Deep Learning-Based Autoencoder for Objective Assessment of Taekwondo Poomsae Movements

Mohamed Amine Chaâbane[1] [a], Imen Ben Said[1,2] [b] and Adel Chaari[3]
[1]MIRACL Laboratory, University of Sfax, Tunisia
[2]Digital Research Center of Sfax, Technopark of Sfax, PO Box 275, 3021 Sfax, Tunisia
[3]hejAL! New Tech Solutions GmbH, Bad Dürkheim, Germany

Keywords: Poomsae Movements, Assessment, Deep Learning, Autoencoder, Skeleton Body Points, Computer Vision.

Abstract: Artificial Intelligence (AI) is revolutionizing sports by enhancing performance, improving safety, and creating richer fan experiences. This paper focuses on leveraging AI in Taekwondo, specifically in assessing athlete performance during Poomsae movements, which are foundational to the sport and crucial for success in competitions. Traditionally, the evaluation of Poomsae has been subjective and heavily reliant on human judgment. This study addresses this issue by automating the assessment process. We propose a deep learning approach that utilizes computer vision to analyse athletes' movements captured in video clips of Poomsae. The proposed approach is based on a model that emphasizes the use of autoencoders for training data representing skeleton body points of correct movements. This model can effectively identify anomalies, i.e., incorrect movements by athletes. The SportLand platform implements the proposed approach, providing coaches and athletes with precise and actionable insights into their performance. This platform can serve as an assistant for self-evaluation, allowing Taekwondo athletes to enhance their skills at their own pace.

1 INTRODUCTION

Nowadays, the use of AI-driven systems in sports, including machine learning (ML), computer vision, and data analytics, has led to a significant paradigm shift in athlete management. These systems transform the landscape of sports science, significantly supporting real-time decision-making in coaching (Pisaniello, 2024). Recently, integrating AI technologies to enhance athlete performance has become a key focus of research (Hong et al., 2021; Michalski et al., 2022).

Taekwondo is a widely practiced martial art, with Poomsae as a fundamental component that consists of choreographed patterns embodying the core techniques and principles of the art. Traditionally, Poomsea movement evaluation relies mainly on subjective visual observation, which can lead to errors and bias in judgment. As a result, there is a growing demand for automated solutions that offer effective and objective methods to assess the quality and accuracy of athletes' Poomsae movements.

Computer vision has emerged as a promising solution to address the subjectivity often associated with manual evaluation methods. Specifically, human action/activity recognition (HAR) plays a crucial role in identifying and labelling activities from videos that capture complete actions (Kong & Fu, 2022) by utilizing AI and various hardware devices, such as cameras. Due to its cost-effectiveness, Sun et al. (2022) argue that skeleton point detection is a more effective data modality for HAR than other modalities, such as red-green-blue (RGB) depth images. The authors also assert that deep learning DL methods, such as convolutional neural networks (CNNs), autoencoders, and long short-term memory (LSTM) networks, have demonstrated significant potential in HAR tasks.

HAR has been widely researched in sports, with its use to analyze and evaluate various athletic performances (Host & Ivašić-Kos, 2022). This paper proposes an AI-based approach to assessing Taekwondo movements, thereby avoiding the subjectivity inherent in traditional evaluation methods. This approach can

[a] https://orcid.org/0000-0002-6831-5530
[b] https://orcid.org/0000-0001-5025-6715

Chaâbane, M. A., Ben Said, I. and Chaari, A.
Deep Learning-Based Autoencoder for Objective Assessment of Taekwondo Poomsae Movements.
DOI: 10.5220/0013706100003988

In *Proceedings of the 13th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2025)*, pages 170-177
ISBN: 978-989-758-771-9; ISSN: 2184-3201

significantly enhance athlete performance by providing real-time insights into movement accuracy and kinematic features, such as force and speed.

More precisely, the proposed approach leverages the power of deep learning autoencoder models and utilizes variations in skeleton point data to evaluate Poomsae movements, effectively distinguishing between correct and incorrect movements. To achieve this, we created a dataset of videos showcasing the correct execution of Poomsae by skilled athletes. This dataset was then used to train and evaluate the autoencoder model, which achieved an impressive average accuracy of 99%.

