J.Natn.Sci.Foundation Sri Lanka 2024 52 (1): 143 - 156
DOI: http://dx.doi.org/10.4038/jnsfsr.v52i1.11986

RESEARCH ARTICLE

Deep Learning

Evaluation of Taekwondo Poomsae movements using skeleton points [†]

M Fernando[*], KD Sandaruwan and AMKB Athapaththu
University of Colombo School of Computing, UCSC Building Complex, 35 Reid Ave, Colombo 00700, Sri Lanka.

Submitted: 01 July 2023; Revised: 04 December 2023; Accepted: 22 December 2023

Abstract: Taekwondo is a widely practised martial art and an Olympic sport. In Taekwondo, Poomsae movements are essential, as they form the foundation of the sport and are fundamental for success in competitions. The evaluation of Poomsae movements in Taekwondo has been a subjective process, relying heavily on human judgments. This study addresses the above issue by developing a systematic approach to evaluate Poomsae movements using computer vision. A long short-term memory-based (LSTM-based) machine learning (ML) model was developed and evaluated for its effectiveness in Poomsae movement evaluation. The study also aimed to develop this model as an assistant for self-evaluation, that enables Taekwondo players to enhance their skills at their own pace. For this study, a dataset was created specially by recording Poomsae movements of Taekwondo players from the University of Colombo. The technical infrastructure used to capture skeleton point data was cost-effective and easily replicable in other settings. Small video clips containing Taekwondo movements were recorded using a mobile phone camera and the skeleton point data was extracted using the MediaPipe Python library. The model was able to achieve 61% of accuracy when compared with the domain experts' results. Overall, the study successfully achieved its objectives of defining a self-paced approach to evaluate Poomsae while overcoming human subjectivity otherwise unavoidable in manual evaluation processes. The feedback of domain experts was also considered to finetune the model for better performance.

Keywords: Long short-term memory (LSTM), machine learning (ML), skeleton points, Taekwondo, video classification.

INTRODUCTION

The integration of technology to enhance athletes' performance has become a prominent research area. However, movement evaluation in sports remains challenging due to limited technological involvement and the high costs associated with accessing required high-end technological infrastructure. Computer vision has also emerged as a promising technology also for evaluating movements in sports training.

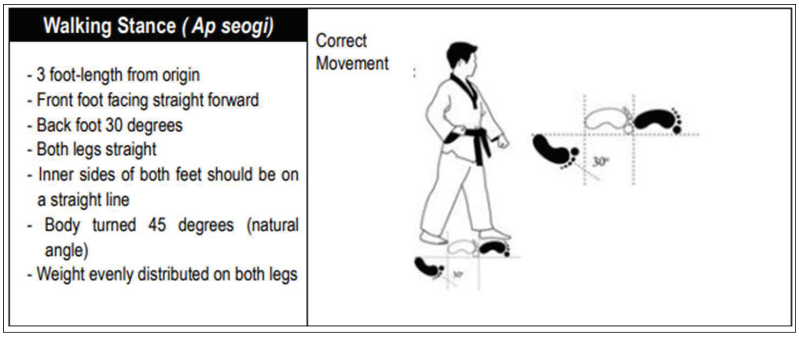
Taekwondo is a globally popular martial art which has also been recognized as an Olympic sport. In Taekwondo, Poomsae movements are essential, as they form the foundation of the sport and are fundamental for success in competitions. Therefore, there is a growing demand for effective and objective evaluation methods to assess the quality and accuracy of players' Poomsae movement. Despite its popularity, less attention has been given to using technologies such as computer vision to assess Taekwondo movements.

There is no systematic way to track players' performance and traditional evaluation methods for Taekwondo movements rely heavily on subjective visual observation, which can lead to errors and biasedness in judgments. Accurate execution of these movements is

[*] Corresponding author (michelleufernando@gmail.com; https://orcid.org/0009-0000-4911-6272)

[†] The abstract was presented at the 23rd International Conference on Advances in ICT for Emerging Regions (ICTer 2023), 8- 9 November 2023, University of Colombo School of Computing, Colombo, Sri Lanka.

also crucial for success in competitions. Figure 1 is an illustration of some movement guidelines given by the World Taekwondo Federation (WTF, 2014) for Poomsae movements.



Walking Stance (*Ap seogi)*

- 3 foot-length from origin
- Front foot facing straight forward
- Back foot 30 degrees
- Both legs straight
- Inner sides of both feet should be on a straight line
- Body turned 45 degrees (natural angle)
- Weight evenly distributed on both legs

Figure 1: WTF guidelines for walking stance in Poomsae (WTF, 2014)

Taekwondo training typically involves multiple sessions per week, each lasting for several hours. With each session coach dealing with more than ten students at a time, it can be challenging to provide individualized attention to each player. This can result in less time devoted to mastering specific movements, and slower the overall progress in skill development. Therefore, self-training is a must to sharpen the skills, and unavailability of a self-paced movement evaluation method is a serious problem in Taekwondo.

Contribution of the paper

In response to the above challenges, this paper presents a method that involves the skeleton data points of players to assess the quality and accuracy of the movements, using supervised deep learning (DL). It utilizes variations in skeleton points data to evaluate movements and distinguish correct from incorrect movements. This approach leverages the power of machine learning (ML) algorithms to accurately identify and classify the quality of movements based on the observed patterns in the skeleton data.

Organization of the paper

The remaining sections of the paper are organized as follows: the next sections delve into existing literature, providing a comprehensive overview of prior research in the field. Following this, the Methodology section details the approach to handling data, outlining the design and development of the classification model employing LSTM networks. Subsequently, the Results and Discussion section presents experimental findings, including model evaluation metrics, and discusses results in conjunction with domain experts' assessments. Finally, the Conclusion summarizes key findings of the study, referring insights presented throughout the paper.

Player performance evaluation

When it comes to player performance evaluation using computerized systems, Human action/activity recognition (HAR) plays a major role. HAR can be referred to as the art of identifying and naming activities from a video containing complete action execution (Kong & Fu, 2022), using artificial intelligence (AI) from the gathered activity raw data, by utilizing various kinds of hardware devices such as wearable sensors, electronic device sensors, camera devices (Kinect, Orbbec, CCTV), and also some commercial off-the-shelf (COTS) equipment like a webcam with an open-source software solution to detect movements.

