Detecting Anomalous Kicks in Taekwondo With Spatial and Temporal Features

DANBI CHO[1], **HYUNYOUNG LEE**[1], **WONJAE LEE**[2], **AND SEUNGSHIK KANG**[1]
[1]Department of Computer Science, Kookmin University, Seoul 02707, Republic of Korea
[2]Department of Sports of Industry and Leisure, Kookmin University, Seoul 02707, Republic of Korea

Corresponding author: Seungshik Kang (sskang@kookmin.ac.kr)

ABSTRACT A new scoring system based on information technology (IT) is an innovative way to judge matches fairly. However, there are reliability and expertise problems due to sophisticated differences in human-machine judging criteria. Taekwondo, a traditional Korean sport, uses an IT-based protector and scoring system. The system records the player's movement data through its sensors, but it has a limitation in that it cannot detect anomalous kicks. Because anomalous kicks are executed with minimal strength outside the typical form of kick patterns, it interrupts fair judgment of the scoring system and reduces the liveliness of Taekwondo matches. To minimize the dispute by anomalous kicks, we aim to detect the anomalous kicks in Taekwondo matches. We construct a kick dataset and propose an attention-based deep learning model. The proposed model has the advantage of simultaneously handling the spatial and temporal features of the kick dataset because it is a model that combines a convolutional neural network (CNN) and bidirectional long-short term memory (BiLSTM) with an attention mechanism. The experimental results showed that the accuracy of the attention-based CNN-BiLSTM is 95.67%, and the false-positive rate is 0.086. We open our dataset and it is freely available on https://github.com/daanVeer/Taekwondo_dataset.

INDEX TERMS Intelligent systems, machine learning, recurrent neural networks, Taekwondo match, anomalous kicks, convolutional neural network, bidirectional long-short term memory, attention mechanism.

I. INTRODUCTION

In most sports, a human judges the match and it leads to controversies such as wrong or biased judgment and manipulation [1]. To prevent such controversies, scoring systems based on information technology (IT) have been introduced. Examples include a ball position tracking system in tennis, a video judgment assistance system in soccer, and a contact detection system in fencing. IT-based scoring systems have been evaluated as an effective judgment method by improving the fairness and objectivity of sports matches [2].

The protector and scoring system (PSS) is commonly used in Taekwondo matches like World Taekwondo Championships and Olympic Taekwondo Competition. Since its introduction in 2006, the PSS contributed greatly to judge Taekwondo matches objectively by capturing the strength and accuracy of the kick [3]–[5]. However, many studies have discussed the flaws of the PSS such as incorrect judgment of

detecting anomalous kicks (e.g., a monkey kick or scorpion kick) that abuse the scoring principles [6]–[9]. Anomalous kicks reduce the liveliness of a Taekwondo match because they are intended to obtain a score from minimal strength. The foot jab is one of the anomalous kicks that easily obtains a score by a light touch that disturbs the correct judgment of the scoring system [10]–[13]. Our goal is to enhance the current PSS to the intelligent PSS by automatically detecting anomalous kicks based on deep learning approach.

The PSS is equipped with sensors in various locations to capture the features of each axis. We constructed a kick dataset from the PSS sensors and applied it to the deep learning model. The PSS sensors record the players' movement data as sequential frames form. The kick dataset includes the spatial and temporal features because it represents the sequential information about a kick motion and location information about sensors. To learn the spatial and temporal features simultaneously in sequential kick dataset extracted from the PSS, we propose a combining model of attention-based CNN-BiLSTM. Attention-based CNN-BiLSTM model

combines a convolutional neural network (CNN) and long short-term memory (LSTM) with an attention mechanism. CNN plays a role of extracting the important features from the dataset and LSTM trains the features including the sequential information [14]–[17]. Attention mechanism compresses the important features and reminds them, so it has been shown to improve performance in prior studies [18]. Considering the influence of attention mechanism, we added the attention mechanism to the CNN-BiLSTM model. We further compared the performances of combining model with and with no attention mechanism. The remainder of the paper is structured as follows. Section 2 investigates the related works. Section 3 proposes a model architecture for our purpose. Section 4 constructs a dataset for detecting anomalous kicks and analyzes the feature patterns. In section 5, the experimental results are described including the evaluation metrics, followed by a summary of the conclusions in section 6.

II. RELATED WORKS

Many studies have utilized statistical methods to process sequential data such as text and audio. Jinqing *et al.* [19] classify the types of sound in sports audio: speech, music, and environmental noise. They extract feature vectors from the spectrum of audio signals and input them into a Gaussian mixture model (GMM). Hanna *et al.* [20] compose the color change rate by averaging the saturation of RGB (red, green, and blue) pixels according to sequential frames in the video data. The color change rate is used as a feature to classify the type of sport and is applied to the hidden Markov model (HMM).

