Fuzzy Logic Interacts with Attention: Trustworthy Tennis Movements Identification for Motion Capture Data

Pawel Powroznik[1,*], Maria Skublewska-Paszkowska[1], Krzysztof Dziedzic[1] and Marcin Barszcz[1]

[1]Lublin University of Technology, Nadbystrzycka 38D, 20-618 Lublin, Poland

Abstract

Innovative models of Human Action Recognition play a primary role in Computer Vision, combining image analysis, machine learning, and time-motion modelling. However, traditional graph convolution methods struggle with accurately capturing subtle spatial-temporal relationships and are sensitive to unwanted fluctuations present in motion capture datasets. Thus, this study proposes a novel Spatial-Temporal Graph Convolutional Neural Network employing a convolution layer with fuzzy logic integrated with a separate residual attention mechanism for classifying tennis movements based on 3D motion capture data, improving the model's robustness against spatial disturbances and strengthening feature extraction. Experiments on THETIS, Tennis-Mocap, and 3DTennisDS datasets confirm the effectiveness of the proposed architecture, achieving remarkable classification performance. Interpretability of the model's decision-making process highlights the key and relevant motion segments influencing the classification of tennis movements. The results are stable and reliable across experiments, which overall supports its trustworthiness.

Keywords

Fuzzy transform, Attention mechanism, Motion capture, ST-GCN, THETIS, Tennis-Mocap, 3DTennisDS

1. Introduction

Understanding sport data is a crucial aspect of athlete and team performance, decision-making processes, and health [1, 2]. Artificial intelligence-based models enable learning of complex human movement patterns leveraging features from various types of data [3]. Vision-based systems utilising sensors have become very popular because they provide accurate three-dimensional data allowing for precise athlete performance, including activity recognition [4, 5, 6, 7]. The fuzzy logic concept has been widely applied to vague reasoning with various degrees of truth which is useful in dynamic and realistic conditions [8]. Fuzzy logic-based classification involves the linguistic rules that are more understandable by humans [9]. It is applied in a great variety of fields, such as modelling, expert systems, forecasting, computer vision, pattern recognition, image processing, and robotics [10, 11]. Denoising videos presenting various sports disciplines, including tennis, utilising fuzzy Support Vector Machine (F-SVM), is stated to be an effective part of a recognition system [12]. Video classification incorporating fuzzy logic with the Big Bang-Big Crunch technique proved to be a sufficient solution for football scene identification, utilising the fuzzy C-mean (FCM) clustering algorithm. Several fuzzy rules were incorporated to predict the result of the basketball game [13]. In [14] a new method involving optical flow and fuzzy rules was proposed for table tennis recognition. Optical flow calculates the serve and receive movements, with the size distribution built using fuzzy logic. The Inflated 3D ConvNet, pre-trained on the ImageNet, Kinetics, and 3D action video datasets, was applied for classification [15]. Fuzzy C-Means utilising the dynamic time warping-based distance has been shown to be an effective tool for clustering basic tennis movements, including forehand, backhand, and volley [16, 17]. These studies involve motion

TRUST-AI: The Second European Workshop on Trustworthy AI. Organized as part of the International Joint Conference on Artificial Intelligence - IJCAI/ECAI 2026. August 2026, Bremen, Germany.

*Corresponding author.
p.powroznik@pollub.pl (P. Powroznik); maria.paszkowska@pollub.pl (M. Skublewska-Paszkowska); k.dziedzic@pollub.pl (K. Dziedzic); m.barszcz@pollub.pl (M. Barszcz)
0000-0002-5705-4785 (P. Powroznik); 0000-0002-0760-7126 (M. Skublewska-Paszkowska); 0000-0003-4909-9980 (K. Dziedzic); 0000-0002-3428-9805 (M. Barszcz)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-09-02

capture (MoCap) data of a tennis racket position. Trapezoidal membership function has been utilised to fuzzify the input data in tennis strokes classification based on motion capture data utilising the Spatial-Temporal Graph Convolution Neural Network (ST-GCN)-based architectures [18, 19, 20, 21]. Their application allowed to overcome the very thin boundary of recognising similar movements, such as forehand and volley forehand, as well as to identify individual phases of these strokes, the boundaries of which are difficult to determine.

Despite extensive use of fuzzy logic in many areas, this is a newly emerging issue in sport, especially in the case of recognition tasks. Various transforms, like Fourier, Laplace, integral or wavelet, possess powerful potentials in classification purposes [22]. In tennis motion analysis, trustworthiness means that a model should support, not replace, the coach or athlete. An unreliable model may provide misleading training feedback, overemphasise incorrect stroke patterns, or fail when marker registration changes. In this work, trustworthiness is therefore considered through predictive reliability, cross-dataset generalisation, calibration, robustness to spatial fluctuations, and marker-level interpretability.