The SportLand platform is a specific tool that implements our DL approach. It assists athletes in enhancing their performance by offering a self-training service designed to sharpen their skills.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive overview of the general context and prior research in the field, delving into the background and existing literature. Following this, Section 3 outlines our deep learning approach for evaluating Poomsae. Section 4 introduces the SportLand platform, designed for delay training to enhance athletes' performance. Finally, Section 5 summarizes the study's key findings and offers directions for future work.

2 BACKGROUND AND RELATED WORKS

2.1 Poomsae in Taekwondo Sports

In Taekwondo, a poomsae is a sequence of basic movements that encompasses both offensive and defensive techniques suitable for competition. Figure 1 depicts the movements of Taegeuk I Jang, one of the foundational poomsae in Taekwondo. This form includes essential actions such as walking and basic techniques like Makki (block) and Chagi (kick). Taegeuk I Jang consists of 18 movements, numbered from 1 to 18, as shown in Figure 1.

Furthermore, to characterize a Poomsae movement as correct, the World Taekwondo Federation (WTF, 2014) provides a set of guidelines, including:

- Pause Between Movements: Athletes should incorporate a brief pause between movements to emphasize control and precision, allowing for a clear distinction between each movement.
- Symmetrical Pattern of Poomsae Line: The Poomsae should follow a symmetrical pattern, ensuring that movements are executed evenly on both sides, reflecting the balance and harmony inherent in Taekwondo.
- Balance of Each Movement: Maintaining balance throughout each movement is crucial, as it ensures stability and effectiveness in techniques, allowing for powerful and controlled execution.

Figure 1: Taegeuk I Jang Poomsae movements.

2.2 Athlete Performance Assessment in Literature

According to the literature review based on HAR for martial arts performance evaluation, several existing approaches utilize video analysis to assess athletes' movements based on skeleton points data.

A paper by Lee and Jung (2020) introduces a reliable Taekwondo Poomsae movement dataset called TUHAD and proposes a key-frame-based CNN architecture for recognizing Taekwondo actions using this dataset.

Barbosa et al. (2021) compare four different deep learning models to classify Taekwondo movements, aiming to identify which model yields the best results. The study found that convolutional layer models, including CNN combined with LSTM and Convolutional Long Short-Term Memory (ConvLSTM) models, achieved over 90% accuracy in classification.

Emad et al. (2020) propose a smart coaching system called iKarate for Karate training, which tracks players' movements using an infrared camera sensor. After a preprocessing phase, the system classifies the data using the fast dynamic time warping algorithm. As a result, the proposed system generates a detailed report outlining each action performed by the player, identifying mistakes in every movement, and providing suggestions for improvement.

Fernando et al. (2023) evaluate Taekwondo movements using an LSTM-based machine learning model, with a focus on classification performance. In their evaluation, the LSTM model achieved an accuracy of 96% on the test dataset and 61% on the validation dataset.

After a thorough analysis of the literature review, several limitations within the current systems have been identified. On one hand, the proposed approaches often involve costly logistical preparations, such as specialized hardware (sensors, infrared trackers, Kinect cameras) or dedicated laboratory environments. On the other hand, these contributions primarily focus on classifying sequences of actions using classification models to evaluate athletes' movements. While these models are well-suited for sequence-based tasks and effectively leverage their memory capabilities to learn from time-dependent data, they are less effective at learning efficient data representations. This is crucial for accurately evaluating movements and distinguishing between correct and incorrect actions.

3 AUTOENCODER FOR POOMSAE MOVEMENTS ASSESSMENT

This section outlines the deep learning approach we propose for assessing athlete performance. Figure 2 illustrates the general process for implementing our deep learning approach to evaluate Poomsae sequences.

In the following subsections, we will detail each step of this process.