As deep learning models are developed, recurrent neural network (RNN) models have been introduced, such as vanilla-recurrent network (vanilla-RNN), long short-term memory (LSTM), and gated recurrent unit(GRU). RNN models are effective in learning the temporal features of sequential data because they save the histories of the previous time steps while transferring the feature vector of the previous time step to the next time step. RNN models have performed successfully in research using not only text and audio data but also video and sensor data.

Sequential image data are similar to text data because the two types of data have sequential information at each step. Sequential image data are used for the action or gesture recognition task and the video classification. Asadi-Aghbolaghi *et al.* (2017) surveyed related works on methods to process sequential image data with temporal features for the action and gesture recognition task. They showed that combining spatial features using a CNN and temporal features using LSTM facilitates handling variable-length images [21]. Chen *et al.* [22] conducted the action recognition by analyzing the spatio-temporal representation abilities of CNN. Many studies have implemented the combining model, which feeds the output weight of the CNN into the RNN models [23]. Wu *et al.* (2015) proposed a hybrid deep learning framework with a two-stream CNN, namely, a spatial stream CNN and a motion stream CNN, and LSTM. They used a sequential-image dataset for the video classification task, and built spatial frames and stacked motion optical flow to train the model. Donahue *et al.* [24] introduced the long-term recurrent convolutional network for activity recognition, image captioning, and video description tasks.

Zhu *et al.* [25] and You and Korhonen [26] aggregated the spatial and temporal features for the video classification task by combining a CNN and RNN models. They added an attention mechanism to integrate the network and the model boost the performance in video classification. Nikolaidou *et al.* [27] recognized the human activities using a novel CNN-LSTM architecture, which reflects the spatial and temporal by the color and texture of the images.

III. PROPOSED ARCHITECTURE

Recent researches on Taekwondo are the comparison of skill actions between analog and electronic protectors, setting standards for kick impacts, and predicting the intention to accept the electronic protection devices. A 2D CNN has been proposed to analyze the spatio-temporal representations for action recognition and the attention mechanism improves the performance of deep learning models in which it emphasizes important features in the dataset by feeding the attention weights. The experiments show that attention mechanism in video classification is effective when it is added to the LSTM model. In this paper, we focused on the attention-based CNN-BiLSTM model for detecting anomalous kicks. We applied the attention mechanism after BiLSTM and the model is organized in the order of CNN, BiLSTM, and the attention mechanism. Fig. 1 summarizes the flow of our model. A 2D CNN is used to learn the spatial features and BiLSTM to learn the temporal features for the kick dataset.

CNN identifies patterns in data with convolution operations in the computer vision processing of image while maintaining the spatial information and it achieves a notable performance in text and video classification tasks. Most CNN models comprise several layers, stacking convolutional layers, pooling layers, and fully connected layers to extract the data features. The CNN then creates a feature map by traversing the input data through a filter while performing a convolution operation. The feature map depends on the layer characteristics, such as filter size, stride, padding, and pooling type. We constructed the 2D CNN to learn the spatial features from the kick data, as shown in Fig. 1. The inner structure of the 2D CNN in Fig. 1 is shown in Fig. 2. It has a total of 101 2D CNNs to process the sequential frames for a kick, where a frame is inputted into a 2D CNN.

We explored three steps process in 2D CNN. First, we separate data from the four sensors on each frame, denoted as (S^1, S^2, S^3, S^4).

- S_1: the torso sensor.
- S_2: the pelvis sensor.
- S_3: the right leg sensor.
- S_4: the left leg sensor.

