BMC Sports Science, Medicine and Rehabilitation

RESEARCH

Open Access

Deep learning-based tennis match type clustering

Hyo-Jun Yun[1], Nara Jang[2*] and Minsoo Jeon[3*]

Abstract

Background This study aims to define and cluster tennis match types based on how they are played.

Methods The research data selected for this study were from the 100th round of 32 matches of the five finals of the 2023 International Tennis Open Tournament. Based on expert knowledge and sports expertise, 27 variables were included across seven areas. Three models were applied and the silhouette coefficient was calculated to identify the optimal number of clusters. A difference test was conducted on the game record variables based on the cluster results.

Results Calculation of the silhouette coefficients for the three models showed that Model 3 (silhouette coefficient: 0.406) had the highest performance. The clustering results for the tennis match types are as follows. First, the NEt Rusher Defensive type, which is defensive and induces net play. Second, the ALl Courter Defensive type, which is either defensive or all-round. Third, the STroke Placement Offensive type, which is aggressive and has strengths in stroke. Fourth, the SErve Placement Offensive type, which is aggressive and has strengths in sub courses.

Conclusion This study's findings are not only provide basic data to cluster game types in tennis matches but also to contribute to establishing game strategies for each game type, thereby further improving performance.

Keywords Deep learning, Transformer, Tennis, Match type

Introduction

Researchers interested in sports analytics have argued that clustering and players' game types are important factors that must be considered when analyzing players [1–4]. This is because when clustering a player's game type, a more accurate analysis can be performed if the game characteristics are reflected in advance. For example, when Type A has an aggressive tendency and Type B has a defensive tendency, the technique of scoring Type A against Type B can be analyzed, which can be used as meaningful information for leaders and players in planning tactics and strategies. Therefore, game-type clustering analysis is essential for sports.

In sports games, the need for game-type clustering analysis varies depending on the rules of the game. For example, in sports such as swimming, weightlifting, shooting, and archery, it is more important to analyze one's own competition tent than to cluster and analyze competition types based on the characteristics of the recorded event. On the other hand, sports such as judo, taekwondo, badminton, and tennis have the characteristic of not only playing against an opponent simultaneously

*Correspondence:
Nara Jang
nara7888@naver.com
Minsoo Jeon
minsu1144@nate.com
[1]Center for Sports and Performance Analysis, Korea National Sport University, Seoul, Republic of Korea
[2]Department of Physical Education, Korea National Sport University, Seoul, Republic of Korea
[3]Department of International Sport, Dankook University, Chungcheongnam-do, Republic of Korea

but also changing the content of one's game depending on the opponent's game; thus, it is important to check the players' game type. Moreover, in sports where there is a high need for game type clustering, it is common to establish tactics and strategies by analyzing the characteristics of opposing players before competition [5–9]. In particular, ac-cording to a report by Choi [5], technical and physical aspects are emphasized to establish tactics, suggesting that these are important factors in distinguishing game types. Therefore, the need for clustering game types is emphasized to plan efficient tactics.

Tennis, a sport that involves simultaneously playing against an opponent. It is a racket sport where the content of one's game changes depending the opponent's game type [10]. Due to the nature of tennis, one's game content and tactics change depending on the opponent's serve direction, technique, and court location, among others. The game strategy is determined by what type of game the player is playing against [11]. In sports similar to tennis, such as badminton and table tennis, there have been analyses of match types and gameplay content considering the opponent [12–15], and this has been reported to be effective in separating and analyzing game types. Considering this, the common view of field leaders and researchers is that in tennis, analyzing game types separately can provide more meaningful information.

Based on previous studies on tennis game analysis, serve analysis [13, 16–18], score analysis [19, 20], and speed analysis [21]. In tennis matches, research is being conducted on match style, strategy analysis, and match outcome prediction using machine learning [22]. In tennis matches, research is being conducted on match style strategy analysis, and match outcome prediction using machine learning [22]. Despite the various academic research aimed at improving player performance in tennis, there is a scarcity of studies that analyze games through the clustering of game types, which are considered important in the field of tennis, most of which are based on simple technical analyses or studies that analyze games as a single variable. Although these studies convey simple information to athletes and coaches, they are insufficient for providing specific information. In other words, in tennis, scoring is achieved through connections, rather than a single variable. There are significant limitations in explaining all movements with a single variable. Therefore, several game content variables in tennis must be analyzed and research on the clustering of game types emphasized in tennis is necessary.

A variety of analysis methods have been introduced to cluster tennis game types; a deep-learning-based transformer can be applied as a recent methodology with high clustering accuracy. Transformers were first introduced by Vaswani et al. in 2017 and have been reported to perform well in natural language processing [23].