3.1 Data Collection

The Data Collection step aims to create the dataset that the proposed model uses. To do so, we collect Poomsae sequences from diverse online sources. This includes videos available on platforms such as YouTube and martial arts websites that are carefully chosen to cover a comprehensive range of movements necessary for Poomsae sequences. In addition, we record video sequences during training sessions supervised by renowned professional trainers. These videos provide examples of movements performed by skilled athletes and offer high-quality data. Specific conditions, such as lighting and camera angles, were adhered to while recording the videos, ensuring the consistent and superior capture of the athletes' movements.

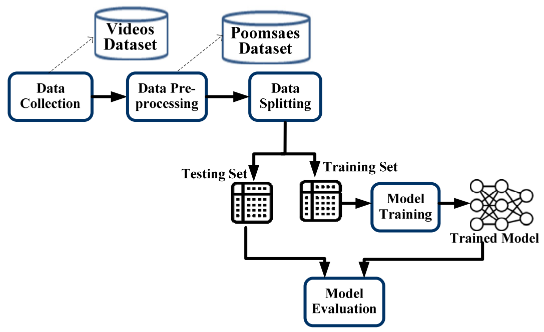


Figure 2: Overview of the Steps Involved in Implementing the Autoencoder approach for Assessing Athlete movement.

3.2 Data Pre-Processing

Data Preprocessing is crucial in transforming and preparing the data collected for effective utilization by the autoencoder model. It involves the following steps, which are detailed in the following sub-sections: Frame Extraction, Body Point Detection, Angle Calculation, Data Storage and Augmentation, and Data Scaling.

3.2.1 Frame Extraction

Each collected video was manually segmented into short video clips containing only one movement. Then, key frames were extracted from the segmented videos to define representative images of each movement in the Poomsae sequences.

It is mandatory to bring all the video clips into the same frame rate during extraction. More precisely, 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, which allows for the precise identification and analysis of specific movements.

3.2.2 Body Point Detection

Once the video frames are extracted, we leverage the MediaPipe[1] library to detect and map each frame's

[1] MediaPipe is a robust and highly accurate open-source framework for real-time pose detection, which allows for precise identification of the human body's critical anatomical landmarks.

key body joints and limb positions. Figure 3 shows an example of a body landmark driven by the MediaPipe library. From this figure, we highlight that MediaPipe provides a total of 33 landmarks for the human body, including major joints and body parts, such as shoulders, elbows, wrists, hips, knees, and ankles.

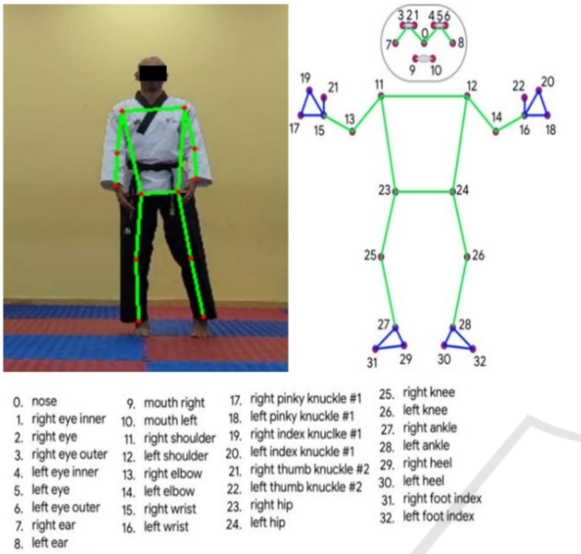


Figure 3: Example of Body Landmark Visualization driven by MediaPipe library.

3.2.3 Angle Calculation

We have developed an algorithm to calculate precise angles between various body landmarks for each extracted frame. For example, in Figure 4, parts (a) and (b) illustrate the angle BÂC, representing the angle between the right arm and forearm. Here, points A, B, and C correspond to landmarks 13, 11, and 15, respectively (Cf. Figure 3). Part (c) of the figure depicts the angle FÊG, which represents the angle between the right upper leg and lower leg, with points E, F, and G corresponding to landmarks 25, 23, and 27, respectively.

3.2.4 Data Storage and Augmentation

The calculated angles serve as features that describe the input variables from which our autoencoder model learns. These angles were stored in a CSV file, ensuring easy access and facilitating in-depth analysis of the athlete's kinematic information. Additionally, this dataset can be utilized to train and evaluate deep learning models for automated movement analysis and assessment.