Second, we carry out the convolution operation of hidden size 32 with *ReLU* as the activation function to summarize
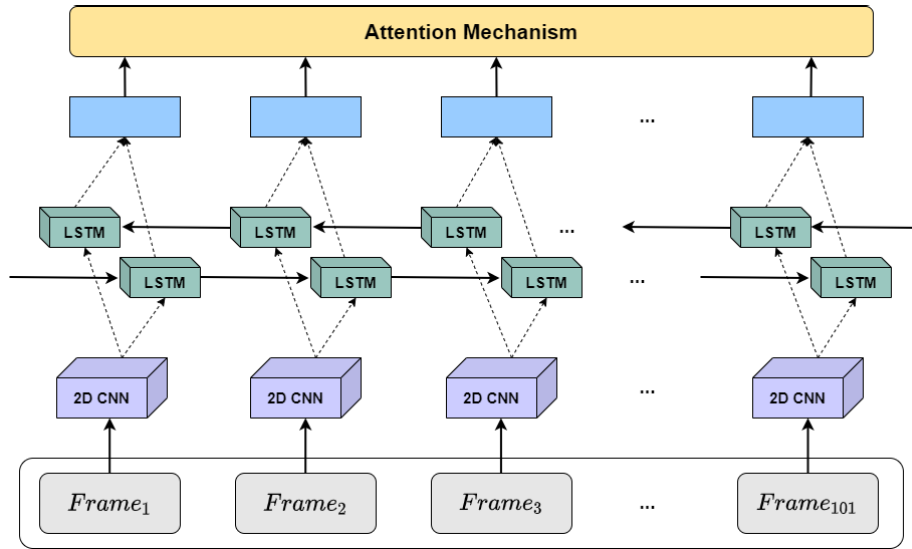
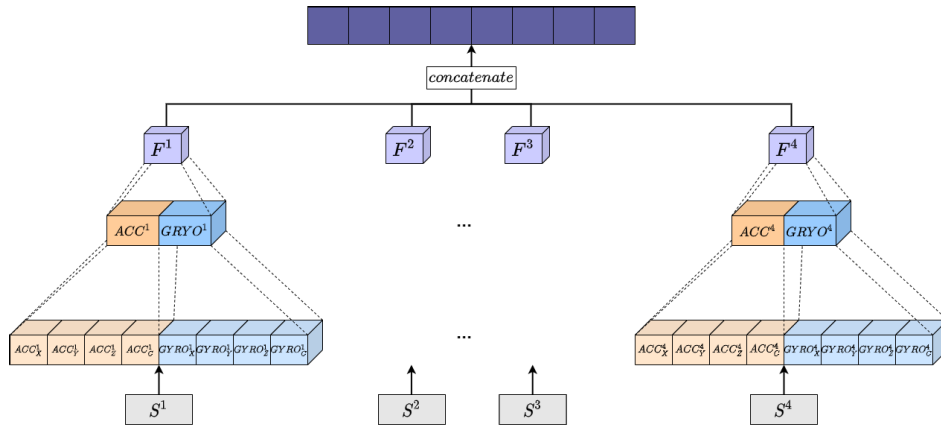**FIGURE 1.** Attention-based CNN-BiLSTM architecture.

FIGURE 2. Inner structure of the 2D CNN for learning the spatial features.

the spatial features of the acceleration and gyro values, respectively, by noting the features of the x, y, and z axis and the composite value in each sensor. For example, ACC^i is an acceleration feature that corresponds to the ACC_X^i, ACC_Y^i, ACC_Z^i, and ACC_C^i for the i-th sensor, and likewise for $GYRO^i$. Finally, we compress the ACC and $GYRO$ features from each sensor as $F^i = [ACC^i, GYRO^i]$ for the i-th sensor with a convolution operation of hidden size 16 with *ReLU* as the activation function. Consequently, one frame data is expressed as four compressed spatial features F^1, F^2, F^3, and F^4. We concatenate all the spatial features as $[F^1; F^2; F^3; F^4]$ and feed it into the LSTM network. We conduct average pooling to reduce the number of parameters after the convolution operation.

RNN models update the weight values using the histories of the sequential data in the process of learning the features. So, RNN models are effective in processing sequential data such as text, audio, and video. Vanilla-RNN is the basic RNN model, but it has a problem: information is lost when the length of the sequential data increases. LSTM improves

vanilla-RNN by adding gates: a forget gate f, input gate i, and output gate o. For the time step t, LSTM learns x_t, h_{t-1} and c_{t-1}, where h_{t-1} is a previous hidden state and c_{t-1} is a previous memory cell. In equation (1), W_{xf} means the weight matrix of input token x in forget gate f, W_{hf} means the weight matrix of hidden state h in forget gate f, and b_f is bias term of forget gate f, and likewise for input gate i, output gate o, information to save g, and memory cell c. $\odot$ indicates the element-wise product of vectors.

$$
\begin{aligned}
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
g_t &= \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t
\end{aligned}
\tag{1}
$$

GRU was introduced as a model to increase memory efficiency by reducing the number of LSTM gates. Unlike LSTM, which has three gates, GRU is trained with two gates: a reset gate r and update gate z. In equation (2), W_{xr} means

the weight matrix of input token x in reset gate r, W_{hr} means the weight matrix of hidden state h in reset gate r, and b_r is bias term of reset gate r, and likewise for update gate z, output gate o, information to save g, and memory cell c.