Table 1 Characteristics of international open competitions selected as research subjects

Competition name	Number of set	Number of players
2023 BNP Paribas Masters	44	37
2023 Miami Masters	30	24
2023 Rolex Shanghai Masters	70	52
2023 Cincinnati Masters	20	18
2023 Canada Masters	36	27
Total	200	99

Table 2 Final selected variables

Category	Variables	Category	Variables
Serve	First serve	Technique	Forehand
	Second serve		Backhand
Serve Location	1		Foreslice
	2		Backslice
	3		Forevolley
Point Type	Winning shot		Backvolley
	Unforced error		Smashing
	Error		Dropshot
Rally Location	1	Error Type	Net
	2		Left and Right of Court
	3		Back of Court
	4		Double Fault
	5	Rally	Number of Rally

The Transformer model is capable of parallel processing, significantly improving the computation speed, and unlike RNNs or LSTMs, it efficiently handles long-term dependencies regardless of the sequence length [23]. The advantage of transformers is that they are known to have a higher clustering accuracy than previously introduced clustering methods; therefore, they are one of the methods commonly used in industrial engineering to perform clustering analysis. Therefore, this study aimed to define and cluster tennis match types based on how they are played. This is considered highly meaningful in the field of sports science, because it analyzes tennis match content based on AI.

Methods

Research materials

The research data selected for this study were from the 100th round of 32 matches of the five finals of the 2023 International Tennis Open Tournament. The characteristics of the inter-national open competitions selected as the research data are listed in Table 1.

Research variables and data collection

To select the research variables, a literature review was first conducted [10, 13, 22–24]. Thereafter, a meeting of seven experts, consisting of tennis players and coaches, was held to finally set 27 variables in seven areas (Table 2). Table 2 shows the final selected variables, and

Figure 1 shows the serve and rally positions. Data collection was conducted by five tennis players who manually recorded match data using Excel while directly reviewing the match videos. To enhance the validity and reliability of the recorded data, they underwent training before data recording. The final dataset used in the study is publicly available (URL: (https://github.com/gywns86/datas/tree/main/data).

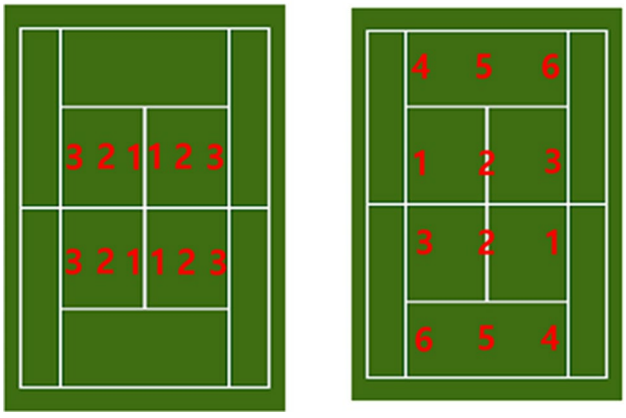

Fig. 1 A code the serve location(left) and rally location(right)

Data analysis

To achieve this study's purpose, three models were applied, and the model with the best performance was selected. The processes of the three models are illustrated in Figure 2.

Model 1 applies K-means clustering to raw data without dimensionality reduction. Model 2 reduces the dimensions of the raw data through principal component analysis and applied the k-means algorithm. Model 3 goes through transformer embedding in raw data, reduces the dimensions through principal component analysis, and applies the k-means algorithm. The reason for reducing the dimensions using principal component analysis in the two models is to conduct principal component analysis to solve the problem of performance deterioration when there are many independent variable inputs in the case of distance-based clustering algorithms, such as the k-means algorithm.

The transformer algorithm applied to Model 3 was first proposed by Vaswani et al. in 2017 and showed excellent performance in natural language processing [25]. The Transformer model is capable of parallel processing, significantly improving the computation speed, and unlike RNNs or LSTMs, it efficiently handles long-term dependencies regardless of the sequence length [26]. The transformer consists of an encoder and a decoder; however, in this study, only the encoder part of the transformer was used. Specifically, the BERT model developed by Google was used. Therefore, the pre-trained weights of the BERT