HAR can be done using various data modalities like Red-Green-Blue (RGB) depth images, skeleton points, infrared, radar, and WiFi signals. Skeleton point detection is more cost-effective compared to other data modalities (Sun *et al*., 2022). Insensitivity to the background, use of COTS hardware devices, and the availability of open-source software are some of the advantages.

There are many different approaches to HAR, including rule-based dynamic models and machine-learning methods (Vrigkas *et al*., 2015). Each approach has its strengths and weaknesses, and the choice of method depends on the specific application. In recent years, deep learning methods such as convolutional neural networks

(CNNs) and Long-Short Term Memory (LSTM) networks have shown great promise in HAR tasks, achieving state-of-the-art results on many benchmark datasets (Sun *et al*., 2022). HAR has been widely researched in the computer science field, with applications ranging from surveillance and robotics to healthcare and sports. In sports, it has been used to analyze and evaluate various athletic performances, such as gymnastics, dance, and martial arts (Emad *et al*., 2020; Host & Ivašić-Kos, 2022).

Use of skeleton joint tracking technology in HAR

In the field of HAR, the use of skeleton joint tracking technology has gained significant attention in recent years. This technology involves tracking the movement of human joints using various sensors and cameras. The resulting data can be used to analyse and understand the biomechanics of human motion, which can be applied in fields such as sports, healthcare, and rehabilitation as described by Vrigkas *et al*., (2015). It offers non-intrusive, cost-effective, and convenient solutions for evaluating body posture and movement. The use of skeleton point data obtained from pose estimation algorithms and motion capture systems has the potential to revolutionize action recognition and motion analysis.

A paper by Chung *et al*., (2022) provides a comprehensive overview and analysis of various state-of-the-art techniques for skeleton-based human pose estimation algorithms. The authors review several algorithms including OpenPose, PoseNet, MoveNet, and MediaPipe-Pose that use different methods, such as DL, graphical models, and optimization techniques, to estimate the human body's 2D or 3D pose from images or videos.

The techniques are evaluated using several benchmark datasets, and their strengths and limitations are discussed in detail. The paper also presents a comparative analysis of these techniques, including their accuracy, computational complexity, and suitability for real-time applications. The authors highlight the importance of considering factors such as pose variability, occlusion, and image resolution when selecting a technique for a specific application. The authors indicate that MediaPipe has performed well in video data.

ML models for player performance evaluation

A paper by Zhang *et al*. (2012) proposes a system to recognize and segment the postures of golf swings consisting of two main components: a Gaussian mixture model (GMM) and a support vector machine (SVM). The GMM was used to detect the golfer's body parts. The detected body parts were then used to calculate the angles and positions of the golfer's body during the swing.

The SVM was trained on the extracted features to classify the golf swing into one of four categories: perfect, good, average, or poor. The experimental results showed that the proposed system achieved an accuracy of 85% in classifying the golf swing into one of the four categories. It can provide a quantitative evaluation of golf swings and identify areas for improvement. One of the weaknesses of this system is its use of the Kinect sensor, which is relatively expensive for general-purpose usage.

In a paper by Piergiovanni & Ryoo (2018), the authors proposed a model for recognizing fine-grained actions in baseball videos using a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The authors demonstrated that their model outperformed several baseline models. However, further research is needed to explore the generalizability of the model to other types of video data. Specifically, the model consists of two main parts: a feature extractor and a classifier.

The classifier is an LSTM network that takes as input the sequence of feature vectors, and outputs a probability distribution over the fine-grained action labels. To train the model, the authors used a large dataset of baseball videos that had been annotated with fine-grained action labels. They trained the model using a cross-entropy loss function and the Adam optimizer. However, it has not used skeleton points for the classification.

In a research paper, Zhao *et al*., (2019) propose a DL model for recognizing human actions using skeleton point data. The paper addresses the challenge of recognizing human actions from skeleton point data by exploiting Bayesian Graph convolutional long short-term memory (GC-LSTM) networks. The proposed model consists of two main components: a graph convolutional layer and a Bayesian GC-LSTM layer that applies graph convolutions on the skeleton data to capture the spatial relationships between the joints. The Bayesian GC-LSTM layer combines the temporal dependencies of the skeleton data with uncertainty modelling using Bayesian inference. The authors evaluated the proposed model on three publicly available datasets: the NTU RGB+D dataset, the Kinetics-Skeleton dataset, and the SBU Interaction dataset. The experimental results showed that the proposed model performed well on all three datasets.

Liu *et al*., (2017) proposed a skeleton-based human action recognition approach utilizing global context-aware attention LSTM networks. The method addresses the challenge of capturing spatial and temporal information from skeleton data by incorporating a context-aware attention mechanism into LSTM networks.

This enables the network to focus on relevant joints while considering the overall context of the action. The approach achieves improved accuracy in action recognition, as demonstrated through experiments on benchmark datasets. However, limitations include the reliance on accurate skeleton joint tracking and the potential sensitivity to noise or missing joint information, which may impact performance in real-world scenarios. Further research is needed to address these limitations and enhance the method's applicability.

According to the literature review based on HAR for martial arts performance evaluation, there are some existing approaches for classifying taekwondo and other martial arts movements using videos.

The Taekwondo unit technique human action dataset with key frame-based CNN action recognition (TUHAD) paper (Lee & Jung, 2020) defined a reliable Taekwondo Poomsae movement dataset called TUHAD and proposed a key-frame-based Convolutional Neural Networks (CNN) architecture to recognize Taekwondo actions using their dataset. TUHAD has 1936 action samples of eight unique actions performed by ten individuals and recorded from front and side views. The suggested model achieved recognition accuracy of up to 95.83%, according to a correlation analysis of the input configuration and precision.

A paper by Barbosa *et al*., (2021) compared 4 different deep-learning approaches to classify Taekwondo movements. To determine which methodology or technique yields the greatest results, it was tested using a dataset that had already been created. It was determined that convolution layer models, such as CNN plus long short-term memory (LSTM) and convolutional long short-term memory (ConvLSTM) DL models, give more than 90% accuracy. It produced the best results given the movements, which were typically fast and frequently high-leg movements.

The paper by Liang & Zuo (2022) proved that movement classifications from videos and mapping them to score a Taekwondo sparring (fighting) match are also possible. They introduced a graph convolution framework to recognize, segment, and evaluate Taekwondo actions with a specific part of the perception structure. The identified Taekwondo movements are marked in a time series using LSTM, and feature extraction is done at the graph convolution level to acquire the spatial and temporal correlations between skeleton joints. Using the manually labelled database, it predicted the action class and then matched the score for each movement. Finally, it is validated using Taekwondo competition data videos. This approach has an average action identification accuracy of 90% and an average action score matching rate of 74.6%.