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$$
$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z)$$
$$g_t = \tanh(W_{xg}x_t + W_{hg}(r_t \odot h_{t-1}) + b_g)$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t)g_t \tag{2}$$

To detect the anomalous kicks in Taekwondo matches, we explored the LSTM and GRU, respectively, and compared their performances. To learn both past and future information, we transferred the bidirectional state $[\vec{h}; \overleftarrow{h}]$ by concatenating $\vec{h}$ and $\overleftarrow{h}$ to the RNN model (BiLSTM and BiGRU), where $\vec{h}$ and $\overleftarrow{h}$ denote the forward state and backward state. We designed the BiLSTM and BiGRU models by bidirectionally encoding the state of the LSTM and GRU. Because the kick data consists of a sequence of frames, we used the RNN models to train the temporal data features. Before learning the temporal features, the spatial features are transferred from the 2D CNN to the RNN models. The RNN models are fed with the features through the 2D CNN and reduce the dimension of the vector space to a hidden size with *tanh* as the activation function.

We applied the weight values of BiLSTM (or BiGRU) to the Bahdanau attention mechanism [28]. Attention mechanism inspects the input data once again in the decoder, along with the context vector expressed from the encoder. Because the attention mechanism focuses on the importance of a particular time step while assigning a high weight value, it effectively minimizes the information loss. When hidden states of an encoder are $[h_1, h_2, \ldots, h_n]$, from 1 to n time step, and the hidden state of a decoder by t time step is s_t, the attention score is calculated as shown in equation (3). In this equation, h_i is a hidden state of the i-th time step of the encoder, and W_a, W_b, and W_c are weight vectors for training model.

$$score(s_{t-1}, h_i) = W_a^T \tanh(W_b s_{t-1} + W_c h_i) \tag{3}$$

Using the BiLSTM (or BiGRU) weight values, we obtained the context vector and attention weights from the attention mechanism with a hidden size of 64. We further added two dense layers after the attention mechanism. One dense layer has a hidden size of 8 with *ReLU* as an activation function, and the other dense layer has a hidden size of 2 with *softmax* as the activation function to output the probability distribution. To find the optimal weight, the Adam optimizer is used with a learning rate of 0.001 and cross-entropy as the loss function. In equation (4), $q(x_i)$ and $p(x_i)$ are the true probability distribution and predicted probability distribution, respectively, for the i-th data x_i. The model was trained for 10 epochs with a batch size of 64. We implemented the models on docker images using Ubuntu 18.04 and TensorFlow version 2.2 as the deep learning framework with

a 3.2 GB NVIDIA Tesla 100.

$$Cross\ Entropy\ Loss = -\sum_{i=1}^{n} q(x_i)log(p(x_i)) \tag{4}$$

IV. DATASET AND DATA ANALYSIS
A. DATASET
To detect anomlaous kicks in Taekwondo matches, we constructed a kick dataset by utilizing sensors from the PSS. In the environment of a Taekwondo match, we attached four sensors (torso, pelvis, right leg, and left leg) to the PSS and performed kicks. One sensor represents eight features as follows.

- ACC_X: acceleration value in x-axis.
- ACC_Y: acceleration value in y-axis.
- ACC_Z: acceleration value in z-axis.
- ACC_C: composite acceleration value.
- $GYRO_X$: gyro value in x-axis.
- $GYRO_Y$: gyro value in y-axis.
- $GYRO_Z$: gyro value in z-axis.
- $GYRO_C$: composite gyro value.

The kick data consists of 32 features from the four sensors. We recorded 101 sequential frames (0.02sec/frame) from a kick and obtained 32 features of 101 frames from one kick. Then, we labeled an anomalous kick as 0 and a normal kick as 1. Monkey kick is an anomalous kick and the roundhouse, back hook, and push kicks are normal kicks. Anomalous kicks have been performed in 279 kicks, and normal kicks have been performed in 285 kicks, resulting in a total of 564 kicks. The 285 normal kicks consists of 97 roundhouse kicks, 100 back hook kicks, and 88 push kicks. To prevent data-based bias, we randomly extracted 270 kicks for each kick type, which resulted in a 540 kick dataset constructed with 270 anomalous kicks and 270 normal kicks. We divided the dataset into training set and test set for model training in a 9:1 ratio from each type of kick.

The video data, $n \times n$ images, are composed of sequential data from k consecutive frames. Because the kick dataset has a format in which 1×32 features are composed of sequential data from 101 frames, we assume that the kick dataset is similar to the video data format in the video classification tasks. Therefore, we treat the kick dataset constructed from the PSS as video data for detecting anomalous kicks.

B. DATA ANALYSIS
To analyze the kick dataset, we present the patterns of anomalous kicks and normal kicks using the training set. The patterns are separated by sensor location, type of kick, and sensor values. Fig. 3 shows the patterns of a sensor attached to the torso position of the four sensors (torso, pelvis, right leg, and left leg). The thick line indicates the average line of the kick.