The main aim of this study is to propose a novel, reliable neural network architecture employing a Gaussian membership function integrated with a custom residual attention mechanism for classifying tennis movements based on 3D motion capture data. Thus, the main contributions are as follows:

- We propose the Spatial-Temporal Graph Fuzzy Convolutional Network with Residual Attention Mechanism (ST-GFCN RAM), which utilises MoCap data for forehand, backhand, and volley recognition. It incorporates fuzzy convolution to enhance indicating spatial-temporal relationships and the residual attention mechanism to extract the most relevant features.
- We assess the developed model utilizing trustworthy evaluation protocol that combines 1) accuracy, precision, recall, and F1-score metrics, 2) marker-level explainability of tennis movement recognition using SHAP, 3) model generalization with cross-dataset testing on THETIS, Tennis-Mocap, and 3DTennisDS, 4) model calibration employing Expected Calibration Error (ECE), and 5) evaluation of the model robustness.
- We perform the state-of-the-art studies to compare our approach with ST-GCN-based architectures for tennis movement identification. This evaluation revealed our model to be highly effective.
- Ablation study reveals that the residual attention module and fuzzy membership are relevant to the model's performance.

2. Materials and methods

2.1. Datasets

Three datasets containing tennis movements, recorded using different markerless and marker-based motion capture systems, were incorporated in this study. Due to various number of points, defining key body positions, all data were simplified to the THETIS model. The **THETIS** [24] consists of 12 tennis strokes (backhand, forehand, service, and smash with their various variants), utilised by 31 amateurs and 24 professional tennis players, registered by Microsoft's Kinect motion capture system with the frequency set to 60Hz. In total, 304 backhand, 310 forehand, 106 volley forehand, and 99 volley backhand moves are gathered in ONI format, defining 15 points of the tennis player model (Fig. 2). The **Tennis-Mocap** dataset [25] consists of tennis MoCap data captured from 17 athletes (5 high performers and 12 regular ones) from the Caldas-Colombia tennis league. Their movements were recorded using the 6-camera optical passive Optitrack Flex V100 with 100 Hz frequency. Each participant had 34 markers attached to the body, defined by Arena software according to the Biovision Hierarchy protocol. The athletes tried to hit the ball as they participated in a real tournament for 30 seconds. In total, 161 backhand, 161 forehand, 144 volley forehand, and 144 volley backhand moves are analysed. The **3DTennisDS** dataset [26] contains forehand, backhand, volley forehand, and volley backhand, from 10 tennis players, recorded using the passive optical 8-camera Vicon motion capture system with the frequency set to 100Hz. The capture was performed using the Plug-in Gait biomechanical model, consisting of 39 retro-reflective markers attached to the defined parts of the body. For this study, 208 backhands, 170 forehands, 121 volley forehands, and 121 volley backhands were gathered.

2.2. Fuzzy convolutional layer

The graph fuzzy convolution layer (GFC Layer) is created to better extract information from MoCap data. Spatial points from 3D data are represented as graph nodes, and their spatial relationships are described using adjacency matrices.

Let $G = (V, E)$ be an undirected spatial graph, where $V = \{v_i\}_{i=1}^{N}$ denotes a set of nodes, and $E \subseteq V \times V$ a set of edges describing the spatial neighbourhood of nodes. Each node v_i is assigned a feature $x_i \in \mathbb{R}^3$ corresponding to the spatial coordinates of 3D points.

Let $\mu_{FT} : \mathbb{R}^3 \times \mathbb{R}^3 \to [0, 1]$ be a fuzzy membership function satisfying symmetry and spatial constraints. In this study, the function is the Gaussian membership function:

$$\mu_{FT}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_{\text{fuzz}}^2} \right), \tag{1}$$

where σ_{fuzz} is a parameter defining the blur range.

The use of a Gaussian membership function in the GFC layer causes the influence of points that are spatially distant from a given point to be significantly reduced, resulting in improved robustness to input data unwanted fluctuations. The properties of the Gaussian function indicate that the larger the distance between points, the smaller the value of $\mu_{FT}(x_i, x_j)$. This significantly limits the impact of unwanted fluctuations, spatially distant points on the convolution result.