Emad *et al*. (2020) proposed a smart coaching system called iKarate for Karate training. It offers a system that will track the players' movements using an IR camera sensor and classify the data (after a pre-processing phase) using the fast dynamic time warping algorithm. As an output, it will generate an accurate report that includes every action the player performed, mistakes made in each movement, and suggestions to improve movements.

A paper by Cunha *et al*. (2021) proposed a user-friendly and affordable approach for quickly evaluating the performance of Taekwondo competitors. They have created a mobile application to monitor the movements made by the player throughout a training session. Sometimes, occlusion occurs due to the rotation of the player or the overlapping of a limb. To overcome this issue, they have used motion sensors and an inertial measurement unit for data collection. Then it will be transmitted to the mobile app through Wi-Fi. Hardware components will be fixed to the players through Velcro strips. They used a special lab environment to conduct the study with the relevant hardware devices.

Identified limitations of existing systems

Upon careful analysis of the literature review, it is discerned that the following limitations exist within current systems: using relatively expensive hardware devices (sensors, IR trackers, Kinect camera), lack of skeleton point data usage for movement evaluation, classifying only one movement without any way to detect a sequence of actions, and the need to set up a special lab environment. Currently there is no systematic way to evaluate a sequence of Poomsae movements.

MATERIALS AND METHODS

The study followed the design science research methodology (vom Brocke *et al*., 2020). Design science methodology is a systematic approach that aims to develop and evaluate innovative solutions to practical problems. The process involves several stages starting with problem identification, followed by the defined objectives, design and development, demonstration, evaluation, and communication of a solution. Figure 2 illustrates the alignment of each step in the research with the design science approach adopted throughout the study. The methodology emphasizes the importance of iteration and feedback in the design and development process, as well as the need for iterative and rigorous evaluation of the solution's effectiveness.

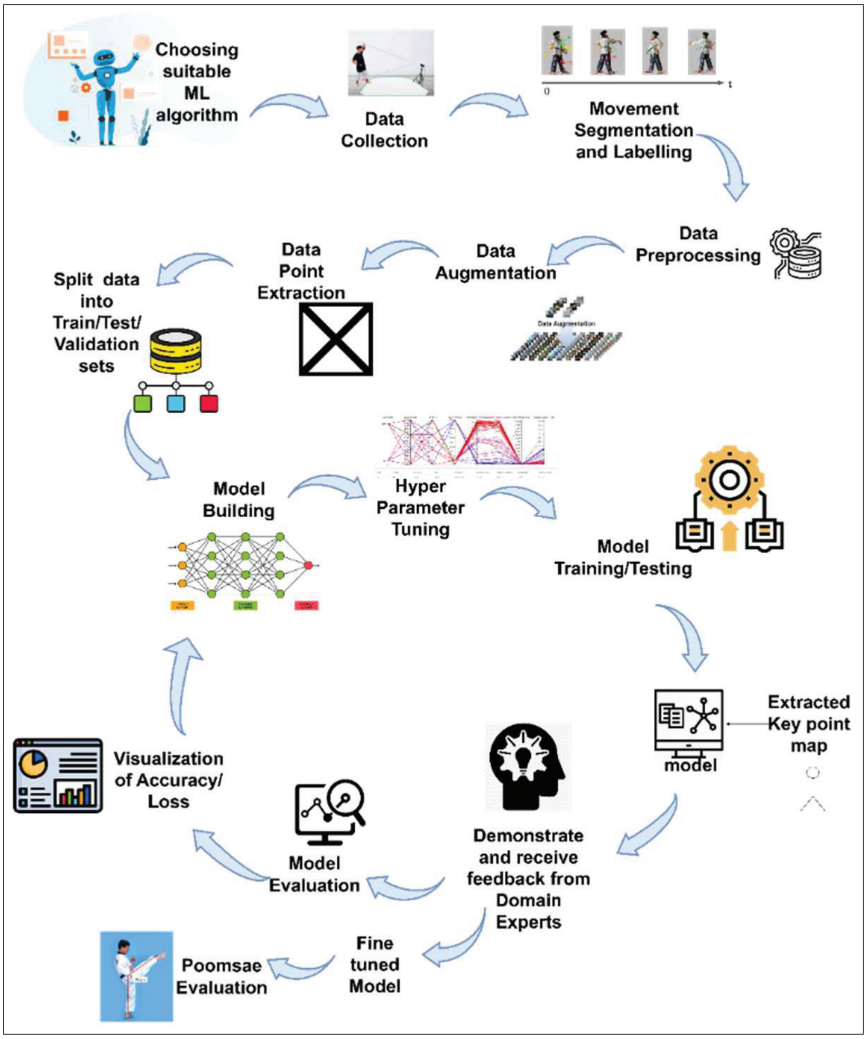


Figure 2: Designing and evaluation of the solution

Choosing suitable ML/DL algorithm

The problem of Taekwondo movement classification using skeleton point data can be mapped to a classification problem, which is a widely researched area in human action recognition. However, due to the sequential and temporal nature of the movement data, simple ML models are not suitable for accurately classifying the movements. Therefore, there is a need for more complex models such as artificial neural networks (ANNs) to handle this video classification task.

Convolutional neural networks (CNNs) are commonly used for image recognition tasks, but they are not well-suited for temporal data analysis, because the model has to store the previous state of the frame to process the current state of the frame. Recurrent neural networks (RNN) are designed to handle sequential data by processing the data sequentially, and they can remember previous inputs using hidden states. However, traditional RNNs may suffer from the problem of vanishing gradients during training, which can affect their performance in video kind of data classification. (Olah, 2015; Mittal, 2019)

Long short-term memory (LSTM) is a type of RNN that addresses the problem of vanishing gradients by introducing memory cells and gates. LSTM has been shown to be effective in dealing with dynamic temporal datasets, making it a suitable model for movement classification problems using skeleton point data (Barbosa *et al*., 2021). Therefore, in this research, LSTM was selected as the best classification technique for accurately classifying Taekwondo movements based on skeleton point data.