Fig. 3(a) shows the torso sensor patterns of anomalous kicks, whereas Fig. 3(b) shows the pattern of normal kicks. Of the eight features on the torso sensor, ACC_X shows the continuous x-axis acceleration value of anomalous kicks and their average as a thick line. Pattern analysis indicates
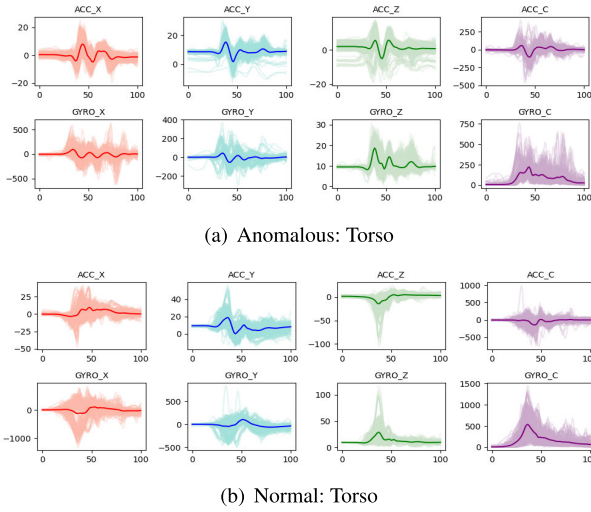
(a) Anomalous: Torso

(b) Normal: Torso

FIGURE 3. Pattern distributions of the torso sensor; (a) anomalous kicks, and (b) normal kicks.

TABLE 1. Experimental results for detecting anomalous kicks in Taekwondo matches.

Model	ACC(%)	FNR	FPR
Baseline Models			
KNN	89.83	0.131	0.071
SVM	91.39	0.088	0.083
Comparable Models			
CNN	86.41	0.197	0.074
CNN+LSTM	91.35	0.00	0.172
CNN+BiLSTM	93.20	0.00	0.172
CNN+GRU	91.35	0.00	0.172
CNN+BiGRU	93.20	0.024	0.111
Proposed Models			
CNN+BiLSTM+Attention	**95.67**	**0.00**	0.086
CNN+BiGRU+Attention	95.06	0.049	**0.049**

anomalous kicks have more variation than do normal kicks. Anomalous kicks show a less clear strike point than normal kicks, so most of the average lines appear to be uneven flows. We found that a particular strike point was noticeable in normal kicks, including roundhouse, back hook, and push kicks, especially on the gyro value of the z-axis (GYRO_Z) and composite gyro value (GYRO_C). However, the average lines of the normal kick were flat. We believe that it is because the normal kicks consist of three kinds of kicks (roundhouse, back hook, and push kicks). Pattern analysis for the remaining three sensors (pelvis, right leg, and left leg), excluding the torso sensor, are presented in the appendix.

V. EXPERIMENTAL RESULTS

Table 1 shows the experimental results of the three types of models: baseline models, comparable models, and proposed models. For the baseline models, KNN and SVM, we use the same experimental results in Cho *et al.* [29]. We compared the results of KNN using the L1 distance and SVM with the results of our models. We implemented the models combining CNN and BiLSTM with an attention mechanism. To confirm whether the attention mechanism improves the performance in our study, we compared the comparable models and proposed models with and with no attention mechanism. The comparable models are CNN, CNN + LSTM,

CNN + BiLSTM, CNN + GRU, and CNN + BiGRU with no attention mechanism. The proposed models are attention-based CNN-BiLSTM and attention-based CNN-BiGRU, denoted as CNN + BiLSTM + Attention and CNN + BiGRU + Attention, respectively, in Table 1.

Our model has been evaluated by using three metrics: accuracy, false-negative rate (FNR), and false-positive rate (FPR).

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad (5)$$

$$FNR = \frac{FN}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{TN + FP} \quad (7)$$

- true positives (TP): normal kicks, that classified positive.
- false positives (FP): anomalous kicks, that classified positive.
- true negatives (TN): anomalous kicks, that classified negative.
- false negative (FN): normal kicks, that classified negative.

FNR is the rate of normal kicks incorrectly classified as anomalous kicks. If a model classifies a normal kick as an anomalous kick, players will not obtain a score. On the other hand, FPR is the rate of anomalous kicks classified incorrectly as normal kicks. If a model classifies an anomalous kick as a normal kick, players will receive an unfair score for non-scoring kicks. Therefore, we evaluated our models using accuracy, FNR, and FPR to confirm how well the model detects the type of kick in Taekwondo matches.

