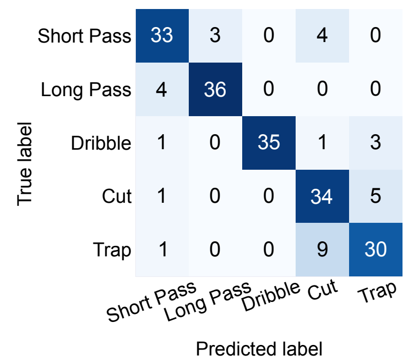
C. Result and discussion

Fig. 2 illustrates the training curve when the network was configured with 19 input items. The classification accuracies recorded were 77.5 % and 84.0 % for the networks with 18 and 19 input items, respectively. Fig. 3 presents the confusion matrices. The enhanced accuracy observed with the 19 input items suggests that incorporating the ball possession rate in the model training contributes to improved classification accuracy of football player actions. Although the classification accuracy of this sensing-based method is marginally lower compared to previous video-based methods [5], [6], the achievement of an accuracy exceeding 80 % indicates that our model attained a substantial level of accuracy. This outcome validates the suitability of the developed input items for the neural network training in recognizing football player actions.

The confusion matrices reveal that the numbers of correctly identified actions were increased for both Cut and Trap in the case of 19 input items. In comparison with prior sensing-based methodologies [1], [10], this study achieved commendable accuracy across all types of player actions, possibly attributable to the inclusion of time-series changes in player and ball locations within the input items.



(a) 18 input items



(b) 19 input items

Fig. 3. Confusion matrices. Each value represents the number of actions classified.

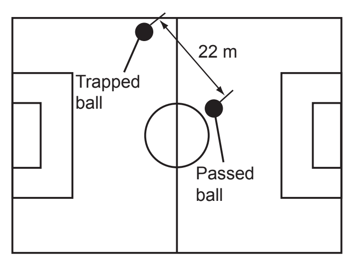
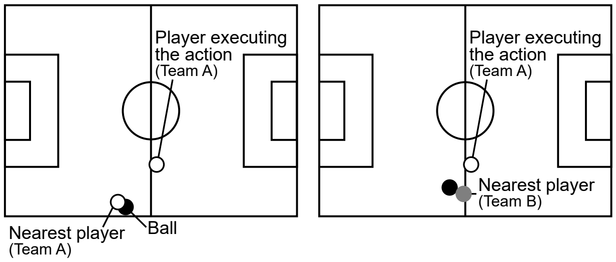


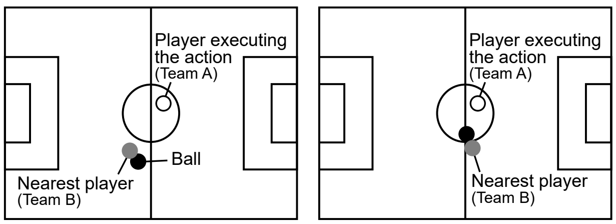
Fig. 4. Example instance where Short pass was incorrectly classified as Long pass.

The count of accurately classified Short pass and Trap actions was inferior to that of other actions. Additionally, there were notable misclassifications between Short pass and Long pass, as well as between Cut and Trap. Fig. 4 showcases an instance where a Short pass action was incorrectly classified as a Long pass. The classification criterion distinguishing between Short and Long passes is based on whether the distance between the passing player and the receiving player exceeds 30 m. The misclassification of this particular short pass as a Long pass likely occurred due to its proximity to the threshold, with a distance of 22 m.



(a) $t = -2$ (b) $t = -1$

Fig. 5. Example instance where Trap was incorrectly classified as Cut.



(a) $t = -2$ (b) $t = -1$

Fig. 6. Example instance where Cut was correctly identified.