Data collection

A collection of recorded videos was obtained from players of the University of Colombo (UOC) Taekwondo team. Before recording videos of each player, the research objectives were explained to the players, and their consent obtained to participate in the study. Fifteen members of the UOC Taekwondo team volunteered, comprising seven senior players and eight junior players. A senior player means he or she should have membership in the Sri Lanka Taekwondo Federation (SLTF), should have a blue or above belt (till 6th GUP), have experience in participating in Poomsae competitions for more than two years, and have won a medal from at least one Poomsae event. A Junior player means he or she should have membership in SLTF, should have a junior green or above belt (8th GUP or below), have experience in participating in one poomsae competition, but may or may not have won a medal from a Poomsae event. Therefore, it was confirmed that all participants were aware of performing a Poomsae.

Both correctly and incorrectly performed movements were captured, providing a comprehensive dataset for analysis. Initially, over 750 videos were collected, encompassing a wide range of skill levels. A 1 m tripod stand and an iPhone XS max mobile phone camera were used for recording videos. (Camera specifications: iPhone XS max, 12-megapixel (f/1.8, 1.4-micron),1920 * 1080 resolution, 30 fps).

Movement segmentation and labelling

The whole performance of the series of Poomsae movements by a player was segmented into short video clips that contained only one movement. The segmentation part was done manually, and, in this phase, each video clip was labelled as to whether it was corrected or not with the help of domain experts (okankop, 2020). The 'incorrect' labelled data pertains to the commonly occurring mistakes made by Poomsae players. Collecting data from different players with different skill levels helped to increase the diversity of the dataset and made the model more robust and generalizable.

Figure 3 shows a player performing 4th movement. In the left side image, the stance is incorrect, and the punch is correct. Therefore, the movement is labelled as i4 (incorrect movement 4). Figure 3 right side image shows a player performing the same movement. The stance is correct, and the punch is correct. Therefore, the movement is labelled as c4 (correct movement 4).

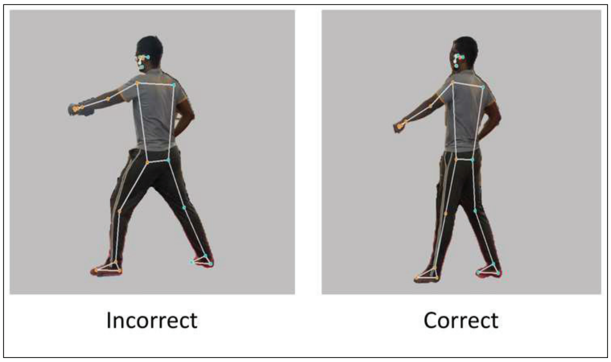


Figure 3: Player performing the fourth movement in two different ways

Data pre-processing

Segmented videos contained different frame numbers. However, the LSTM model requires the same sample size when it comes to the model training process. Therefore, bringing all the video clips into the same frame rate was necessary.

The approach involved using a video processing library to standardize the frame rate and duration of the video clip. Specifically, the frame rate was adjusted to 30 frames per second (fps) with a consistent one-second duration per video clip. This process ensured temporal uniformity, facilitating precise and systematic analysis of the video content. The method was chosen to provide a reliable foundation for research and evaluation. A simple Python script written with the 'moviepy' library was used.

Data augmentation

The data augmentation technique was utilized to increase the size of the dataset and improve the accuracy of the model as described by Park & Sohn (2020). To achieve this, salt and pepper noise and colour inverter techniques were applied. These techniques helped to diversify the dataset by adding additional variations to the original images (Nida *et al*., 2022). The resulting augmented dataset was then mixed with the original dataset and increased to 1500+ video clips. That dataset was used for training the DL model. The 'Vidaug' (okankop, 2020) Python library was used for the augmentation process.

Skeleton point data extraction

Skeleton point data was extracted using the MediaPipe Pose (Google, n.d.) solution. It is a Python library that provides highly accurate landmark (skeleton joint) detection. The extracted coordinates were stored as binary files in a well-defined folder structure, considering read/write efficiency. Figure 3 shows how the MediaPipe Pose detection library identifies human skeleton points.

The dataset contained 8 different action classes with 188 videos for each class, resulting in a total of 1504 videos. Each video has a one-second duration and consists of 30 frames. When extracting key points, one frame had 132 coordinates, which corresponded to 33 joints and 4 coordinates (x and y coordinates of the point in the image frame, the depth or z coordinate, and a visibility value that indicates whether the landmark is visible in the image or not) for each joint. Therefore, each video clip contained 132 * 30 key points. Therefore, the total number of key points in the dataset can be calculated as:

1504 videos * 30 frames per video * 33 joints per frame * 4 coordinates per joint = 5,955,840 key points.

The collected data was divided into two sets of datasets: training and testing. Random seed was set to 42, to ensure that the data would be split into the same training and testing sets each time. The number 42 is often used as a random seed in ML because it is a reference to the book "The Hitchhiker's Guide to the Galaxy," in which 42 is famously referred to as the "Answer to the Ultimate Question of Life, the Universe, and Everything."

Hyper parameter tuning

Hyperparameter tuning was done for the Long Short-Term Memory (LSTM) network by adjusting the number of layers, the number of neurons per layer, dropout rate, activation function, and learning rate for different model architectures and checking the validation accuracy increment in several attempts.

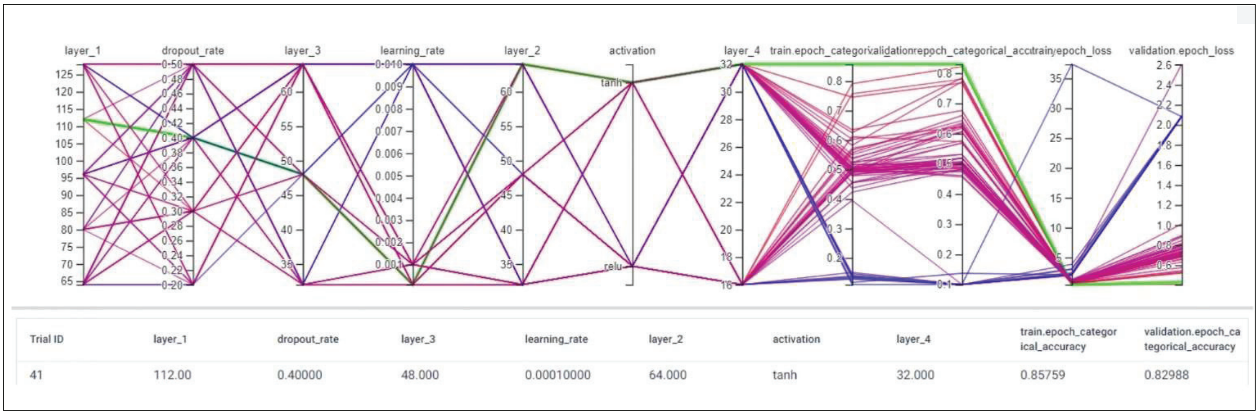
Selected hyperparameters for the tuning process are listed below.