To prevent the problem of overfitting, we divided the training set by 8:2 ratio to obtain a validation set and validated the performance of the model while tuning the model parameters. Using the model with the highest performance when training a model with the validation set, we evaluated the test set. We presented the results in Table 1 by averaging the experiments of three times on the test set. The proposed models outperformed the baseline and comparable models. Baseline models, KNN and SVM, had accuracies of 89.83% and 91.39%, respectively. The SVM performance was better than the CNN performance (86.41%) in the comparable models. In terms of error rate, SVM had a lower FNR (0.088) than KNN (0.131), but the FPR was lower with KNN (0.071) than SVM (0.083). The comparable models other than CNN performed better than the baseline models. In particular, the bidirectional models such as BiLSTM and BiGRU (both 93.20%) outperformed the unidirectional models such as LSTM and GRU (91.35% and 91.97%, respectively). The FNR was lowest with the CNN + LSTM, CNN + BiLSTM, and CNN + GRU models. However, the CNN FPR was lower than that of the other comparable models.

The models with attention mechanism outperform the comparable models with no attention mechanism by about 2%p accuracy. It shows that the attention mechanism improves the performance on the task of detecting anomalous
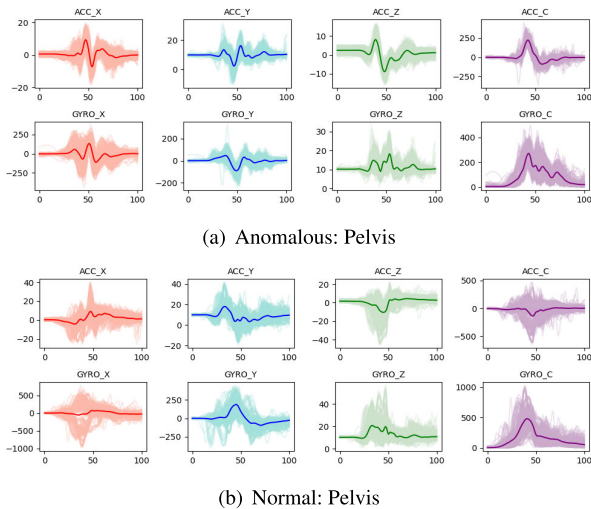
(a) Anomalous: Pelvis

(b) Normal: Pelvis

FIGURE 4. Pattern distributions of the pelvis sensor; (a) anomalous kicks, and (b) normal kicks.

kicks. The attention-based CNN-BiLSTM achieved a higher accuracy of 95.67% and a lower FNR than the attention-based CNN-BiGRU. However, the attention-based CNN-BiGRU had a better FPR at 0.049 than the attention-based CNN-BiLSTM. Through these results, we found that the BiLSTM and BiGRU perform similarly, but BiLSTM is better than BiGRU on accuracy and FNR, and BiGRU is better on FPR.

VI. CONCLUSION

The PSS is widely used in domestic and overseas Taekwondo matches like World Taekwondo Championships and Olympic Taekwondo Competition, improving the fairness and accuracy of scoring Taekwondo fights. However, IT-based scoring system caused a problem of anomalous kicks of obtaining an easy sensor-specialized score. Thus, we explored an intelligent PSS by automatically detecting anomalous kicks through deep learning approach by analyzing player's movements. We constructed a kick dataset and proposed an attention-based CNN-BiLSTM for detecting anomalous kicks in Taekwondo matches. The proposed model has the advantage of handling sequential sensor values by training spatial and temporal features simultaneously through combining a 2D CNN, BiLSTM/BiGRU, and an attention mechanism. The attention-based CNN-BiLSTM achieved 95.67% accuracy, and a 0.086 FPR. Furthermore, the attention-based CNN-BiGRU had a better FPR at 0.049 than the attention-based CNN-BiLSTM. Experimental results show that the attention-based CNN-BiLSTM model improved 2%p accuracy compared to the CNN-BiLSTM with no attention mechanism. We conducted a research in cooperation with KP&P, a PSS manufacturing company, and our goal is to apply our model to the real Taekwondo matches for judging the domestic and overseas Taekwondo competition. So, it is expected that our research result will practically contribute to the fair judgment in Taekwondo

matches. We make our Taekwondo dataset publicly available on GitHub for further researches.

APPENDIX A
PATTERN DISTRIBUTIONS OF OTHER SENSORS

We analyzed the patterns of kick in accordance with the location of sensor. Fig. 4, 5, and 6 present the pattern distributions of sensor values attached on the pelvis, right leg, and left leg, respectively.
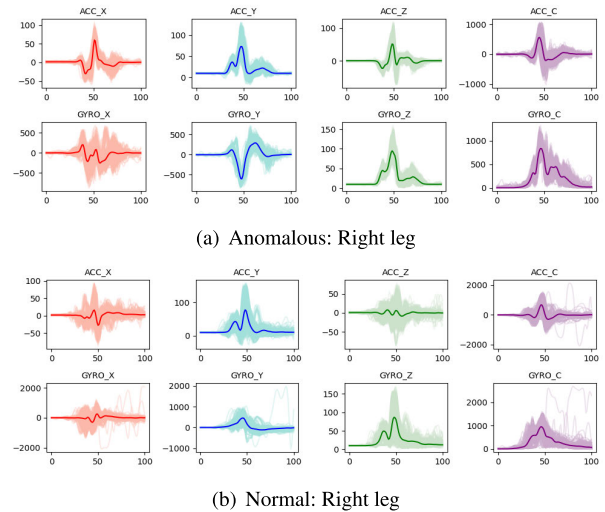

(a) Anomalous: Right leg

(b) Normal: Right leg

FIGURE 5. Pattern distributions of the right leg sensor; (a) anomalous kicks, and (b) normal kicks.