Once the initial CSV dataset is created, it is important to enhance the stored kinematic data using data augmentation techniques. Specifically, we employ linear interpolation to generate additional data points between each pair of adjacent rows in the dataset. More precisely, we propose to add Gaussian noise to support small random variations to our numerical data. This is accomplished by adding values drawn from a normal distribution (Gaussian distribution) centred around zero to each data point.

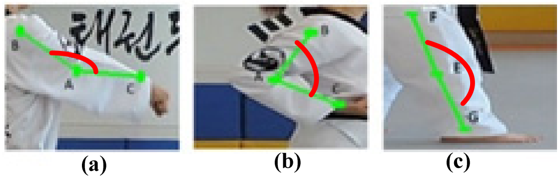


Figure 4: Example of Calculated Angles.

This augmentation process increases the overall size and diversity of the dataset, which can improve the performance of any DL models trained on this data.

3.2.5 Data Scaling

To ensure consistency across the dataset and improve the performance of any DL model trained on this data, we scaled the athlete kinematic features using the MinMaxScaler module from the scikit-learn preprocessing library. This scaling technique normalizes the data to a common numerical range, typically between 0 and 1. This data normalization process helps to achieve better convergence and stability during the training of deep learning models, as it ensures that all features are on a similar scale and have a comparable influence on the learning process.

3.3 Data Splitting

The dataset was divided into training and testing sets. We typically use 70% of the data for training and 30% for testing. This division enables the model to acquire knowledge from most of the data while being evaluated on a subset.

3.4 Model Training

3.4.1 Model Building

The deep learning approach we propose leverages autoencoders to detect incorrect movements of athletes. Specifically, we define and train an autoencoder model on data representing correct movements, enabling it to reconstruct athlete movements accurately.

The architecture of the proposed autoencoder model is given in Figure 5. The autoencoder contains three main components: the Encoder, the latent space, and the Decoder. The architecture diagram illustrates the flow from input to output through these components. More precisely, the Encoder component encompasses the following layers:

- Input Layer: The model begins with an input layer that accepts data with features corresponding to calculated angles.
- First Dense Layer (256): The input data is processed through a dense layer of 256 neurons. This layer applies L2 regularization with a coefficient of 0.01, followed by batch normalization to stabilize and accelerate the training process. LeakyReLU's activation function introduces non-linearity while allowing a small gradient when the unit is not active.
- First Dropout Layer (0.3): A dropout layer with a rate of 0.3 is applied to randomly set 30% of the input units to zero during training. This will further help to prevent overfitting.
- Second Dense Layer (128): This is followed by a dense layer with 128 neurons, which again applies L2 regularization, batch normalization, and LeakyReLU activation.
- Second Dropout Layer (0.3): A second dropout layer with a rate of 0.3 is included.
- Third Dense Layer (64): The final layer of the encoder has 64 neurons, with the same regularization and activation techniques as the previous layers. This reduces the input to a smaller dimensional space.

Latent Space (code representation): The latent space representation captures the essential features of the input data in a lower-dimensional form.

The decoder is composed of the following layers:

- First Dense Layer (128): The latent space is redirected through a dense layer comprising 128 neurons, utilizing the identical L2 regularization, batch normalization, and LeakyReLU activation.
- Second Dense Layer (64): The next layer further decreases the dimensionality with 64 neurons, continuing with regularization and activation techniques.
- Final Dense Layer (4): The final dense layer contains 4 neurons, matching the input dimension. It uses the sigmoid activation function to ensure that the output values are normalized between 0 and 1.

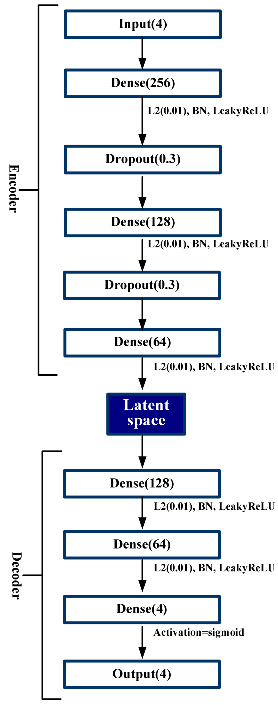


Figure 5: Architecture of the Autoencoder model.