The standard graph convolution operation is defined as:

$$h_i^{(l+1)} = \phi\left(\sum_{j \in \mathcal{N}(i)} A_{ij} W^{(l)} h_j^{(l)} \right), \tag{2}$$

where A denotes a symmetric adjacency matrix, while graph fuzzy convolution is given by:

$$h_i^{(l+1)} = \phi\left(\sum_{j \in \mathcal{N}(i)} A_{ij} \mu_{FT}(x_i, x_j) W^{(l)} h_j^{(l)} \right), \tag{3}$$

where $h_i^{(l)}$ denotes the representation of node v_i at level l and $W^{(l)}$ the weight matrix for training. Here, $\mathcal{N}(i)$ is defined by the anatomical adjacency matrix A, whereas $\mu_{FT}(x_i, x_j)$ softly reweights only these anatomically admissible connections according to the normalised spatial distance between markers. Thus, the fuzzy term does not create arbitrary edges. It modulates the strength of existing graph connections. In Eqs. (2)-(3), $\phi(\cdot)$ denotes a generic nonlinear activation function, whereas σ_{fuzz} is reserved for the Gaussian membership width.

Assuming that the membership function μ_{FT} is symmetric, the GFC layer preserves the spatial symmetry of the input data. It follows directly from the symmetry of the fuzzy membership function and the symmetry of the adjacency matrix.

The GFC layer can be applied with any fuzzy membership function μ_{FT} that satisfies basic spatial constraint and symmetry conditions. The quality of data classification depends on the proper selection of the μ_{FT} function and the fuzzy parameters. The generalisation follows directly from the definition of the GFC layer and the arbitrary choice of the fuzzy membership function, which provides the appropriate mathematical and spatial properties. The GFC layer logically results from the need to apply a fuzzy mechanism to improve the classification of spatial motion data. By selecting a fuzzy membership function, this layer can be flexibly adapted to the specifics of the analysed data, significantly improving its resistance to unwanted fluctuations and classification quality.

2.3. Residual attention layer

Residual Attention Layer (RA Layer) is introduced to effectively enhance important features and mitigate the influence of less important features through a selective attention mechanism. This layer is especially useful in contexts where data is noisy or complex, such as spatio-temporal data.

For a feature representation in layer l, denoted as $H^{(l)} \in \mathbb{R}^{N \times F}$, where N denotes the number of nodes and F denotes the number of features, the RA operation is defined as:

$$H^{(l+1)} = H^{(l)} + \rho(Conv(H^{(l)})) \odot H^{(l)} \tag{4}$$

where $\rho(\cdot)$ is the sigmoid activation function, $Conv(\cdot)$ is a 1D convolution operation, and $\odot$ denotes element-wise multiplication.

The residual attention mechanism improves the stability of network training by providing a direct gradient flow, which mitigates the problem of vanishing gradients. The residual attention layer contains a residual structure (skip connection) that allows for a direct gradient flow. In practice, due to the direct path $H^{(l)}$, the gradient of the loss function with respect to $H^{(l)}$ is the sum of the gradient passed directly and the gradient passed through the attention layer, which prevents its value from decreasing rapidly. Formally, if we denote the derivative of the loss with respect to the output of the RA layer as $\partial L/\partial H^{(l+1)}$, then the residual path implies

$$\frac{\partial L}{\partial H^{(l)}} = \frac{\partial L}{\partial H^{(l+1)}}\big(I + M\big), \tag{5}$$

where M is a diagonal matrix whose entries are the attention values $\rho(Conv(H^{(l)}))$. All diagonal entries lie in $(0, 1)$, the smallest eigenvalue of $I + M$ is at least 1, it prevents vanishing gradients [27].

The residual attention layer can be interpreted as an adaptive feature scaling, where the values of the attention function $\sigma(Conv(H^{(l)}))$ decide which features will be enhanced or weakened. The sigmoid function in the attention layer generates values in the range $(0, 1)$, which directly affect the importance of individual features. The application of the RA layer leads to improved classification results by selectively enhancing important features and suppressing irrelevant ones, especially in the case of spatio-temporal data. It has been experimentally shown that the residual attention mechanism effectively increases the importance of key features of input data, which is translated into higher classification accuracy. The empirical evidence is presented in a number of works related to attention mechanisms [28, 29].

2.4. Model architecture

Let $\mathcal{X} = \{x_t^{(i)}\}_{t=1,\ldots,T}^{i=1,\ldots,N} \subset \mathbb{R}^3$ denote a motion capture sequence of T frames and N markers. Each frame is normalised by its root–joint position and scale, then mapped to an undirected spatial graph $G = (V, E)$ whose vertices correspond to markers and whose edges follow an anatomically plausible connection scheme extracted from the metadata of the recording system. The resulting adjacency matrix $A \in \{0, 1\}^{N \times N}$ is kept fixed during training. Temporal ordering is preserved, so that the pre-processed input is a tensor $X \in \mathbb{R}^{T \times N \times 3}$ accompanied by A.

