Identification and Prediction of Tennis Players' Technical and Tactical Behaviors Based on Transformer Model

Juan Pu[1], Xianyi Zha[2]
[1]Physical Education College, Jiangxi Normal University, Nanchang 330022, Jiangxi, China
[2]Department of Physical Education, Qingdao Agricultural University, Qingdao 266109, Shandong, China
Corresponding author: Juan Pu (e-mail: PUJUAN127@163.com)

ABSTRACT This research addresses the challenge of automatically identifying complex tennis techniques and tactics, dynamically predicting shot intentions, and assessing round outcomes early from structured event sequences. By constructing a Transformer sequence model, the study provides a data-driven intelligent solution for tactical decision-making. The model uses structured shot sequences from Match Charting Project logs, integrating attributes like shot type, direction, and match status. It employs a six-layer Transformer encoder with a multi-head self-attention mechanism for multi-task learning, jointly optimizing technical/tactical classification, shot intention prediction, and early outcome assessment. Experiments validate its robust performance, achieving a Macro-F1 of 0.935 for tactical classification and a round outcome AUC@5 of 0.923. The model significantly outperforms BiLSTM and CNN-LSTM baselines, with ablation studies confirming the critical contributions of its embedding and multi-task design. The proposed framework enables end-to-end deep analysis and effective prediction of tennis technical and tactical behaviors without visual input, providing a scalable, data-driven solution for intelligent competitive analysis, principles that may also inform structured event-stream prediction in other non-visual engineering datasets.

INDEX TERMS tennis analytics, transformer model, tactical behavior recognition, sequence prediction, event-stream modeling

I. INTRODUCTION

TENNIS is a competitive sport that relies heavily on a combination of techniques and tactics, where athletes' shot selection, landing point control, and rhythm changes constitute a complex decision sequence. This need for continuous, optimal sequencing mirrors the complexity found in automated manufacturing, where realtime decisions regarding material feed rates, tension control, and pattern execution must be perfectly coordinated to ensure product quality and production efficiency [1,2]. Traditional technical and tactical analysis mainly relies on manual video observation and frequency statistics, which have limitations such as strong subjectivity, low efficiency, and difficulty in capturing long-term dependencies [3,4]. With the structured accumulation of professional tennis tournament data, intelligent analysis based on event sequences has become possible [5]. However, existing methods mostly focus on isolated shot statistics or short-term pattern recognition, lacking the ability to dynamically model the entire chain of "technique-tactics-results". How to achieve automatic recognition of complex technical and tactical behaviors, dynamic prediction of shot intentions, and early judgment of round results from discrete shot event sequences [6,7], these three progressive problems together constitute the core challenges of current data-driven competitive analysis. A sequence modeling framework that can comprehensively model long-term dependencies, integrate multi-dimensional context, and support multi-task joint learning is urgently needed to systematically solve them.

In the field of sports technical and tactical analysis, early studies mostly used Markov models or decision trees to model shot sequences, analyzing serve and receive strategies by using transition probabilities. However, their strong Markov assumptions make it difficult to characterize the long-range dependencies of multishot layouts [8,9]. With the rise of deep learning, recurrent neural networks and their variants (such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit)) [10,11] have been used for tactical recognition in sports such as football and basketball. In the field of tennis, Anderson et al. [12] analyzed round win rates based on MCP (Match Charting Project) data, but did not model sequence dynamics; related studies attempted to use neural networks to identify shot types, focusing only on the technical level of a single shot and ignoring tactical context [13]. Recently, graph neural networks have been applied to two-player confrontation modeling, such as constructing player positions as dynamic graphs to predict movement trajectories, but such methods rely on visual or tracking data and cannot be applied to pure event logs [14,15]. In general, existing sports intelligence research is either limited by the model's expressive capabilities or relies on highdimensional sensory input, and has not yet achieved end-to-end sequential understanding of tennis technical and tactical behaviors under non-visual conditions. In terms of sequence modeling, Transformer has made breakthroughs in natural language processing with its self-attention mechanism and has gradually penetrated into the field of time series analysis. Temporal Fusion Transformer [16,17] effectively processes multivariate time series by separating static and dynamic variables, and performs well in energy and financial forecasting, but its applicability to discrete event sequences has not yet been verified. Some studies have attempted to use Transformer [18,19] for motion analysis. Transformer is used to process badminton video action recognition, but it relies heavily on visual features; Transformer is used for e-sports player behavior prediction, although it is based on event logs, but it does not distinguish between technical and tactical dimensions. In tennis-specific research, only a few works have explored sequence models: using a single-layer Transformer to predict the landing point of a serve [20], the task granularity is relatively coarse; combining LSTM with attention mechanism for action classification [21,22], but the attention mechanism is local and cannot be learned. It is worth noting that current research generally separates technical and tactical recognition, intention prediction and result judgment, lacks a unified multi-task learning framework, and does not fully integrate key contexts such as game status and player style. In summary, existing work has obvious deficiencies in the sequential modeling capabilities of the model architecture, the technical and tactical coupling of task design, and the depth of integration of contextual information.