(a) Anomalous: Left leg
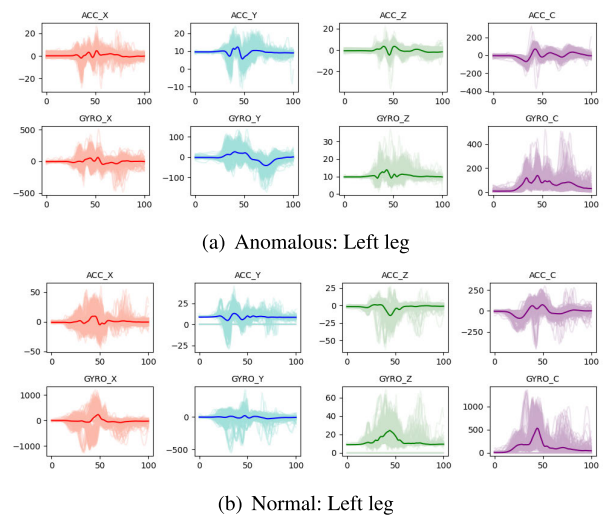

(b) Normal: Left leg

FIGURE 6. Pattern distributions of the left leg sensor; (a) anomalous kicks, and (b) normal kicks.

REFERENCES

[1] B. Dyer, "The controversy of sports technology: A systematic review," *SpringerPlus*, vol. 4, no. 1, pp. 1–12, Sep. 2015.
[2] S. J. Haake, "The impact of technology on sporting performance in Olympic sports," *J. Sports Sci.*, vol. 27, no. 13, pp. 1421–1431, Nov. 2009.
[3] J.-H. Park, "Study on the psychology for the improvement of competitions in Taekwondo," *Korean J. Sports Sci.*, vol. 29, no. 2, pp. 423–439, Apr. 2020.

[4] E.-H. Cho, H.-J. Eom, and S.-Y. Jang, "Comparison of patterns of skill actions between analog and electronic protectors in Taekwondo: A log-linear analysis," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, p. 3927, Jun. 2020.

[5] S. U. Park, D. K. Kim, and H. K. Ahn, "A predictive model on the intention to accept Taekwondo electronic protection devices," *Appl. Sci.*, vol. 11, no. 4, p. 1845, Feb. 2021.

[6] E. H. Cho and H. J. Eom, "Validity of electronic-protector scoring system in Taekwondo game," *Korean J. Meas. Eval. Phys. Educ. Sports Sci.*, vol. 9, no. 2, pp. 93–104, Dec. 2007.

[7] D. S. Yang, "Games management strategy and improvement plan search in accordance with the introduction of electronic protector in Taekwondo tournament," *J. Korean Soc. Philosophy Sport, Dance, Martial Arts*, vol. 16, no. 4, pp. 189–202, Dec. 2008.

[8] S. M. Park and D. S. Yang, "Field views and means of improvement on usage of electronic protective gears in Taekwondo tournaments," *J. Korean Soc. Philosophy Sport, Dance, Martial Arts*, vol. 17, no. 4, pp. 217–234, Dec. 2009.

[9] C.-H. Choi, H. Oh, and M. Jeon, "Adequacy of setting standards for kick impact in the Taekwondo electronic scoring system: Comparison of a reference group model based on empirical data," *BMC Sports Sci., Med. Rehabil.*, vol. 13, no. 1, pp. 1–7, Dec. 2021.

[10] W. J. Lee and C. H. Choi, "Issues in news articles related to Taekwondo protector & scoring system: Topic modeling and semantic network analysis approach," *Korean J. Meas. Eval. Phys. Educ. Sports Sci.*, vol. 22, no. 4, pp. 41–53, Dec. 2020.

[11] M. C. Kim, J. H. Yu, and H. Y. Jung, "An analysis of the relationship among perceived quality, trust, technology acceptance, and intention to use in electronic body protectors of Taekwondo," *Taekwondo J. Kukkiwon*, vol. 10, no. 3, pp. 283–304, 2019.