The network is a composition of L identical *ST-GFC-RA* blocks followed by a spatio-temporal pooling and a classification head. Write $H^{(0)} = XW_{\text{in}}$, where $W_{\text{in}} \in \mathbb{R}^{3 \times F}$ is a learned linear projection to the initial feature dimension F. For $l = 1, \ldots, L$ the l-th block performs three successive transformations defined by Eq. (6):

$$H^{(l)} = \mathrm{RA}\Big(\mathrm{TC}\big(\mathrm{GFC}(H^{(l-1)}, A, \Theta_l)\big)\Big), \tag{6}$$

where Θ_l collects the trainable parameters of the block. The first operator GFC implements the graph fuzzy convolution (Section 2.2). It aggregates spatial information with adaptive fuzzy weights $\mu_{FT}(x_i, x_j)$ instead of the binary entries of A, producing features $\tilde{H} \in \mathbb{R}^{T \times N \times F}$. The second operator TC is a one-dimensional convolution of kernel size 3 applied along the temporal axis of $\tilde{H}$ with shared weights across nodes, thereby capturing short-range dynamics while keeping the spatial graph fixed. The RA is described in Section 2.3. By the identity skip path, it redistributes importance among channels without impeding the gradient flow, as guaranteed by the relation $\partial L/\partial H^{(l-1)} = \partial L/\partial H^{(l)}(I + M)$ where M is the diagonal attention mask.

After the stack of blocks, the tensor $H^{(L)}$ is pooled in two steps. A temporal average collapses the T dimension, followed by a graphwise average that collapses the N vertices, yielding a vector

$h_{\text{pool}} \in \mathbb{R}^F$. This representation is passed through a fully connected layer $W_{\text{cls}} \in \mathbb{R}^{F\times K}$, producing logits $z = W_{\text{cls}} h_{\text{pool}}$ that are normalised by a Softmax, where K is the number of tennis-stroke classes. Formally, the architecture realises a function $f_\theta\colon \mathbb{R}^{T\times N\times 3} \to \Delta^{K-1}$, $\theta = \{W_{\text{in}}, \Theta_1, \ldots, \Theta_L, W_{\text{cls}}\}$, that is trained end-to-end with a cross-entropy objective. The presence of fuzzy–weighted graph convolutions equips f_θ with robustness to spatial jitter inherent in marker trajectories, while residual attention compensates for temporal misalignments and erratic marker drop-outs, yielding a model that outperforms the vanilla ST-GCN baseline of [30] in our ablation studies. The design remains compatible with alternative membership functions and deeper attention schemes, making it a flexible backbone for future research on motion capture understanding. The entire model architecture is depicted in Fig. 1.

3. Experiments and results

All experiments are executed on a workstation equipped with an AMD Ryzen Threadripper PRO 7965WX CPU and a single NVIDIA RTX 4090 GPU. The implementation relies on PyTorch, and it is available on GitHub (https://github.com/AnonymousCodesForPapers/ST-GFCN). We assess the developed model utilising a trustworthy evaluation protocol that combines: 1) standard metrics, 2) Shap values visualisation, 3) model generalisation, 4) model calibration employing ECE, and 5) evaluation of the model robustness.

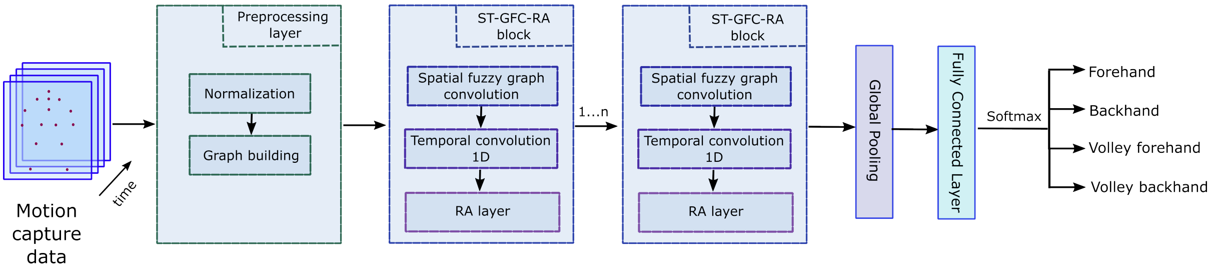

Figure 1: Architecture diagram of the ST-GFCN RAM model applied in the presented studies.