To address these limitations, this paper proposes a Transformer-based multi-task sequence modeling framework specifically for the recognition and prediction of technical and tactical behaviors in tennis. First, unlike existing methods that focus solely on single-shot techniques or isolated statistics, this paper considers each point as a "behavior sequence" consisting of hitting events. It explicitly defines composite technical and tactical behavior categories encompassing technique type, hitting direction, and tactical effect, enabling finegrained recognition of coupled technical and tactical patterns. Second, for the task of intention prediction, it abandons static assumptions and leverages the autoregressive capabilities of the Transformer to dynamically predict the next shot's attributes (type, direction, and finishing risk), effectively capturing long-range tactical planning. Furthermore, the model further enables early prediction of round outcomes, serving as a comprehensive validation of the aforementioned recognition and prediction capabilities. Third, to enhance the model's contextual awareness, score status, break point indicators, and player IDs are embedded into the sequence input, enabling the model to learn decision-making differences in high-pressure situations. Finally, through a shared encoder, it achieves joint learning of technical and tactical recognition and hitting intention prediction. Based on the learned representations, it naturally derives the ability to predict round outcomes, avoiding information loss caused by task fragmentation. This study not only verifies the feasibility of deep technical and tactical analysis of pure event sequences without visual conditions but also constructs the first end-to-end Transformer benchmark for understanding tennis technical and tactical behavior. This scalable methodological framework for intelligent competitive analysis finds a direct parallel in its potential to process non-visual, high-speed sensor data for autonomous defect identification and predictive quality control within complex manufacturing processes.

In order to systematically solve the core challenges in tennis technical and tactical behavior modeling, this paper focuses on three clearly defined and progressive multi task objectives:

(1) technical and tactical classification - identify the macro technical and tactical patterns of the whole round, and output six types of coupling technical and tactical behavior labels (such as "serve online", "baseline stalemate", etc.);

(2) Hitting intention prediction: Based on the historical hitting sequence, dynamically predict the technology type, hitting direction and ending risk of the next shot, so as to achieve fine-grained and forward-looking behavior deduction;

(3) Early round outcome assessment - when the round is up to the k-th shot ($K \geq 1$), predict the final winner of the round based on the existing information as a comprehensive verification of the above recognition and prediction ability.

The three correspond to the logical chain of "behavior understanding intention deduction result prediction" respectively, and together form an end-to-end intelligent analysis framework of technology and tactics.

II. TECHNICAL AND TACTICAL ANALYSIS MODEL BASED ON TRANSFORMER

A. DATA SOURCE AND SEQUENCE CONSTRUCTION

This study uses the Match Charting Project (MCP) public dataset (https://github.com/JeffSackmann/tennis_MatchChartingProject), which contains shot-by-shot event logs. MCP has the advantages of high temporal resolution, strong structure, and good annotation consistency, and has been widely adopted in many tennis analysis studies.

In MCP raw data, each point is recorded as an independent event sequence, and each beat contains multiple fields. This study extracts core fields from it to construct a technical and tactical behavior sequence: server: server player ID; returner: receiver player ID; p1_score, p2_score: current score; game, set: current game and set number; serve_width: serve direction ("B"=Body, "C"=Center, "W"=Wide); serve_depth: serve depth ("D"=Deep, "N"=Normal); serve_type: serve type ("F"=Flat, "S"=Slice, "T"=Topspin);

shot#_type (shot1_type, shot2_type, ...): shot type of the nth shot ("F"=Forehand, "B"=Backhand, "V"=Volley, "O"=Overhead, "S"=Serve, "X"=Unforced Error, "Z"=Forced Error, "W"=Winner); shot#_dir: shot direction ("D"=Down-the-line (DTL), "C"=Cross-court, "M"=Middle); point_victor: winner of the point (1=server wins, 2=receiver wins); rally_count: total number of reps in the point.