- Number of layers - deep neural networks (DNN) should use a minimum number of Layers to keep the generalized behaviour as mentioned in the papers by Hobs (2015) and Eckhardt (2018).
- Number of neurons per layer – The number of neurons in the input layer is the size of the input data (features), and the number of neurons in the output layer should be equal to the number of classes, The number of neurons in the hidden layers is determined by the complexity of the problem as discussed by Heaton (2008).
- Activation function - Tanh (hyperbolic tangent) and Relu (rectified linear unit) are widely used for classification tasks when using neural networks. Tanh maps the input to the range of -1 to 1 while Relu maps the input to the range of 0 to infinity (Lee & Jung, 2020).
- Dropout rate – Adding dropout layers prevents overfitting. Typically, the dropout rate value between 0.1 and 0.5 performs well in video data classification (Brownlee, 2017).
- Learning rate - controls the step size at each iteration and smoothens the convergence (Eckhardt, 2018).
- Optimizer - Adam optimizer adjusts the learning rate adaptively during training, allowing it to converge faster than other optimization algorithms.

Model building

Initially, the model was designed with an LSTM input layer and a dense output layer. The number of hidden layers was decided with the hyperparameter tuning as described in the previous section. Two hidden LSTM layers and two hidden dense layers were used. Each layer contains a dropout layer to avoid overfitting. The final model was designed with the hyperparameters chosen in the above section. Figure 4 indicates how the hyperparameter tuning process was carried out using a parallel coordinate graph.

Table 1 and Figure 5 visualize the deep learning model architecture. The input layer of the neural network comprises 132 neurons, which correspond to the 132 coordinates utilized as features for the classification task. Notably, these coordinates serve as the fundamental input attributes for the network's classification model. It is important to highlight that the sample size is precisely 30. This sample size is representative of a single video clip, each of which contains 30 frames. Each frame, in turn, encapsulates 132 distinct features. The utilization of this input structure is pivotal in enabling the network to process and classify video data effectively.



Trial ID	layer_1	dropout_rate	layer_3	learning_rate	layer_2	activation	layer_4	train.epoch_categorical_accuracy	validation.epoch_categorical_accuracy
41	112.00	0.40000	48.000	0.00010000	64.000	tanh	32.000	0.85759	0.82988

Figure 4: Hyperparameter tuning visualized in a parallel coordinate view

Table 1: Model architecture

Layers	Type & no. of neurons	Hyper parameters
Input	LSTM Layer with 132 Neurons	Input size = 33 landmarks * 4 coordinates = 132 Sample size = 30 frames
Hidden	LSTM layer with 112 Neurons Dense layer with 64 Neurons LSTM layer with 48 Neurons Dense layer with 32 Neurons	dropout rate = 0.4 activation function = tanh
Output	Dense Layer with 8 neutrons	SoftMax activation function (8 movement categories)

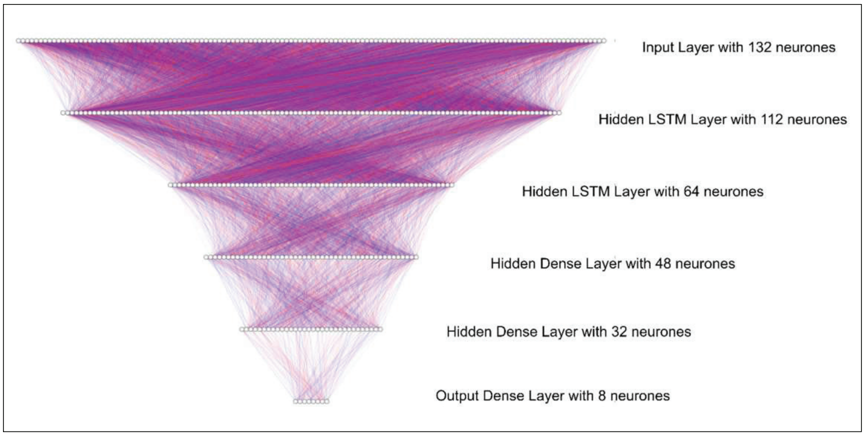

Input Layer with 132 neurones
Hidden LSTM Layer with 112 neurones
Hidden LSTM Layer with 64 neurones
Hidden Dense Layer with 48 neurones
Hidden Dense Layer with 32 neurones
Output Dense Layer with 8 neurones

Figure 5: Deep learning architecture diagram

The output layer of the neural network was configured as a dense layer, employing the SoftMax activation function. This choice of activation function was deliberate, as it offers the capacity to provide action class probabilities in the context of multi-class classification. In this movement classification task, there are a total of 8 distinct movements, and accordingly, the output layer was composed of 8 neurons. Each neuron in the output layer is responsible for generating a probability estimate for its corresponding class, thereby facilitating the classification of input data into one of the 8 defined movement classes.

The input data for the model consisted of preprocessed skeleton point data, as described in the data preprocessing section. Therefore, the first input layer has 132 neurons and has a [30, 132] input data vector. In the end, it was expected to have a multi-class classification of 8 classes, therefore the output dense layer had 8 neurons.

Model training / testing

Model training involves feeding the preprocessed and extracted skeleton point data into the ML model. The models are then trained to recognize patterns in the data. In this case, it could be defined as identifying the relevant pose within 30 frames of skeleton point data. Finally, the models are tested using a separate dataset to evaluate their accuracy and generalization ability.

The model was trained using a categorical cross-entropy loss function since this is a multi-class classification model and optimized using the Adam optimizer due to its ability towards fast convergence and adaptive learning rate. The training result is described in Table 2.

Table 2: Model training summary after 333 epochs

Result description	Score
Categorical accuracy for the training dataset	0.995
Loss for the training dataset	0.03722
Categorical accuracy for the testing dataset	0.96013
Loss for the testing dataset	0.1730

The model training and testing process was conducted using a local machine that solely utilized a CPU for training. The average duration for training the LSTM model with a minimum of four layers, using the collected dataset, was 30 to 40 minutes. However, hyperparameter tuning required more time, as it involved evaluating various potential combinations to improve accuracy. This task took an average of five hours to complete 60 trials and output a good hyperparameter combination.