Figs. 5 and 6 depict scenarios where a Trap action was erroneously classified as a Cut and where a Cut was correctly identified, respectively. The differentiation between Cut and Trap rests on whether the passing player and the receiving player belong to the same team. A comparison of these figures indicates almost identical positions of players and ball, except for the team affiliation of the player nearest to the ball. In this case, the ball possession rate failed to effectively determine the team identity of the passing player, leading to misclassification. To enhance the classification precision, the consideration of additional input items, such as ball acceleration and details about the team executing the preceding action, may prove beneficial. This approach could provide a more comprehensive dataset, potentially improving the accuracy of the neural network in distinguishing between closely related actions like Cut and Trap.

V. Conclusions

This research undertook the classification of football player actions using sensor data, serving as an initial step toward multimodal learning that integrates match video and sensing data. Initially, a sensor array was employed to capture the two-dimensional coordinates of players and the ball during football matches. Utilizing these recordings, a dataset comprising over 800 labeled instances of five distinct player actions was compiled. Neural networks, trained with 19 input items accounting for time-series changes in player and ball locations, exhibited a classification accuracy of 84.0 %. This outcome suggests that sensing data is efficacious for classifying football player actions.

In this study, the scope of classified actions was confined to five types. In future, we will aim to expand the range of actions to 12 for assessing the effectiveness of the sensing data in classifying a broader spectrum of actions. Additionally, to enhance classification accuracy, multimodal learning that consolidates sensing data and match video will be pursued.

As a prospective extension of this research, integrating the classifier developed in this study with action detection algorithms or detectors could facilitate the automated classification of football player actions. Moreover, the methodology employed in this study could be extrapolated to other team sports such as rugby or basketball, not exclusively football, to explore the classification of player actions in these domains.

References

[1] T. Imai, A. Uchiyama, T. Magome and T. Higashino, “Play recognition using soccer tracking data based on machine learning,” Advances in Network-Based Information Systems. NBiS 2018. Lecture Notes on Data Engineering and Communications Technologies, vol. 22, pp. 875–884, August 2018.

[2] C. Cuevas, D. Quilón and N. García, “Techniques and applications for soccer video analysis: a survey,” Multimedia Tools and Applications, vol. 79, pp. 29685–29721, August 2020.

[3] S. Akan and S. Varlı, “Use of deep learning in soccer videos analysis: survey,” Multimedia Systems, vol. 29, pp. 897–915, June 2023.

[4] P. Oskouie, S. Alipour and A. -M. Eftekhari-Moghadam, “Multimodal feature extraction and fusion for semantic mining of soccer video: a survey,” Artificial Intelligence Review, vol. 42, pp. 173–210, August 2014.

[5] A. Sen, S. M. M. Hossain, M. A. Russo, K. Deb, K. -H. Jo,“Fine-grained soccer actions classification using deep neural network,” Proceedings of 2022 15th International Conference on Human System Interaction, pp. 1–6, 2022.

[6] K. Muhammad, H. Ullah, M. S. Obaidat, A. Ullah, A. Munir, M. Sajjad and V. H. C. de Albuquerque, “AI-driven salient soccer events recognition framework for next-generation IoT-enabled environments,” IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2202–2214, February 2023.

[7] H. Zhu, J. Liang, C. Lin, J. Zhang and J. Hu, “A transformer-based system for action spotting in soccer videos,” Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports, pp. 103–109, October 2022.

[8] V. Khaustov and M. Mozgovoy, “Recognizing events in spatiotemporal soccer data,” Applied Sciences, vol. 10, no. 22, November 2020.

[9] D. Fassmeyer, G. Anzer, P. Bauer and U. Brefeld, “Toward automatically labeling situations in soccer,” Frontiers in Sports and Active Living, vol. 3, pp. 137–150, November 2021.

[10] F. Vidal-Codina, N. Evans, B. E. Fakir and J. Billingham, “Automatic event detection in football using tracking data,” Sports Engineering, vol. 25, September 2022.

[11] Japan Professional Football League, “J STATS,” https://aboutj.jleague.jp/corporate/terminology/stats/, (accessed Nov. 9, 2023).

[12] Stats Perform, “OPTA EVENT DEFINITIONS,” https://www.statsperform.com/opta-event-definitions/, (accessed Nov. 9, 2023).