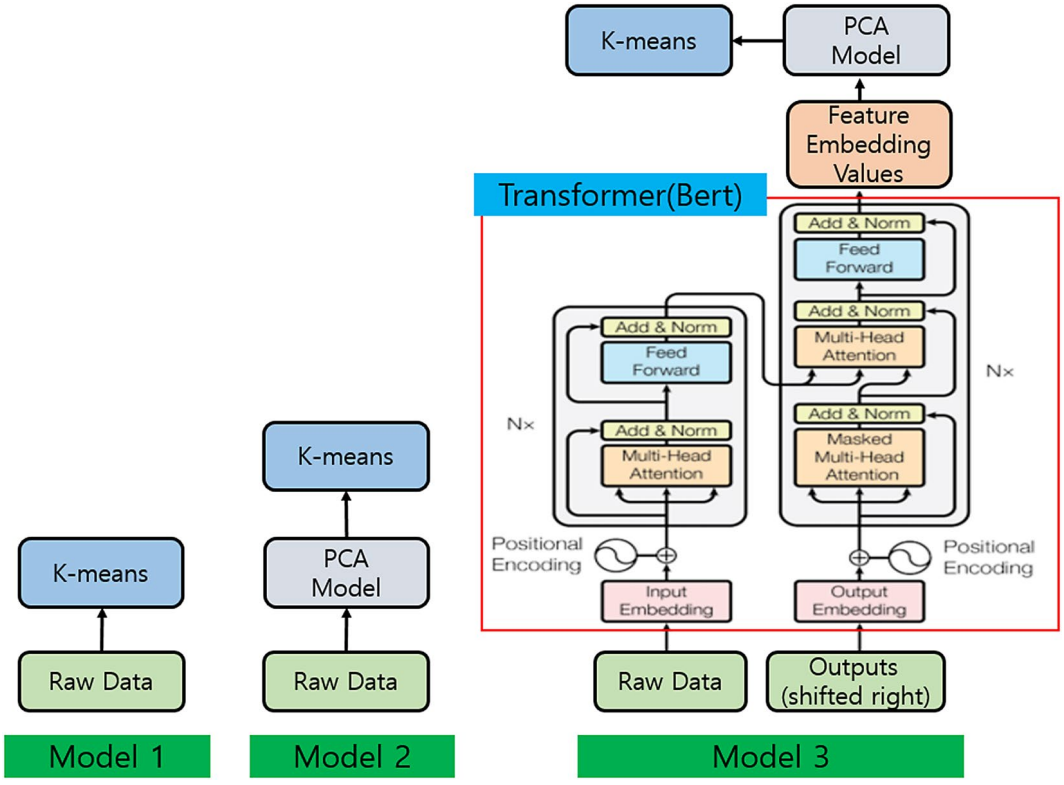

Fig. 2 The processes of the three models

Table 3 Silhouette coefficient calculation results for selecting the optimal number of clusters for each model

number of clusters (k)	1st			2nd	
	Model 1	Model 2	Model 3	Model 3-1	Model 3-2
2	0.246	0.402	0.406	0.392	0.440
3	0.190	0.368	0.396	0.380	0.435
4	0.155	0.350	0.351	0.364	0.413
5	0.130	0.343	0.211	0.356	0.429
6	0.136	0.362	0.145	0.359	0.398
7	0.129	0.255	0.091	0.364	0.407
8	0.125	0.216	0.079	0.378	0.411
9	0.116	0.214	0.123	0.370	0.415

model were used from the values provided by Google, and the features extracted through the BERT model were embedded into a 768-dimensional space from the initial 27 dimensions.

The k-means algorithm of the unsupervised learning series was used to cluster game types. An unsupervised learning algorithm must determine the number of clusters in a dictionary, and the results appear to differ depending on the number of clusters. Therefore, in this study, a silhouette coefficient was calculated to identify the optimal number of clusters. Specifically, the number of clusters was adjusted from two to nine to calculate the silhouette coefficient, and the model and number of clusters with the highest silhouette coefficients were selected. Formula 1 presents the silhouette coefficient calculation formula. a(i) refers to the average value of the distance to data within the cluster to which ì belongs, and b(i) refers to the minimum value of the average distance to data from the cluster to which ì does not belong.

$$s(\grave{\imath}) = (b(i) - a(i)) / (\max(a(i), b(i))) \quad (1)$$

The researchers subjectively judged the names of the clusters. Therefore, a difference test was conducted on game record variables based on the cluster results. This was performed to identify the characteristics of the clusters, and an independent sample t-test was used to test the differences. All statistical significance levels were set at 0.05, and Python 3 and SPSS Ver 25.0, were used for analysis.

Results

Analysis of tennis player's game type

To determine the optimal number of clusters, the number of clusters was adjusted from two to nine to calculate the silhouette coefficient. The results are summarized in Table 3.

For all three models, the silhouette coefficient was the highest when the number of clusters was set to two, with Model 3 achieving the highest silhouette coefficient. Therefore, the performance of the model that reduced the dimension through principal component analysis and applied the k-means algorithm after transformer embedding from the raw data was found to be excellent, and setting the number of clusters to two was found to be appropriate. Finally, in Model 3, the number of clusters was set to two, and the game types were analyzed. There were 248 and 152 clusters in Clusters 1 and 2, respectively. The visual results of the clustering are shown in Fig. 3.

Based on the results of the first round, to confirm the detailed clustering of Model 3, the number of clusters was adjusted from two to nine in the second round to calculate the silhouette coefficient. The silhouette