[12] S.-Y. Kim and A.-R. Lee, "Status of Taekwondo competition," *J. Martial Arts*, vol. 14, no. 2, pp. 23–38, May 2020.

[13] T.-T. Liu, Y.-C. Lin, W.-T. Tang, J. Hamill, and J.-S. Chang, "Lower-limb kinematic characteristics of Taekwondo kicks at different attack angles," *Int. J. Perform. Anal. Sport*, vol. 21, no. 4, pp. 519–531, Jul. 2021.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–14.

[16] W. Yang, J. Li, F. Fukumoto, and Y. Ye, "HSCNN: A hybrid-siamese convolutional neural network for extremely imbalanced multi-label text classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 6716–6722.

[17] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," 2016, *arXiv:1611.06639*.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 5998–6008.

[19] L. Jiqing, D. Yuan, H. Jun, Z. Xianyu, and W. Haila, "Sports audio classification based on MFCC and GMM," in *Proc. 2nd IEEE Int. Conf. Broadband Netw. Multimedia Technol.*, Beijing, China, Oct. 2009, pp. 482–485.

[20] J. Hanna, F. Patlar, A. Akbulut, E. Mendi, and C. Bayrak, "HMM based classification of sports videos using color feature," in *Proc. 6th IEEE Int. Conf. Intell. Syst.*, Sofia, Bulgaria, Sep. 2012, pp. 388–390.

[21] M. Asadi-Aghbolaghi, A. Clapes, M. Bellantonio, H. J. Escalante, V. Ponce-Lopez, X. Baro, I. Guyon, S. Kasaei, and S. Escalera, "A survey on deep learning based approaches for action and gesture recognition in image sequences," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Washington, DC, USA, May 2017, pp. 476–483.

[22] C.-F.-R. Chen, R. Panda, K. Ramakrishnan, R. Feris, J. Cohn, A. Oliva, and Q. Fan, "Deep analysis of CNN-based spatio-temporal representations for action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6175–6465.

[23] Z. Wu, X. Wang, Y.-G. Jiang, H. Ye, and X. Xue, "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 461–470.

[24] H. Donahue, L. A. Hendricks, S. Guadarrama, and M. Rohrbach, "Long-term recurrent convolutional networks for visual recognition and description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017.

[25] L. Zhu, D. Tran, L. Sevilla-Lara, Y. Yang, M. Feiszli, and H. Wang, "FASTER recurrent networks for efficient video classification," in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 13098–13105.

[26] J. You and J. Korhonen, "Attention boosted deep networks for video classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Abu Dhabi, UAE, Oct. 2020, pp. 1761–1765.

[27] S. S. Nikolaidou, I. Vernikos, E. Mathe, E. Spyrou, and P. Mylonas, "A novel CNN-LSTM hybrid architecture for the recognition of human activities," in *Proc. 22nd Eng. Appl. Neural Netw. Conf.*, 2021, pp. 121–132.

[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.

[29] D. B. Cho, H. Y. Lee, W. J. Lee, and S. S. Kang, "Machine learning based classification of anomalous kicks in Taekwondo games," in *Proc. Korea Comput. Congr.*, Jeju, Republic of Korea, 2021, pp. 798–800.

DANBI CHO received the B.A. degree in big data analytics business statistics from Kookmin University, Seoul, South Korea, in 2020, where she is currently pursuing the M.S. degree in computer science. Her research interests include natural language processing, machine learning, deep learning, and text mining.

HYUNYOUNG LEE received the B.S. and M.E. degrees in computer science from Kookmin University, Seoul, South Korea, in 2016 and 2019, respectively, where he is currently pursuing the Ph.D. degree in computer science. His research interests include natural language processing, machine learning, deep learning, and recommendation systems.

WONJAE LEE received the bachelor's degree in sport science from Seoul National University, Seoul, South Korea, in 2000, the M.B.A. degree from Korea University, Seoul, in 2005, the master's degree in sport management from AISTS, Lausanne, Switzerland, in 2008, and the Ph.D. degree from Seoul National University, in 2012. From 2000 to 2003, he served the Republic of Korea Navy as a Lieutenant. From 2004 to 2019, he worked as a Sport Administrator at various organizations, including the Korean Olympic Committee and the World Taekwondo Federation, Seoul. Since 2019, he has been an Associate Professor at Kookmin University, teaching and studying sport management.

SEUNGSHIK KANG received the B.S., M.S., and Ph.D. degrees in computer science from Seoul National University, Seoul, South Korea, in 1986, 1988, and 1993, respectively. He worked as an Associate Professor at Hansung University, from 1994 to 2001. He is currently working at Kookmin University as a Full Professor. His research interests include natural language processing, information retrieval, text mining, big data processing, and machine learning.

• • •