This study preprocesses the raw data and constructs sequences. Round extraction: Only complete rounds with rally_count $\geq$ 1 and point_victor $\in$ {1,2} are retained, eliminating interrupted or invalid points. Shot event normalization: Each shot is represented as a uniformly structured tuple (shot_type, direction, outcome, depth). Here, shot_type merges the original fields and maps them to 12 technical labels; direction is discretized into three categories; outcome is inferred from shot#_type; depth is set to "Unknown" only for nonserving rackets; serving rackets are classified according to serve_depth.

Sequence alignment: Starting with the serving racket, the sequence is arranged according to shot order, forming an event sequence of length rally_count. Context extraction: Calculates tactical context variables such as whether the current score is a break point (is_break_point) or game point (is_game_point). Ultimately, each round is represented as a structured sequence, and all discrete attributes are one-hot encoded or embedded. The specific encoding methods are shown in Table 1.

TABLE 1. Encoding methods

Attribute category	Encoding method
shot_type	Custom category mapping+One hot
direction	Rule mapping+One hot
outcome	Based on type inference+One hot
depth	Mapping+One hot
player_roles	Embedding
game_status	Logical rule calculation+One hot

Although the match charting project dataset has the advantages of high structure and good annotation consistency, it should be pointed out that its batting event records rely on manual observation and annotation, which may have subjective judgment bias or inconsistent annotation granularity. For example, the distinction between "forced error" and "unforced error" may vary according to the annotator's experience. In addition, the types of events, player styles and site conditions covered by MCP data have historical bias, which may not fully represent the distribution of current or future high-frequency and highly automated event data. Therefore, although this model shows excellent performance on MCP, its generalization ability on machine collected data completely generated by sensors or computer vision systems still needs to be further verified by cross data source migration experiments. Future work will explore fine-tuning strategies on automatic annotation datasets to enhance the adaptability of the model to non manual annotation sequences.

B. SEQUENCE REPRESENTATION AND CONTEXTUAL EMBEDDING

To model tennis matches [23,24] as learnable sequences, the original event logs need to be converted into structured embedding tensors and fused with tactical context information to support the Transformer's deep understanding of technical and tactical behavior. The textual representation and contextual embedding of a one-point sequence are shown in Figure 1.

Figure 1 illustrates the complete conversion process from raw MCP data to model input tensors. A CSV (Comma-Separated Values) record containing a score, player ID, and shot-by-shot information is parsed into structured rounds. Each shot is discretized into four-tuple tokens according to the rules in Table 1, with the special symbols [CLS]/[SEP] added. Subsequently, an embedding layer maps the token attributes, player ID, and match status to dense vectors. Finally, all embeddings are summed with the learnable positional encoding to form the final input tensor. This representation explicitly integrates technical and tactical behavior with situational context, providing a semantically complete sequence input for Transformer modeling. The input eigenvector of this model is fused by six kinds of attributes, and finally projected into a 256 dimensional unified embedded vector. Among them, the hitting type (12 types), direction (3 types), result (3 types) and depth (3 types) are all spliced into 55 dimensional sparse vectors by one hot coding, and then mapped to 256 dimensions by linear projection layer; The player role (server/receiver) can obtain two 64 dimensional vectors respectively by looking up the table through the learned player ID embedding table, and the projection is 256 dimensional after splicing; The game status (such as break point, bureau point, and even point) generates a three-dimensional binary flag through logical rules, which is also upgraded to 256 dimensions through the embedded layer; Finally, all attributes are embedded and added to the 256 dimensional sine position coding of the standard transformer to form the final input tensor. This process ensures that each batting event has a unified vector representation while preserving semantic integrity, providing a high reproducibility basis for sequence modeling.