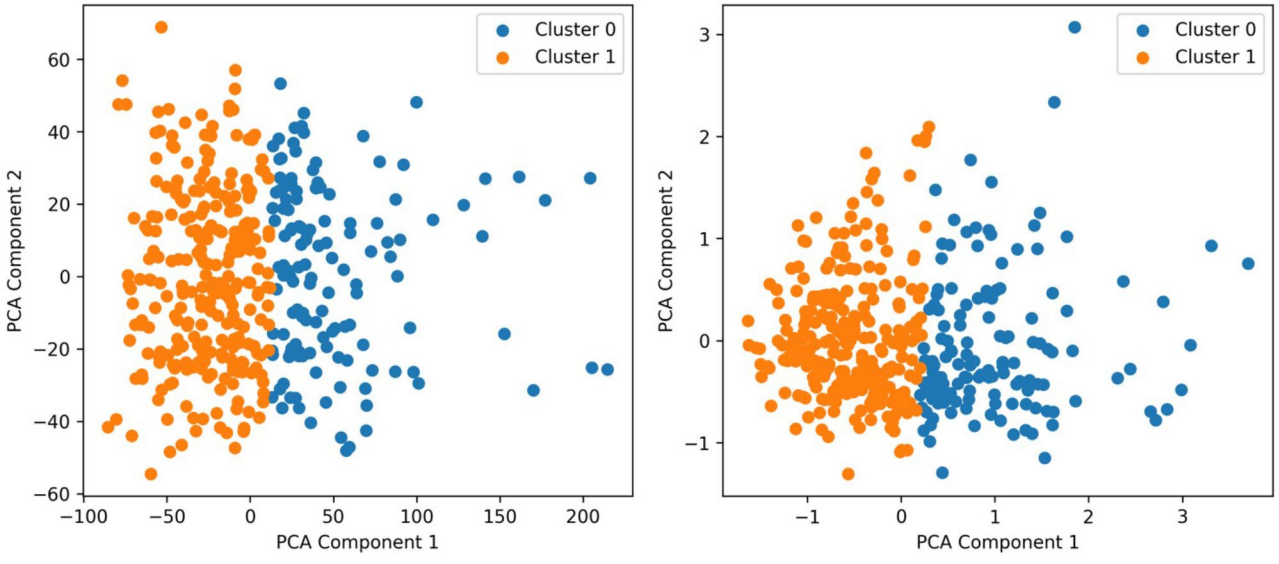

Fig. 3 K-means analysis visualization of Model 1 (left) and Model 2 (right)

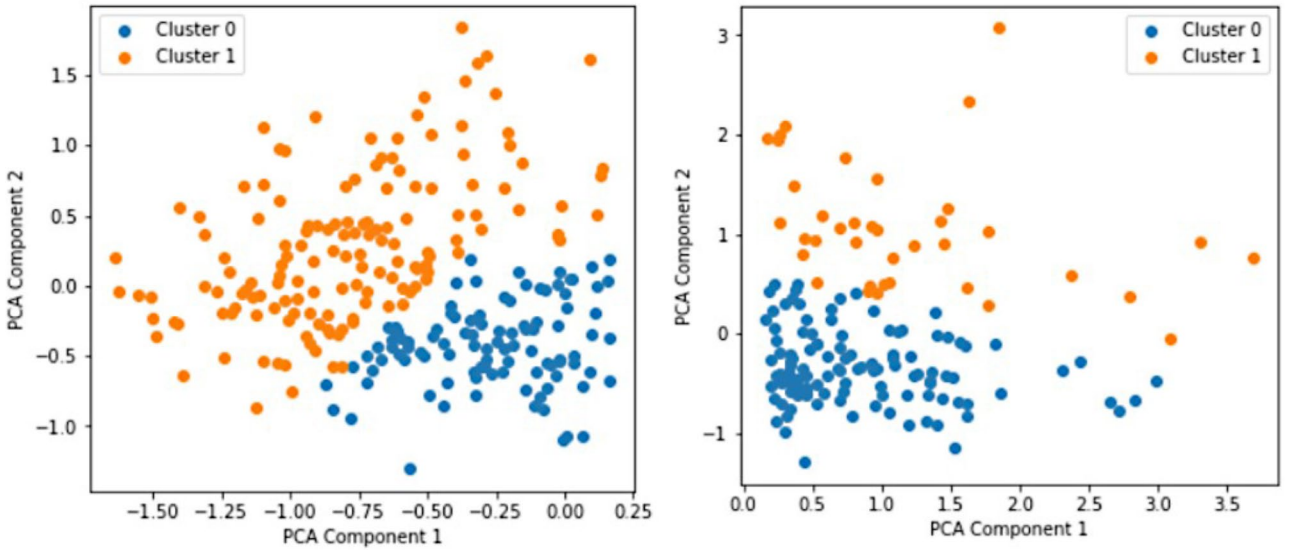

Fig. 4 K-means analysis visualization of cluster 2 – 1 (left) and cluster 2–2 (right)