3.4.2 Reconstruction Algorithm

The reconstruction algorithm's first step is to decompose the Poomsae video into individual movements for evaluation. The algorithm receives an initial set of 30 frames representing one movement, extracts key points for each frame, and calculates the corresponding angles. All calculated angles, along with their respective frame numbers, are then input into the autoencoder model. The model subsequently performs the reconstruction task. This process is repeated for the entire video.

3.4.3 Model Compilation

The autoencoder model is compiled using the Adam optimizer along with a Mean Squared Error (MSE) loss function. The Adam optimizer was chosen for its efficient performance with large datasets and its adaptive learning rate mechanism, which adjusts the learning rate for each parameter. The MSE loss function is particularly well-suited for regression tasks, as it minimizes the reconstruction error. This minimization is crucial for effective anomaly detection, allowing the model to accurately reconstruct normal data patterns and identify deviations representing incorrect movements.

Furthermore, to prevent overfitting and optimize training, we implemented the following callbacks:

- EarlyStopping: This callback monitors the validation loss during training and stops the training process if the validation loss does not improve for a specified number of epochs. This helps to avoid overfitting and ensures that the model maintains its best performance.
- ModelCheckpoint: The model with the lowest validation loss is saved by this callback. The saved model can then be used for further evaluation and testing.

3.4.4 Training Results

The autoencoder models were trained using the training dataset with the following parameters: 100 epochs, a batch size of 32, and a learning rate of 0.002. The model was trained with the objective of minimizing the reconstruction error. Furthermore, to ensure the effectiveness of the training process, the loss and validation loss performance measures were monitored and plotted over the epochs. This revealed how well the model was learning and whether it was over-fitting or under-fitting.

Figure 6 illustrates the training results of the first four movements of Taegeuk I Jang Poomsae. The interpretation of the results is as follows: Initially, both training and validation losses went down quickly in the beginning, which showed that the model was learning well from the input data. The loss error stabilized and reached near-zero values, showing that the models could reconstruct the input data with minimal error. The best epoch, marked by the lowest validation loss, was identified for each movement, making sure that optimal weights were saved for later evaluation.

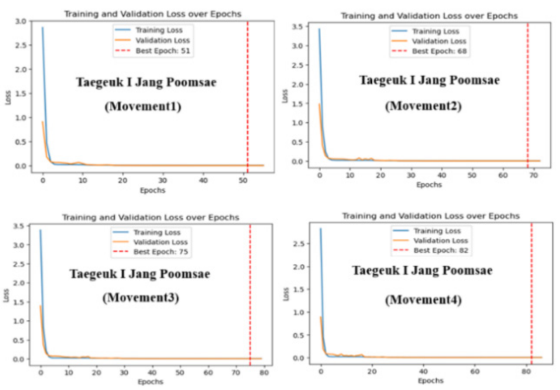


Figure 6: Training and Validation Loss Over Epochs for three different movements of Taegeuk I Jang Poomsae.

The overall training results were highly satisfactory, with each movement achieving low reconstruction errors and demonstrating robust generalization capabilities. The model's ability to accurately reconstruct the input data while maintaining low validation loss highlights its effectiveness in capturing the correct patterns for each movement in Taegeuk I Jang Poomsae.

3.5 Model Evaluation

The model's effectiveness is assessed using the Mean Squared Error (MSE) and the coefficient of determination (R2) for both the training and testing datasets. Results for the first three movements related to Taegeuk I Jang Poomsae are summarized in Table 1, presented below.

According to the presented results, the train and test MSE values are consistently low, indicating that the model performs well on both training and test datasets. The Train and Test R^2 values are close to 1, indicating that the model explains a large proportion of the variance in the data, both for training and testing. The performance metrics across different movements are consistent, indicating that the model can adapt to various types of movements.