Classification algorithm

In the classification algorithm, the initial step is to break down a sequence of movements into individual movements for evaluation. This means the algorithm gets an initial 30 frames as a movement and extracts key points for each frame. Then all the coordinates with respective frame numbers will be fed into the LSTM classification model. Then the model will perform the classification task. If the input coordinates lack the usual characteristics needed for classification, the algorithm follows a specific protocol. It dequeues the initial frame that was initially selected as the input frame sequence and enqueues the subsequent frame from the video file. Then again, the newly selected input frame sequence is fed to the LSTM classification model. This process is repeated for the entire video.

RESULTS AND DISCUSSION

The final movement evaluation algorithm was designed in a way that users could upload a video file of a player's Poomsae performance. The output would be the classified movement sequence, indicating whether each movement is correct or not. As shown in Figure 6 it detects the first movement as correct (c1) and in Figure 7 it detects the second movement as incorrect (i2) for the given video. The classified action sequence result is displayed in the top left corner of the window.

Model evaluation

The evaluation of the ML model can be done in two different ways. The first approach involves the use of standard classification metrics to measure the accuracy and performance of the model on a test data set. The second approach is to evaluate the model's performance with the help of a domain expert to ensure that it is addressing the identified research problem.

During the training process, the model was optimized using various hyperparameters to ensure that it was accurately detecting correct and incorrect patterns. The model achieved an average accuracy of 96% for the test data set. After training, the model was applied to a new validation data set of 11 videos and the results compared with a domain expert. The comparison resulted in the ML model classifying the correct movements with 66% accuracy as shown in the Table 4.

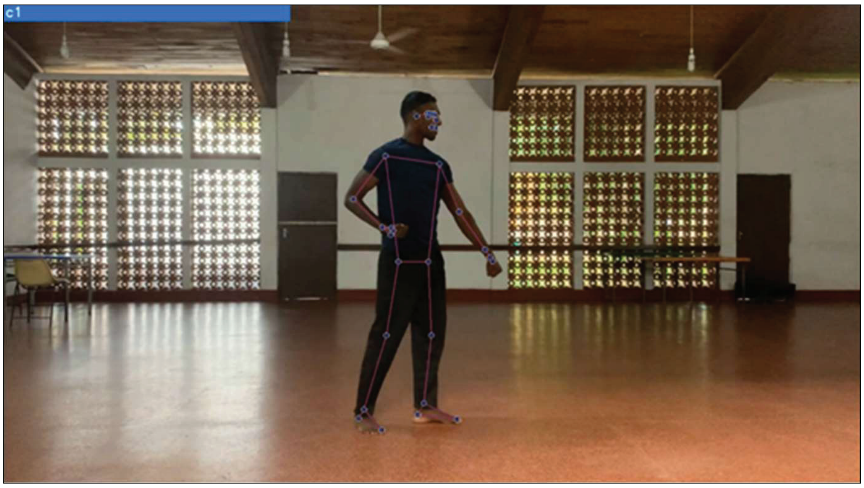


Figure 6: Screen capture obtained from the ML model output video - first movement evaluation.

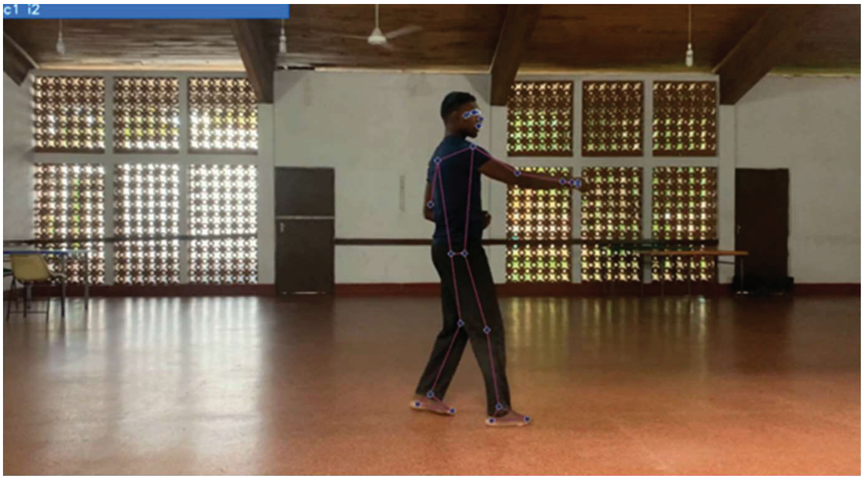


Figure 7: Screen capture obtained from the ML model output video - second movement evaluation.

Classification metrics

Accuracy is defined as the fraction of all correct predictions, which includes true positives and true negatives, out of all the predictions made by the model for a certain class of movements. However, it is important to note that high accuracy alone does not necessarily imply good model performance, as it may not reflect a class imbalance in the data. To account for this, additional evaluation measures such as precision, recall, and the F1 score are commonly used.

These measures provide insights into the model's performance concerning correctly identifying the positive class (precision) and capturing all positive instances in the data (recall), and the harmonic mean of these two measures (F1 score) provides a more balanced view of the model's performance.

The utilization of a neural network architecture with four hidden layers resulted in the highest achieved accuracy of 96%. This configuration also demonstrated a consistent and smooth convergence in the train/test accuracy, as visually depicted in Figure 8. Whereas, Figure 9 illustrates a progressive reduction in both training and testing loss over the course of more than 300 epochs.