To ensure full replicability, all studies are run with a fixed random seed and exactly the same configuration file. For every dataset, the player-wise 80% and 20% split recommended by the original authors is respected, ensuring that athletes in the training set never occur in test sets. Five random splits are repeated and averaged through stratified Monte-Carlo cross-validation. Sequences are clipped to two seconds, zero-padded where necessary, and temporally down-sampled to 120 frames. The blur parameter σ_{fuzz} is initialised at 50 and expressed during training as $\sigma_{\text{fuzz}}(e) = \sigma_{\text{fuzz}}(0)\exp(-0.05\,e)$, where e denotes the epoch index.

Models are trained with AdamW, an initial learning rate of 5×10^{-4}, cosine decay, a batch size of 32, weight decay of 10^{-2}, and a DropPath rate of 0.1. Training ends in 120 epochs. A two-stage curriculum stabilises fuzzy weights: during the first 20 epochs, temporal convolutions and attention are disabled, allowing fuzzy masks to settle on plausible anatomical links. All layers are then unfrozen for the remaining 100 epochs. A cyclical learning-rate schedule with warm restarts shortens convergence by about 15% compared with monotone cosine decay while preserving peak accuracy. The weights achieving the best validation accuracy are finally evaluated on the held-out test fold. The ST-GFCN RAM performance is compared against four strong baselines trained under identical preprocessing: the original ST-GCN, ST-GCN augmented only with residual attention (A3T-GCN) [31], ST-GCN containing only the fuzzy convolution (4s-ShiftGCN) [32], and Feature Fusion Graph Consecutive-Attention Network (FFGCAN) that incorporates Adaptive Consecutive Attention Module and Graph Self-Attention module [33]. All baseline hyperparameters are tuned on the validation set.

The proposed ST-GFCN RAM achieves its highest performance on the 3DTennisDS dataset, illustrating the benefit of fuzzy weights (Table 1). Paired permutation tests with $p < 0.01$ confirm that the improvements over vanilla ST-GCN are statistically significant on every benchmark.

The ablation protocol used the same preprocessing, player-wise splits, random seeds, and hyperparameters for all variants. Starting from the full ST-GFCN RAM, we removed one component at a time: fuzzy weighting, residual attention, or adaptive σ_{fuzz} annealing. Each variant was retrained from scratch and evaluated on the same five Monte Carlo splits (Table 1). While the fuzzy transform is replaced by a standard graph convolution that uses binary entries of the adjacency matrix A, the macro-averaged scores drop approximately $6\% - 8\%$. A paired permutation test on five Monte Carlo partitions gives $p = 3.2 \times 10^{-3}$, indicating that the improvement provided by fuzzy weighting is statistically significant. Removing the RA layer while keeping fuzzy convolutions decreased macro-averaged scores by about $3\% - 4\%$. The variance across splits also increased from 0.6 to 1.1 percentage points, suggesting reduced robustness. The experiment with Gaussian noise further illustrates the robustness of the model to small coordinate perturbations (Fig. 4). Keeping the fuzzy transform and RA layer intact, but fixing $\sigma_{\text{fuzz}} = 50$ throughout training, produces macro-averaged scores over 89%. Although the accuracy gap to the full system is modest ($\approx 0.9\%$), early-epoch loss stagnates for roughly ten epochs longer, confirming that annealing expedites convergence. A two-way ANOVA on the four-model pool reveals that all layers interact positively ($F(1, 16) = 9.3$, $p = 7.8 \times 10^{-3}$). In other words, the presence of attention amplifies the benefit of fuzzy weighting and vice versa. Stacking both mechanisms is therefore essential to achieve performance. Although the increase in accuracy over ST-GCN is statistically significant, the leave-one-dataset-out test revealed a 2.2% decrease, indicating that further study is needed to improve generalisation across registration protocols.

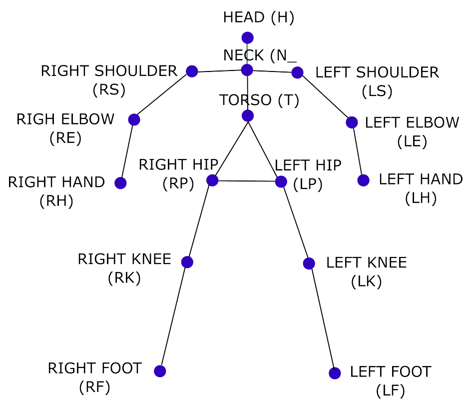

Figure 2: THETIS tennis player model with body key positions.