Table 1: Results of model evaluation.

Taegeuk I Jang movements	Train MSE	Test MSE	Train R^2	Test R^2
Movement1	0.00025	0.00025	0.99451	0.99442
Movement2	0.00026	0.00026	0.99225	0.99216
Movement3	0.00022	0.00023	0.99464	0.99445

4 SPORTLAND PLATFORM

SportLand[2] is a sports tech platform that connects athletes across various disciplines, particularly in martial arts such as Taekwondo, karate, kung fu, and kickboxing. The platform features **SportLand**$_{\mathbf{AICoach}}$ module, which is an AI-powered tool designed to evaluate athlete performance. It analyses an athlete's movements during training, providing actionable insights to optimize technique and maximize performance gains. This tool serves as a self-evaluation assistant, enabling Taekwondo athletes to enhance their skills at their own pace.

[2] https://www.sportland.ai

More precisely, the **SportLand**$_{\textbf{AICoach}}$ module offers two distinct methods for evaluating athlete performance: Single-Movement Evaluation and Full-Sequence Evaluation

The next subsections provide detailed information on how these evaluations are conducted within the SportLand platform. We also provide an online demonstration video [3] illustrating these evaluations.

4.1 Single-Movement Evaluation

The **SportLand**$_{\textbf{AICoach}}$ platform provides a comprehensive interface for uploading, displaying, and analyzing videos, offering real-time feedback on movement correctness and instantaneous speed and force metrics. Athletes can upload pre-recorded videos or use their webcam to capture live footage.

Once a video is uploaded or recorded, it is decomposed into 30 frames representing the movement to be evaluated. Next, body landmarks are extracted, and angles are calculated from each frame. All calculated angles and their respective frame numbers are input into the trained model to compute the reconstruction error. If this error exceeds a predetermined threshold, the movement is flagged as incorrect. Conversely, if the error is within acceptable limits, the movement is deemed correct, and the algorithm subsequently calculates the athlete's speed and force. Figure 7 illustrates an example of a correctly evaluated movement.

4.2 Full-Sequence Evaluation

SportLand$_{\textbf{AICoach}}$ offers the ability to analyze the complete sequence of movements in a Poomsae. To achieve this, the athlete must upload a video, like the method used for evaluating individual movements. The uploaded video is then broken down into individual movements for assessment, and the previously described steps for assessing a single movement are applied to all segments.

The athlete's score increases with each correctly identified movement, while incorrectly identified movements receive 0 points. Additionally, performing the movements in the correct order is essential, as each movement must accurately follow its predecessor. A movement is marked as incorrect if it is not detected within 60 frames, ensuring timely execution and thorough evaluation. Each correct movement adds 2 points to the athlete's score, while incorrect movements or those performed out of sequence score 0 points. The

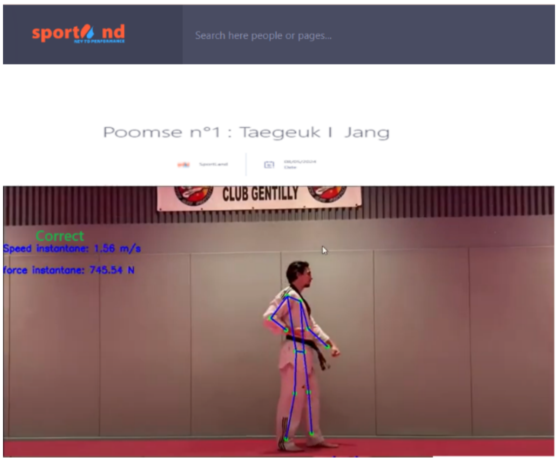


Figure 7: Example of evaluation result.

total score, which reflects the athlete's accuracy and performance, is calculated by summing the individual scores for each movement.

Figure 8 below illustrates the interface displaying a Poomsae score. The interface provides athletes with their total score and features a comprehensive table detailing each movement performed. This table displays the results for every individual movement, indicating whether each movement was assessed as correct or incorrect, along with the corresponding score for each movement.