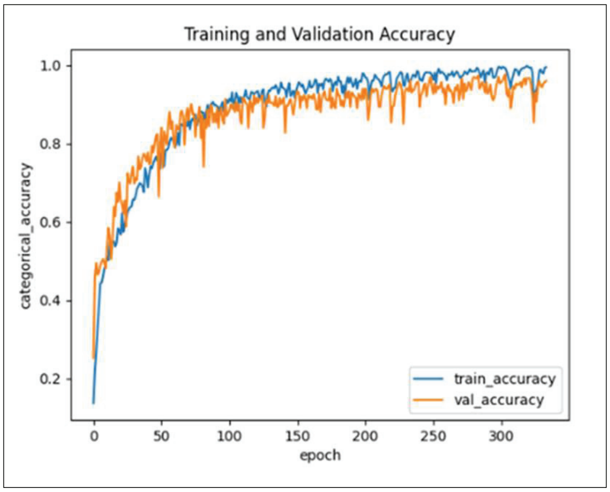


Figure 8: Training and testing (validation) accuracy for the LSTM Model

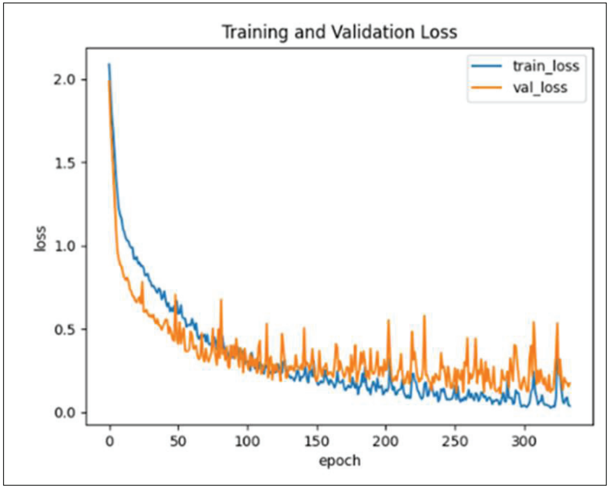


Figure 9: Training and testing (validation) Loss for the LSTM Model

The test dataset comprises a total of 301 video clips. The classification report depicted in Table 3 provides an overview of the performance metrics of the trained model on the test dataset. The report includes accuracy, precision, recall, and f1-score measures, which are essential for evaluating the effectiveness of the model.

Table 3: Classification report containing precision, recall, and f1-score for each movement class

	Precision	Recall	F1-score	Support
c1	0.83	0.94	0.88	32
c2	1.00	0.95	0.98	43
c3	0.97	1.00	0.99	37
c4	0.97	1.00	0.99	37
i1	0.94	0.85	0.89	39
i2	0.95	1.00	0.98	40
i3	1.00	0.98	0.99	42
i4	1.00	0.97	0.98	31
Accuracy			0.96	301
Macro Average	0.96	0.96	0.96	301
Weighted Average	0.96	0.96	0.96	301

Model evaluation with the domain experts

Table 4 simply presents the evaluation results of a Poomsae movement by the domain expert and the ML model. The experiments were done using 11 validation video files from several Poomsae players. Domain experts' results are considered the 'ground truth' when evaluating the model performance.

Table 4: ML model evaluation results with the domain experts

Video No.	Domain expert's * result sequence				ML model classification sequence				No. of correctly detected movements by the ML model	Detection percentage
1	c1	c2	i3	c4	i1	i2	c3	c4	1	25%
2	i1	i2	c3	i4	i1	i2	i4	i4	3	75%
3	i1	c2	i3	i4	i1	c2	c3	c4	2	50%
4	c1	c2	c3	c4	c1	c2	c3	c4	4	100%
5	i1	c2	c3	i4	i1	c2	c3	c4	3	75%
6	c1	c2	i3	i4	c1	i2	i3	i4	3	75%
7	i1	i2	c3	i4	i1	i2	i4	i4	3	75%
8	c1	i2	c3	i4	i1	c2	c3	c4	1	25%
9	i1	c2	i3	c4	i1	i2	c3	i4	1	25%
10	i1	c2	c3	c4	i1	c2	c3	c4	4	100%
11	i1	i2	i3	i4	i1	i2	i3	i4	4	100%
Average Accuracy										66%

[*] Domain expert: an experienced individual in Poomsae evaluation, and their assessments (holding a Black Belt 6th Dan (Kukkiwon) and also an International Referee of the World Taekwondo Federation.) participated in this study.

The letter 'c' indicates that the movement is correct, and 'i' indicates that the movement is incorrect. The numbers 1 to 4 indicate what the movement was that the player performed. The last two columns of the above table list how many of the movements were correctly evaluated by the ML model and then the classification percentage for a given video by the ML model. Finally, the accuracy of the model for the validation data set was presented as a percentage. The ML model achieved 66% of the average accuracy score for the given dataset of 11 videos.

CONCLUSION

In conclusion, this study aimed to evaluate Taekwondo movements using an LSTM-based ML model in terms of its classification performance. In the evaluation of the LSTM model on the test dataset, it achieved a noteworthy accuracy of 96%. When benchmarked against domain experts, the model maintained an average accuracy of 61%. This comparative performance highlights the model's robustness and potential applicability in real-world scenarios. The substantial accuracy suggests that the proposed classification model could serve as a valuable tool for assessing player performance.

The model was able to classify the performance of the movement by observing a set of frames and considering the behaviour of the movement, rather than solely focusing on the terminal pose. It was noticed that this aspect of the ML model could be useful in Poomsae movement evaluation, as it can capture not only the accuracy of the movement but also the strength and expression of energy. While this study focused only on the accuracy of the Poomsae movements, future work could consider these additional aspects in the evaluation process. Furthermore, the model was tested on various video data to ensure its generalizability to real-world scenarios. The feedback of domain experts was also considered to fine-tune the models for better performance.

A major research problem addressed in this study was the issue of subjectivity in the domain of Poomsae evaluation using traditional methods. The ML model performed well in the context without human intervention, indicating that improving these models could lead to a systematic approach to evaluating Poomsae by eliminating human subjectivity. By providing a reliable and objective evaluation tool, this research project contributes significantly to the advancement of the Poomsae evaluation domain.

The lack of self-evaluating methods was identified as another problem that this study aimed to address. The research project successfully developed a model that could be easily integrated with mobile apps or web apps as a self-paced approach for Poomsae evaluation, utilizing commercial off-the-shelf hardware devices and open-source software solutions.

Additionally, the model was insensitive to background distractions, ensuring that the focus remains on the subject of interest. Skeleton point data optimized train/test data storage by requiring less space compared to traditional video file preservation methods, reducing storage costs and resource requirements.

Looking ahead, several exciting avenues exist for extending and enhancing the impact of this research. One promising direction involves the implementation of models tailored for real-time performance evaluation, potentially revolutionizing how movements are assessed as they occur. Additionally, a diverse range of datasets will be employed to further test and refine the models, ensuring applicability and robustness across varied contexts. Furthermore, this research can be extended for a deeper exploration into the capabilities of the ML model for evaluating not only basic movement accuracy but also other criteria such as the expression of energy and presentation skills. These efforts not only aim to improve current research, but also lead to new uses and understandings in movement analysis.