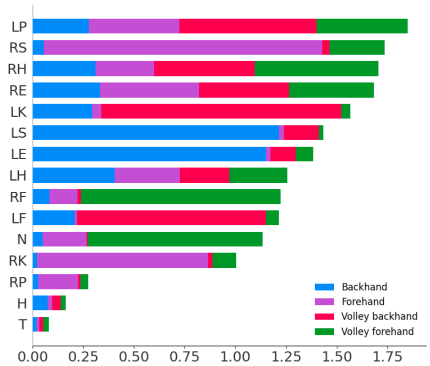

Figure 3: Markers average impact on tennis movements identification utilising SHAP.

The probabilistic reliability of the four-class classifier was assessed using the top-label ECE. For each sample, the predicted label was obtained as the class with the maximum softmax probability, and the corresponding confidence was defined as this maximum probability. The confidence interval $[0, 1]$ was divided into ten equal-width bins. For each bin, the empirical accuracy and average confidence were computed, and ECE was calculated as: $\text{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{n} \left|\text{acc}(B_m) - \text{conf}(B_m)\right|$, where B_m denotes the set of samples assigned to the m-th confidence bin, n is the total number of samples, $\text{acc}(B_m)$ is the empirical accuracy in the bin, and $\text{conf}(B_m)$ is the mean confidence.

Temperature scaling was used as a post-hoc calibration method. This parameter was estimated during cross-validation by minimising the negative log-likelihood and was subsequently applied to the independent test set. Since the available outputs were stored as softmax probabilities, temperature scaling was implemented in the equivalent probability-space form: $\hat{p}_i^{(T)} = \frac{\hat{p}_i^{1/T}}{\sum_{j=1}^{K} \hat{p}_j^{1/T}}$, where $K = 4$ denotes the number of classes. Values of $T < 1$ sharpen the predictive distribution, while values of $T > 1$ soften it.

The results in Table 1 show that the classifier achieved high test accuracy on all datasets while fitted temperatures were consistently below one, indicating that the original model tended to underestimate its confidence relative to its empirical correctness. SHAP values were computed for the input marker trajectories and aggregated over time, coordinate channels, and correctly classified test samples to

Table 1
Macro-averaged results (%) over five splits. Acc. denotes accuracy, Prec. - precision, F1 - F1 score, Full denotes ST-GFCN RAM model, Fuzzy means fuzzy transform layer, Att - residual attention layer, ECE

Dataset	Model	Acc.	Prec.	Recall	F1
THETIS	**Full**	**93.7±0.5**	**93.2±0.5**	**93.0±0.5**	**90.9±0.5**
	- Fuzzy	89.6±0.6	89.0±0.7	88.0±0.7	88.5±0.7
	- Att	90.7±0.7	90.1±0.7	89.3±0.8	89.7±0.8
	Full	91.1±0.6	90.5±0.6	89.3±0.6	89.9±0.7
	Calibration.: ECE before = 0.171, ECE after = 0.109, T = 0.798				
Tennis -Mocap	**Full**	**91.6±0.6**	**91.0±0.6**	**90.0±0.8**	**86.5±0.7**
	- Fuzzy	85.8±0.8	85.1±0.9	84.0±1.1	84.6±1.0
	- Att	86.9±0.7	86.2±0.8	85.1±1.0	85.8±0.9
	Full	89.5±0.7	89.0±0.7	88.4±0.8	88.7±0.8
	Calibration.: ECE before = 0.171, ECE after = 0.107, T = 0.835				
3DTDS	**Full**	**96.4±0.5**	**95.8±0.5**	**95.0±0.7**	**95.3±0.6**
	- Fuzzy	85.9±0.9	85.1±1.0	84.1±1.2	84.7±1.0
	- Att	88.2±0.8	87.5±0.9	86.0±1.0	87.0±0.9
	Full	92.8±0.7	92.3±0.7	91.7±0.7	92.0±0.8
	Calibration: ECE before = 0.171, ECE after = 0.109, T = 0.798				

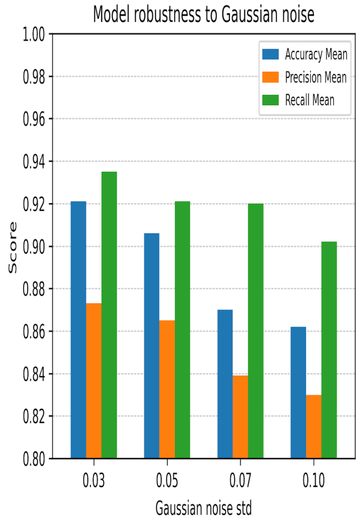

Figure 4
The impact of Gaussian noise on classification results.