Table 4 Results of the variable difference test between clusters 1 and 2

Variables	cluster	M	SD	Variables	cluster	M	SD	Variables	cluster	M	SD
First serve	1	66.4	9.7	Forevolley***	1	1.7	1.8	Rally Location 6	1	30.5	10.0
	2	65.3	11.7		2	1.1	1.4		2	30.1	9.1
Second serve	1	33.6	9.7	Backvolley***	1	1.4	1.5	Winning shot	1	34.2	13.9
	2	34.7	11.7		2	0.9	1.3		2	36.2	16.2
Serve Location 1**	1	40.6	10.7	Smashing***	1	1.0	1.1	Unforced error	1	10.8	11.9
	2	43.5	10.8		2	0.5	0.9		2	9.6	12.1
Serve Location 2*	1	15.5	10.9	Dropshot**	1	1.2	1.4	Error	1	55.1	18.6
	2	13.2	10.1		2	0.8	1.1		2	54.1	19.2
Serve Location 3	1	43.9	11.7	Rally Location 1**	1	3.9	2.7	Double Fault	1	4.8	5.3
	2	43.3	10.5		2	3.1	2.9		2	3.8	6.3
Forehand*	1	46.7	7.7	Rally Location 2	1	7.3	6.3	Net	1	40.9	11.7
	2	48.5	7.8		2	6.9	5.8		2	38.9	14.1
Backhand	1	37.8	9.6	Rally Location 3	1	5.1	3.4	Left and Right of Court	1	22.2	10.4
	2	38.9	8.7		2	4.4	3.9		2	21.5	11.9
Foreslice	1	2.8	3.1	Rally Location 4	1	22.3	7.6	Back of Court**	1	32.0	12.2
	2	2.4	3		2	22.9	6.7		2	35.7	12.8
Backslice	1	7.4	6.3	Rally Location 5	1	30.8	10.1	Number of Rally***	1	169.8	45.9
	2	6.9	6.2		2	32.6	10.9		2	136.9	41.9

coefficients of Clusters 1 and 2 were highest when the number of clusters was set to two. Therefore, the number of clusters was set to two and detailed clustering was performed. Cluster 1 (1–1) of Cluster 1 was clustered with 102 items, Cluster 2 (1–2) with 146 items, and Cluster 1 (2–2) of Cluster 2 with 102 items. Cluster 1) was clustered into 114 cases and Cluster 2 was clustered into 38 cases. The visual results of the clustering are shown in Fig. 4.

Verification of differences in major variables by tennis player game type

To cluster tennis players' game types and identify the characteristics of the clustered game types, differences in key variables according to game type were tested. The results of testing the differences in economic variables between Clusters 1 and 2 are shown in Table 4. The results showed that serve position 1 (t=-2.650, p=.008), serve position 2 (t=2.167, p=.031), forehand (t=-2.247, p=.025), fore volleys (t=3.655), p<.001), back volley (t=3.653, p<.001), smashing (t=3.988, p<.001), drop shot (t=3.241, p=.001), position 1 (t=2.876), p=.004), behind-the-error (t=-2.865, p=.004), and number of rallies (t=7.184, p<.001) showed statistically significant differences. Specifically, the variables of serve position 1, forehand, and behind-the-error were found to be higher in Cluster 2 than in Cluster 1; Cluster 1 was found to be higher than in Cluster 2, while for the variables of serve

Table 5 Results of the variable difference test between clusters 1-1 and 1-2

Variables	cluster	M	SD	Variables	cluster	M	SD	Variables	cluster	M	SD
First serve	1-1	66.8	9.5	Forevolley*	1-1	1.9	1.9	Rally Location 6	1-1	30.6	10.6
	1-2	65.8	10.1		1-2	1.4	1.6		1-2	30.4	9.1
Second serve	1-1	33.2	9.5	Backvolley***	1-1	1.7	1.7	Winning shot	1-1	35.3	13.2
	1-2	34.2	10.1		1-2	1.0	1.1		1-2	32.5	14.8
Serve Location 1	1-1	41.1	11.0	Smashing*	1-1	1.1	1.2	Unforced Error*	1-1	12.0	13.2
	1-2	39.8	10.3		1-2	0.7	0.9		1-2	9.0	9.7
Serve Location 2	1-1	15.3	10.7	Dropshot	1-1	1.3	1.5	Error*	1-1	52.7	17.9
	1-2	15.9	11.2		1-2	1.0	1.3		1-2	58.5	19.1
Serve Location 3	1-1	43.6	12.3	Rally Location 1*	1-1	4.2	2.8	Double Fault*	1-1	5.4	5.6
	1-2	44.2	10.8		1-2	3.4	2.5		1-2	4.0	4.7
Forehand	1-1	46.3	8.0	Rally Location 2*	1-1	8.1	6.7	Net	1-1	40.4	11.7
	1-2	47.4	7.3		1-2	6.3	5.4		1-2	41.7	11.8
Backhand	1-1	37.2	9.5	Rally Location 3	1-1	5.2	3.4	Left and Right of Court	1-1	22.4	9.5
	1-2	38.6	9.7		1-2	4.9	3.4		1-2	21.9	11.5
Foreslice*	1-1	3.2	3.4	Rally Location 4*	1-1	21.5	7.2	Back of Court	1-1	31.8	12.8
	1-2	2.3	2.5		1-2	23.6	7.9		1-2	32.3	11.3
Backslice	1-1	7.3	5.8	Rally Location 5	1-1	30.4	10.6	Number of Rally	1-1	170.5	47.1
	1-2	7.5	6.8		1-2	31.4	9.3		1-2	168.7	44.5