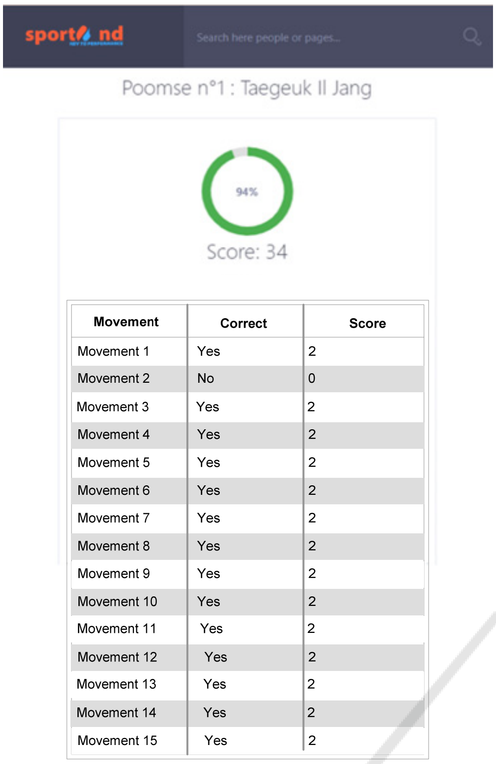
5 CONCLUSION

The primary research problem addressed in this study was the subjectivity inherent in traditional methods of evaluating Taekwondo Poomsae. This paper proposed a deep learning approach based on an autoencoder model, which effectively mitigates human intervention and demonstrates the potential for a systematic evaluation of Poomsae.

The proposed autoencoder model identified anomalies in an athlete's movements by analyzing a sequence of frames and considering movement behavior through skeleton point data, rather than focusing solely on the final pose. Additionally, the model was tested on a diverse set of video data to evaluate its generalizability to real-world scenarios. In this assessment, the autoencoder exhibited remarkable performance, achieving a coefficient of determination (R^2) close to 99% and a very low Mean Squared Error (MSE) of approximately 0.0006.

Furthermore, the absence of self-evaluation methods was identified as another challenge this

[3] https://www.youtube.com/watch?v=cOXxqcG1v8A



Movement	Correct	Score
Movement 1	Yes	2
Movement 2	No	0
Movement 3	Yes	2
Movement 4	Yes	2
Movement 5	Yes	2
Movement 6	Yes	2
Movement 7	Yes	2
Movement 8	Yes	2
Movement 9	Yes	2
Movement 10	Yes	2
Movement 11	Yes	2
Movement 12	Yes	2
Movement 13	Yes	2
Movement 14	Yes	2
Movement 15	Yes	2

Figure 8: Interface displaying the total score and detailed results for each Poomsae movement.

study aimed to address. To resolve this issue, we successfully developed the **SportLand**$_{\textbf{AICoach}}$, a reliable and objective self-paced evaluation tool. This tool leverages the autoencoder model, significantly advancing the field of Poomsae evaluation.

While this study primarily focused on the accuracy of Poomsae movements, future research could extend to evaluating player movements during Taekwondo combat. The goal of this evaluation would be to assist judges in more accurately computing athletes' scores during matches. By leveraging skeleton point data, researchers could analyze movement patterns in real-time to ensure that scoring reflects the precision and effectiveness of techniques used in combat. This approach could enhance the fairness and accuracy of scoring and contribute to a deeper understanding of performance dynamics in competitive Taekwondo.

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.

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

Fernando, M., Sandaruwan, K. D., & Athapaththu, A. M. K. B. (2024). Evaluation of Taekwondo Poomsae movements using skeleton points.

Hong, S., Park, J., Lim, J. (2021). AI-powered wearables for real-time performance feedback: A study in cycling. Sports Technology Journal, 15(1), 45-57.

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.

Michalski, R., Jones, T., Liu, C. (2022). Understanding AI model transparency in sports analytics. Journal of Applied Sports Science, 10(3), 95-108.

Pisaniello, A. (2024). The Game Changer: How Artificial Intelligence is Transforming Sports Performance and Strategy. *Geopolitical, Social Security and Freedom Journal*, *7*(1), 75-84.

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.2022.31 83112.

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