Table 2
ST-GFCN RAM cross-dataset testing

Training dataset	Testing dataset	Acc.	Prec.	Recall	F1
THETIS	Tennis -Mocap	88.5 ±0.5	88.9 ±0.6	87.8 ±0.6	88.3 ±0.7
	3DTDS	88.1 ±0.7	87.4 ±0.8	88.1 ±0.8	87.7 ±0.8
Tennis -Mocap	THETIS	86.6 ±0.7	87.3 ±0.9	88.8 ±0.8	88.0 ±0.9
	3DTDS	86.9 ±0.8	87.6 ±1.1	88.4 ±0.9	87.0 ±1.0
3DTDS	Tennis -Mocap	88.6 ±0.6	87.8 ±0.8	87.3 ±0.6	87.6 ±0.7
	THETIS	89.3 ±0.9	88.7 ±0.8	87.4 ±1.1	88.0 ±0.9

Table 3
Comparison with the state-of-the-art

Dataset	Model	Acc.
THETIS	ST-GCN	88.9±0.7
Tennis-Mocap		84.9±0.9
3DTDS		84.5±1.0
THETIS	4s-ShiftGCN	87.0±0.6
Tennis-Mocap		85.2±0.7
3DTDS		88.6±0.7
THETIS	A3T-GCN	87.3±0.6
Tennis-Mocap		89.0±0.7
3DTDS		91.8±0.7
THETIS	FFGCAN	92.3±3.1
Tennis-Mocap		92.3±0.8
3DTDS		95.1±0.7

obtain marker-level importance (Fig. 3). The highest contributions are observed for pelvis-, shoulder-, and upper-limb-related markers. However, SHAP is used here as evidence of plausibility rather than as causal biomechanical proof. Explanations may still be affected by dataset geometry, handedness distribution, and marker normalisation. The cross-dataset results indicate promising but incomplete generalisation across MoCap protocols (Table 2, Table 3). The observed performance drop confirms that registration-specific marker layouts remain an important deployment limitation.

4. Conclusions and future works

This study addresses the innovative ST-GFCN RAM with a residual attention mechanism for tennis movements recognition. This new model incorporates fuzzy convolution to enhance extracting spatial-temporal features as well as a residual attention mechanism σ_{fuzz} for indicating the most relevant features for classification. The ST-GFCN RAM was evaluated utilising three well-known motion capture datasets: THETIS, Tennis-Mocap, and 3DTennisDS.

From a trustworthiness point of view, the model exhibits predictive reliability, calibration, cross-dataset evaluation, and SHAP-based marker-level plausibility. However, it should be considered a coach-in-the-loop decision-support component rather than an autonomous assessment system. Four-class stroke taxonomy, partial performance loss across registration protocols, empirical tuning required for the fuzzy schedule, and computational requirements for deployment are the main limitations. Future research will focus on extending the label space to include stroke phases, testing markerless pose estimation settings, and investigating adaptive membership functions.

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] P. D'Urso, L. De Giovanni, L. Federico, V. Vitale, Network and attribute-based clustering of tennis players and tournaments, Computational Statistics (2024) 1–24.

[2] B. Razavian, S. M. H. Seyedbeiglou, E. B. Tirkolaee, F. Anka, A novel two-stage fuzzy classification method with different weight permutations for optimal gis-based placement of wellness and sports centers, Expert Systems with Applications 278 (2025) 127282.

[3] Z. Zhao, Y. Chen, P. Goh, N. S. Ahmad, Parameterized synthetic wifi csi data generation for trustworthy human activity recognition: A sim2real approach with edge deployment analysis, Internet of Things (2026) 101910.

[4] J. Patalas-Maliszewska, I. Pajak, P. Krutz, G. Pajak, M. Rehm, H. Schlegel, M. Dix, Inertial sensor-based sport activity advisory system using machine learning algorithms, Sensors 23 (2023) 1137.

[5] B. Bačić, Extracting player's stance information from 3d motion data: A case study in tennis groundstrokes, in: Image and Video Technology, Springer, 2015, pp. 307–318.

[6] B. Bačić, Towards the next generation of exergames: Flexible and personalised assessment-based identification of tennis swings, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–8.

[7] S. Vinyes Mora, W. J. Knottenbelt, Deep learning for domain-specific action recognition in tennis, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 114–122.