*$p<.05$, **$p<.01$, ***$p<.001$

Table 6 Results of the variable difference test between clusters 2-1 and 2-2

Variables	cluster	M	SD	Variables	cluster	M	SD	Variables	cluster	M	SD
First serve	2-1	65.7	11.7	Forevolley	2-1	1.1	1.5	Rally Location 6	2-1	29.6	8.5
	2-2	64.0	11.8		2-2	0.9	1.3		2-2	31.5	10.8
Second serve	2-1	34.3	11.7	Backvolley*	2-1	1.0	1.5	Winning shot	2-1	35.7	17.0
	2-2	36.0	11.8		2-2	0.6	0.7		2-2	37.7	13.7
Serve Location 1	2-1	43.0	10.4	Smashing	2-1	0.5	0.9	Unforced Error	2-1	9.7	10.9
	2-2	45.2	12.0		2-2	0.6	0.9		2-2	9.4	15.5
Serve Location 2**	2-1	14.4	10.4	Dropshot	2-1	0.8	1.2	Error	2-1	54.5	19.7
	2-2	9.5	8.4		2-2	0.6	0.9		2-2	52.9	17.9
Serve Location 3	2-1	42.6	10.3	Rally Location 1	2-1	3.2	2.9	Double Fault	2-1	4.2	6.5
	2-2	45.4	11.0		2-2	2.7	2.6		2-2	2.7	5.5
Forehand	2-1	48.6	7.8	Rally Location 2	2-1	7.1	5.8	Net	2-1	39.8	13.9
	2-2	48.2	8.1		2-2	6.4	5.6		2-2	36.5	14.4
Backhand	2-1	38.8	9.2	Rally Location 3	2-1	4.7	4.2	Left and Right of Court	2-1	21.4	11.3
	2-2	39.4	7.4		2-2	3.6	2.5		2-2	21.8	13.6
Foreslice	2-1	2.4	3.0	Rally Location 4	2-1	22.4	6.8	Back of Court	2-1	34.6	13.0
	2-2	2.5	2.9		2-2	24.3	5.9		2-2	39.1	11.8
Backslice	2-1	6.7	6.2	Rally Location 5	2-1	33.0	10.6	Number of Rally	2-1	137.6	44.2
	2-2	7.3	6.1		2-2	31.5	11.8		2-2	134.8	34.8

*$p<.05$, **$p<.01$, ***$p<.001$

position 2, fore volleys, back volleys, smashing, drop shots, position 1, and the number of rallies.

Table 5 presents the results of testing the differences in the key variables between Clusters 1-1 and 1-2. Statistically significant differences were found in the variables of the fore slices(t=2.538, p=.012), fore volleys(t=2.069, p=.04), back volleys(t=3.752, p<.001), position 1(t=2.274, p=.024), position 2(t=2.280, p=.023), position 4(t=-2.159, p=.032), unforced error(t=2.078, p=.039), error (t=-2.471, p=.014) and double faults (t=2.043, p=.042). Specifically, location 4 and the error variables were higher in Cluster 1-2 than in Cluster 1-1. However, Cluster 1-1 was found to be higher than Cluster 1-2 in the variables for fore slices, fore volleys, back volleys, locations 1 and 2, unforced errors, and double faults.

Table 6 presents the results of testing the differences in the key variables between Clusters 2 – 1 and 2–2. There was a statistically significant difference in the variables of sublocation 2(t = 2.655, p = .009) and back volley (t = 2.311, p = .022), Cluster 2 – 1 was found to be higher than Cluster 2–2 in both sublocation 2 and back volley.

Discussion

In this study, tennis game types were distinguished to complement the limitations of the previous research on tennis game analyses conducted so far. The results of this study are as follows: The model using deep learning exhibited excellent performance, and the tennis players were clustered into four types. Cluster results were obtained through mathematical calculations; however, the name of the cluster was provided directly by the researcher based on the results. Based on the results of this study, the researchers assigned names to each cluster as follows: First, in the case of Cluster 1–1, the game type was named "NEt Rusher Defensive type: NERD." In the case of NERD, the number of rallies was relatively low, and the game was played near the court. Among the four game types, the winning rate analysis by overall type showed that it had the highest winning rate at 58.8%, and the representative player, "Taylor Fritz," was clustering as the NERD type. "Taylor Fritz" is known for his style of playing the game by using a strong serve and mainly net dashing when the opponent is in front of the court. Taylor Fritz is commonly known for his aggressive play; however, in this study, he was clustered as a net rusher defensive type of NERD. The biggest reason for this is believed to be the NERD type, which is defensive rather than offensive because net dashes are mainly performed.

For Clusters 1–2, the game type was named "ALl Courter Defensive type: ALCD." ALCD is a game that involves both defensive and overall activities. The ALCD was an all-court game and had the lowest winning rate (47.3%) of the four game types. This reason is relatively defensive-oriented; however, it is believed that errors occur frequently and affect win rates. Representative players with the ALCD type in this study were "Carlos Alcaraz," "Daniil Madvedev," and "Holger Rune." Daniil Madvedev is a representative player in the ALCD. He is known as a player with a high ability to handle the court and a good eye to the ball, so he does not often get a hit with winning shots. This type of player is clustered as having an excellent ability to pass balls that touch their racket onto the opponent's court without making errors.