Acknowledgments

Gratitude is extended to Dr Hiran Ekanayake, Dr Enosha Hettiarachchi, and Dr Gihan J. Mendis for providing valuable feedback to carry out the research project. Special thanks to Mr. Priyantha Kumara and UOC Taekwondo team players for providing videos, and essential domain knowledge and helping with model evaluation.

REFERENCES

Barbosa P., Cunha P., Carvalho V. & Soares F. (2021). Classification of taekwondo techniques using deep learning methods: First insights. *Proceedings of the BIODEVICES 2021 - 14th International Conference on Biomedical Electronics and Devices;* Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, 11–13 January, Vienna, Austria, pp. 201–208.

Brownlee J. (2017). Dropout with LSTM networks for time series forecasting. In: *Machine Learning Mastery*. Available at *https://machinelearningmastery.com/use-dropout-lstm-networks-time-series-forecasting/*

Chung J.-L., Ong L.-Y. & Leow M.-C. (2022). Comparative analysis of skeleton-based human pose estimation. *Future Internet* **14**(12): 380.
DOI: https://doi.org/https://doi.org/10.3390/fi14120380

Cunha P., Barbosa P., Ferreira F., Fitas C., Carvalho V. & Soares F. (2021). Real-time evaluation system for top taekwondo athletes: project overview. *Proceedings of the BIODEVICES 2021 - 14th International Conference on Biomedical Electronics and Devices;* Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, 11–13 January, Vienna, Austria, pp. 209–220.
DOI: https://doi.org/10.5220/0010414202090216

Eckhardt K. (2018). *Choosing the right Hyperparameters for a simple LSTM using Keras*. Available at *https://towardsdatascience.com/choosing-the-right-hyperparameters-for-a-simple-lstm-using-keras-f8e9ed76f046*

Emad B., Atef O., Shams Y., El-Kerdany A., Shorim N., Nabil A. & Atia A. (2020). Ikarate: Karate Kata guidance system. *Procedia Computer Science* **175**(2019): 149–156.
DOI: https://doi.org/10.1016/j.procs.2020.07.024

Google for Developers. (2023). *Pose landmark detection guide*. Available at *https://developers.google.com/mediapipe/solutions/vision/pose_landmarker*

Heaton J. (2008). *Introduction to Neural Networks with Java*. Heaton Research, Cop.

Hobs. (2015). How to choose the number of hidden layers and nodes in a feedforward neural network? In *Stackexchange*. Available at *https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw/136542#136542*

Host K. & Ivašić-Kos M. (2022). An overview of human action recognition in sports based on computer vision. *Heliyon* **8**(6): e09633.
DOI: https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e09633

Kong Y. & Fu Y. (2022). Human action recognition and prediction: a survey. *International Journal of Computer Vision* **130**(5): 1366–1401.
DOI: https://doi.org/10.1007/s11263-022-01594-9

Lee J. & Jung H. (2020). TUHAD: Taekwondo unit technique human action dataset with key frame-based CNN action recognition. *Sensors* **20**(17): 4871.
DOI: https://doi.org/https://doi.org/10.3390/s20174871

Liang J. & Zuo G. (2022). Taekwondo action recognition method based on partial perception structure graph convolution framework. *Scientific Programming* **2022**: 1838468.
DOI: https://doi.org/10.1155/2022/1838468

Liu J., Wang G., Hu P., Duan L. Y. & Kot A. C. (2017). Global context-aware attention LSTM networks for 3D action recognition. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21–26 July, Honolulu, HI, USA, 3671–3680.
DOI: https://doi.org/10.1109/CVPR.2017.391

Mittal A. (2019). Understanding RNN and LSTM. Available at *https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e*

Nida N., Yousaf M. H., Irtaza A. & Velastin S. A. (2022). Video augmentation technique for human action recognition using genetic algorithm. *ETRI Journal* **44**(2): 327–338.
DOI:https://doi.org/https://doi.org/10.4218/etrij.2019-0510

okankop. (2020 May). *okankop/vidaug*. GitHub. Available at *https://github.com/okankop/vidaug*

Olah C. (2015, August). *Understanding LSTM Networks – Colah's Blog.* Available at *https://colah.github.io/posts/2015-08-Understanding-LSTMs/*

Park C.-I. & Sohn C.-B. (2020). Data augmentation for human keypoint estimation deep learning based sign language translation. *Electronics* **9**(8): 1257.
DOI: https://doi.org/https://doi.org/10.3390/electronics9081257

Piergiovanni A. J. & Ryoo M. S. (2018). Fine-grained activity recognition in baseball videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 18–22 June, Salt Lake City, UT, USA, pp. 1821–18218.
DOI: https://doi.org/10.1109/CVPRW.2018.00226

Sun Z., Ke Q., Rahmani H., Bennamoun M., Wang G. & Liu J. (2022). Human action recognition from various data modalities: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(3): 3200–3225.
DOI: https://doi.org/https://doi.org/10.1109/tpami.2022.3183112

Vrigkas M., Nikou C. & Kakadiaris I. (2015). A review of human activity recognition methods. *Frontiers in Robotics and Artificial Intelligence* **2**: 28.
DOI: https://doi.org/10.3389/frobt.2015.00028

Vom Brocke J., Hevner A. & Maedche A. (2020). Introduction to design science research. In: *Design Science Research* (eds. J. vom Brocke, A. Hevner & A Maedche), pp. 1–13. Springer International Publishing, New York, USA.
DOI: https://doi.org/10.1007/978-3-030-46781-4_1

Zhang L., Hsieh J.-C., Ting T.-T., Huang Y.-C., Ho Y.-C. & Ku L.-K. (2012). A Kinet based golf swing score and grade system using GMM and SVM. *2012 5th International Congress on Image and Signal Processing*, 16–18 October, Chongqing, China, pp. 711–715.
DOI: https://doi.org/10.1109/CISP.2012.6469827

Zhao R., Wang K., Su H. & Ji Q. (2019). Bayesian graph convolution LSTM for skeleton based action recognition. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 27 October–02 November, Seoul, South Korea, pp. 6881–6891.
DOI: https://doi.org/10.1109/ICCV.2019.00698

WTF (2014). *Poomsae Scoring Guidelines for International Referees.* World Taekwondo Federation. Available at *https://d17nlwiklbtu7t.cloudfront.net/983/document/Poomsae_scoring_guidelines.pdf*