[8] H. Novatchkov, A. Baca, Fuzzy logic in sports: a review and an illustrative case study in the field of strength training, International Journal of Computer Applications 71 (2013) 8–14.

[9] W. Song, H. Hagras, A big-bang big-crunch fuzzy logic based system for sports video scene classification, in: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2016, pp. 642–649.

[10] F. Faizah, A. Triwiyatno, R. R. Isnanto, Fuzzy logic implementation on motion of tennis ball picker robot, in: 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), IEEE, 2021, pp. 57–63.

[11] S. Jia, J. Yan, J. Fan, X. Li, L. Gao, Multimodal intelligent wheelchair control based on fuzzy algorithm, in: 2012 IEEE International Conference on Information and Automation, IEEE, 2012, pp. 582–587.

[12] L. Gao, Y. Zhao, Application of sports video image analysis based on fuzzy support vector machine, Scalable Computing: Practice and Experience 25 (2024) 1790–1798.

[13] K. Trawiński, A fuzzy classification system for prediction of the results of the basketball games, in: International conference on fuzzy systems, IEEE, 2010, pp. 1–7.

[14] C.-J. Wang, J.-R. Chang, H. Y. Lin, C. J. Lu, Frames extraction from table tennis competition videos for action classification using optical flow and fuzzy rules, in: 2023 9th International Conference on Applied System Innovation (ICASI), IEEE, 2023, pp. 223–225.
[15] I. Hayashi, M. Fujii, T. Maeda, J. Leveille, T. Tasaka, Extraction of knowledge from the topographic attentive mapping network and its application in skill analysis of table tennis, Journal of human kinetics 55 (2017) 39.
[16] M. Skublewska-Paszkowska, P. Karczmarek, E. Lukasik, Tennis multivariate time series clustering, in: 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2021, pp. 1–6.
[17] M. Skublewska-Paszkowska, P. Powroznik, P. Karczmarek, E. Lukasik, J. Smolka, Fuzzy c-means clustering for motion capture tennis time-series data, IEEE Access (2024).
[18] M. Skublewska-Paszkowska, P. Powroznik, E. Lukasik, Learning three dimensional tennis shots using graph convolutional networks, Sensors 20 (2020) 6094.
[19] M. Skublewska-Paszkowska, P. Powroznik, P. Karczmarek, E. Lukasik, Aggregation of tennis groundstrokes on the basis of the choquet integral and its generalizations, in: 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2022, pp. 1–8.
[20] M. Skublewska-Paszkowska, P. Powroznik, E. Lukasik, Attention temporal graph convolutional network for tennis groundstrokes phases classification, in: 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2022, pp. 1–8.
[21] M. Skublewska-Paszkowska, P. Karczmarek, P. Powroznik, E. Lukasik, J. Smolka, M. Dolecki, Aggregation of tennis multivariate time-series using the choquet integral and its generalizations, in: 2023 IEEE International Conference on Fuzzy Systems (FUZZ), IEEE, 2023, pp. 1–6.
[22] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy sets and systems 157 (2006) 993–1023.
[23] T. T. Nguyen, T. Le Nguyen, G. Ifrim, Robust explainer recommendation for time series classification: Tt nguyen et al., Data Mining and Knowledge Discovery 38 (2024) 3372–3413.
[24] S. Gourgari, G. Goudelis, K. Karpouzis, S. Kollias, Thetis: Three dimensional tennis shots a human action dataset, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2013, pp. 676–681.
[25] C. K. Valencia-Marin, J. D. Pulgarin-Giraldo, L. F. Velasquez-Martinez, A. M. Alvarez-Meza, G. Castellanos-Dominguez, An enhanced joint hilbert embedding-based metric to support mocap data classification with preserved interpretability, Sensors 21 (2021) 4443.
[26] M. Skublewska-Paszkowska, P. Powroznik, E. Lukasik, J. Smolka, Tennis patterns recognition based on a novel tennis dataset–3dtennisds, Advances in Science and Technology. Research Journal 18 (2024).
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[29] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3156–3164.
[30] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.
[31] M. Skublewska-Paszkowska, P. Powroznik, Temporal pattern attention for multivariate time series of tennis strokes classification, Sensors 23 (2023) 2422.
[32] X. Li, J. Kang, Y. Yang, F. Zhao, A lightweight attentional shift graph convolutional network for skeleton-based action recognition, INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL 18 (2023).
[33] P. Powroznik, M. Skublewska-Paszkowska, K. Dziedzic, M. Barszcz, Feature fusion graph consecutive-attention network for skeleton-based tennis action recognition, Applied Sciences 15 (2025) 5320.