In the case of Cluster 2 – 1, the game type was named "STroke Placement Offensive type: STPO." STPO is an offensive type, in which the stroke sends the ball to the opponent's empty court space. STPO was found to have the second highest win rate (50.0%) among the four cluster match types. Representative players for the STPO's game type were "Novak Djokovic," "Jannik Sinner," "Andrey Rublev," "Stefanos Tsitsipas," and "Alexander Zverev." It was clustered as a type of match with most players in the top 10 rankings. The STPO type has five players in the top 10, including the world's number-one Novak Djokovic player. Novak Djokovic's characteristics include stroke placement; he is a high-power player who can play by taking control the game. He also often had an upper hand in stroke competitions with opposing players at baseline. Because they dominate the forehand and backhand, which are the most commonly used tennis skills, it is believed that there are quite a few players with excellent performance in this type.

In the case of Cluster 2–2, the game type was named "SErve Placement Offensive type: SEPO." In the case of the SEPO, the offensive type sends the server to the 1st and 3rd court areas. SEPO has the lowest win rate (36.8%) among the four cluster matching types. The representative player for the SEPO game was "Hubert Hurkacz." Hubert is a representative player in the SEPO. It was confirmed that players with excellent serve placement had a high scoring rate when making a successful first serve and a high scoring rate when making a successful second serve. In addition, indicators related to serve are ranked first; therefore, the probability of maintaining a serve game is high, and a player leads the game by starting with a serve game. Clustering match types can help in planning specific techniques and strategies in advance based on the type of match in the actual game. For example, if Type A plays against Type B, strategies can be planned in advance by reviewing factors such as service and court positioning to create opportunities for scoring, which could positively contribute to enhancing performance.

This study was conducted to conduct a deep learning-based analysis of game content according to tennis game type. Although studies related to tennis have been continuously reported [27–29], it is significant in that the content of the game was analyzed by dividing the game type of tennis players. Most previous studies have focused on analyzing match data or clustering match types using a single variable to explain players [24, 30]. For example, previous studies have analyzed tennis matches or classified match types using single variables, such as serve, rally, match duration, or scoring points. However, this study is significant in that it explains match types using multiple variables rather than relying on a single variable. This comprehensive approach enhances the explanatory power of the study by providing a broader understanding of the players' overall match types. In particular, this study employed the transformer model, a recent deep learning-based methodology, to classify player match types, contributing to improved classification accuracy. The application of this advanced analytical method to the fields of physical education and sports further highlights the significance of this study. Nevertheless, it should be noted that this study classified match types using data from only five major tournaments, which presents a limitation. Moreover, as tennis match types can vary depending on the opponent, future studies should consider the relational aspects of match types to achieve more comprehensive findings. Additionally, the K-means algorithm

used in this study is an efficient and widely adopted clustering method; however, its performance may be limited depending on the shape or density of the clusters. Therefore, future research should apply various algorithms, such as density-based clustering (DBSCAN) and hierarchical clustering, to compare their performances and conduct an in-depth analysis of how each method affects the classification of tennis match types. Such follow-up studies are expected to enable more accurate and meaningful classification of match types.

Conclusion

Tennis players can be classified into four types: First, it can be clustered into the NERD (NEt Rusher Defensive (NERD) type, which is defensive and induces net play. Second, the ALCD (ALl Courter Defensive) type can be clustered as either defensive or all-round. Third, it can be clustering as the STPO (STroke Placement Offensive) type, which is aggressive and has strengths in stroke. Fourth, the SEPO (SErve Placement Offensive) type can be clustering as aggressive and has strengths in sub courses. This conclusion is expected to not only be used as basic data to cluster game types in tennis matches but also to contribute to establishing game strategies for each game type and further improving performance.

Acknowledgements
Not applicable.

Author contributions
Study concept and design: Yun, H., Jeon, M. Acquisition of data: Jang, N.Analysis and interpretation of data: Yun, H.Drafting of the manuscript: Yun, H., Jeon, M.Critical revision of the manuscript for important intellectual content: Jang, N.Statistical analysis: Yun, H., Jeon, M.Administrative, technical, and material support: Yun, H., Jeon, M.Study supervision: Jang, N.All authors read and approved the final manuscript.

Funding
Not applicable.

Data availability
The data that support the findings of this study are available on online (https://github.com/gywns86/datas/tree/main/data).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 December 2024 / Accepted: 4 April 2025
Published online: 28 April 2025

References

1. Kim S, Do J. An assessment of relative performance by clustering of strength by professional team in the Korean basketball league. Korean J Sport Sci. 2015;24(3):1589–603.
2. Kim Y, Park H. Cluster analysis of players through Korean women's professional golf game records. Korean J Sport Sci. 2021;30(2):1025–32.
3. Sarmento H, Marcelino R, Anguera MT, Campaniço J, Matos N, Leitão JC. Match analysis in football: a systematic review. J Sports Sci. 2014;32(20):1831–43.
4. Bunker R, Susnjak T. The application of machine learning techniques for predicting match results in team sport: A review. J Artif Intell Res. 2022;73:1285–322.
5. Kim D, Jeong K. The enhancement of Taekwondo competition performance by analyzing between Korean and foreign athletes in the Liu olympic. Sport Sci. 2019;36(2):117–24.
6. Lee S. Analysis of table tennis game and point system on Ace China players. Korean J Meas Eval Phys Educ Sport Sci. 2012;14(3):79–93.
7. Choi G. The relationship between coaches-athletes relations and team atmosphere, depending on interactions between coaches and athletes. Korean J Sport Sci. 2020;29(6):625–37.
8. Ha S. Development and utilization of playbooks to improve basketball team tactics, strategies and teamwork. J Coaching Dev. 2022;24(5):164–74.
9. Alamar B. Sports analytics: A guide for coaches, managers, and other decision makers. Columbia University; 2024.
10. Hong S, Kim J, Noh G. Preliminary study on analysis of tennis game. Korean J Meas Eval Phys Educ Sport Sci. 2010;12(2):68–75.
11. Kim S, Lee K, Do J. Applicability and limitation of hidden Markov model and sports coding in tennis. Korean J Sport Sci. 2017;26(5):1325–34.
12. Kim H, Hong Y, Kim Y. Investigation of psychological hindrance type and coping strategies in badminton players. Korean J Sport Sci. 2017;28(4):1006–19.
13. Bermejo JP, Ruano MA. Entering tennis Men's grand slams within the top-10 and its relationship with the fact of winning the tournament. RICYDE Rev Int Cienc Deporte. 2016;12(46):410–22.
14. Song H, Li Y, Pan P, Yuan B, Liu T. Multilayer network framework and metrics for table tennis analysis: integrating network science, entropy, and machine learning. Chaos Solitons Fractals. 2025;191:115893.
15. Bunker R, Yeung C, Susnjak T, Espie C, Fujii K. A comparative evaluation of Elo ratings-and machine learning-based methods for tennis match result prediction. Proc Inst Mech Eng P J Sports Eng Technol. 2024;238(4):305–16.
16. Lee Y, Choi S. The kinematic analysis of tennis slice service motion. Korean J Sport Sci. 2013;22(4):1289–95.
17. Yoo H, Hwang Y. Effects of the factors related to serve, receive, and break point on Men's doubles in tennis. J Korean Soc Study Phys Educ. 2012;17(2):13–22.
18. Crespo M, Martínez-Gallego R, Filipcic A. Determining the tactical and technical level of competitive tennis players using a competency model: a systematic review. Front Sports Act Living. 2024;6:1406846.
19. Kim H, Park J, Shin B. Notational analysis of Woman's grand slam tennis game. Korea Sport Res. 2007;18(2):321–32.
20. Lee K, Lee Y, Lee G. The notational analysis of the domestic Man's single tennis game. Korean J Phys Educ. 2004;43(3):903–11.
21. Lee M, Kim M, Lee C. An analysis of the relationship between leisure attitude and quality of life of active seniors participating in tennis: focusing on the mediating effect of stress-related growth. J Leisure Recreat Stud. 2020;44(2):17–29.
22. Renò V, Mosca N, Nitti M, D'Orazio T, Guaragnella C, Campagnoli D, et al. A technology platform for automatic high-level tennis game analysis. Comput Vis Image Underst. 2017;159:164–75.
23. Torres-Luque G, Ramirez A, Cabello-Manrique D, Nikolaidis TP, Alvero-Cruz JR. Match analysis of elite players during paddle tennis competition. Int J Perform Anal Sport. 2015;15(3):1135–44.
24. Cui Y, Liu H, Liu H, Gómez MA. Data-driven analysis of point-by-point performance for male tennis player in grand slams. Motricidade. 2019;15(1):49–61.
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
26. Kim T, Kim S, Lim K. A study on utilization of vision transformer for CTR prediction. Knowl Manag Res. 2021;22(4):27–40.
27. Shin K, Lee Y, Ko J. Biomechanics of open stance forehand stroke in tennis: expert vs. novice. Asian J Phys Educ Sport Sci. 2023;11(6):71–80.
28. Choi G, Jang N. The influence of perceived coach-athlete interaction on training engagement behaviors and self-efficacy in tennis players. Sport Sci. 2023;41(2):189–96.

29. Subagja DS, Kusmaedi N, Komarudin K. The effect of learning media and coordination to forehand top spin accuracy on table tennis. JUARA J Olahraga. 2019;4(2):220–8.
30. Fan X, Li Z, Zhang X, Yang R, Tan H, Chen Y et al. Data-Driven Tennis Strategy Evaluation through Hierarchical Markov Models. In: Proc 2024 9th Int Conf Intell Inf Process. 2024:219–25.

Publisher's note