C. MULTI-TASK TRANSFORMER ARCHITECTURE

1) Model Architecture

The multi task design of this model strictly corresponds to the above three goals: Head-1 is responsible for technical/tactical classification, and uses [CLS] representation as input to judge six types of technical and tactical modes for the whole round; Head-2 performs hitting intention prediction, and uses autoregressive method to predict the ternary attribute of the $t + 1$ beat at each step t; Head-3 realizes the early result evaluation and predicts the round winner (two categories) after any k-shot. The three share the underlying transformer encoder to ensure the consistency of representation, while maintaining the target specificity through the task specific header.

FIGURE 1. Textual representation and contextual embedding of a sequence

To achieve a comprehensive understanding of tennis skills and tactics, this paper designs a multi-task Transformer architecture [25,26]. By using a shared encoder, it jointly learns three tasks: skill and tactic recognition, shot intention prediction, and round outcome analysis. This method enhances representation generalization while maintaining model efficiency. The multi-task Transformer model structure is shown in Figure 2.

Figure 2 illustrates the overall architecture of the multi-task Transformer model proposed in this paper. The model takes a structured sequence of shot events as input. First, an embedding layer fuses attributes such as shot type, direction, effect, and depth for each shot with player ID, match status, and absolute position encoding to generate a context-enhanced input tensor. This tensor is fed into a six-layer shared Transformer Encoder, which outputs a sequence of contextual representations containing global dependencies. Given an input sequence $X = [x_0, x_1, \ldots, x_L] \in \mathbb{R}^{(L+1)\times d}$, where x_0 is the [CLS] token, the shared encoder generates a context-aware representation $H = [h_0, h_1, \ldots, h_L] = \mathrm{Encoder}(X)$ via $L_e = 6$ stacked Transformer blocks. Based on this representation, three task-specific heads operate in parallel: Head-1 utilizes the [CLS] token to classify six fine-grained technical and tactical behaviors across the entire round; Head-2 dynamically predicts the next shot's shot type, direction, and finishing risk based on an autoregressive mechanism and a teacher forcing strategy; and Head-3 predicts the final winner of the round at any K shots using the [CLS] representation. The three major tasks use cross entropy, multi-attribute cross entropy and binary cross entropy as loss functions respectively, and through weighted joint optimization, achieve in-depth analysis and forward-looking deduction of technical and tactical behaviors.

2) Shared Transformer Encoder

The core of this model is a 6-layer stacked standard Transformer Encoder. Each layer consists of a multi-head self-attention mechanism and a position feed-forward network, supplemented by residual connections and layer normalization. In each encoder layer, the multi-head self-attention sublayer linearly projects the input into a query, key, and value matrix:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{1}$$

In formula (1), $W^Q, W^K, W^V \in \mathbb{R}^{d\times d_k}$, $d_k = d/h = 32$ and $h = 8$ are the number of attention heads. Single-head attention is calculated as:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

The multi-head mechanism [27,28] concatenates the outputs of h heads and linearly transforms them:

$$\mathrm{MHSA}(X) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W^O \tag{3}$$

In formula (3), $W^O \in \mathbb{R}^{d\times d}$.

Residual connection and layer normalization:

$$Z = \mathrm{LayerNorm}(X + \mathrm{MHSA}(X)) \tag{4}$$

Feedforward network sublayer: It consists of two linear transformations and GELU (Gaussian Error Linear Unit) activation:

$$\mathrm{FFN}(Z) = ZW_1 + b_1 \xrightarrow{\mathrm{GELU}} A \tag{5}$$

$$\mathrm{Output} = AW_2 + b_2 \tag{6}$$

In formulas (5)-(6), $W_1 \in \mathbb{R}^{d\times d_{ff}}$, $W_2 \in \mathbb{R}^{d_{ff}\times d}$, $d_{ff} = 1024$.

Second residual connection:

$$X_{\mathrm{out}} = \mathrm{LayerNorm}(Z + \mathrm{FFN}(Z)) \tag{7}$$

D. TASK-SPECIFIC HEAD DESIGN

Head-1 (Technical and Tactical Behavior Classification) is designed to identify macro-technical and tactical patterns across a complete round. The input is the representation h_0 corresponding to [CLS], which is mapped to 6-dimensional logits via two MLP (Multi-Layer Perceptron) layers (including GELU and Dropout):

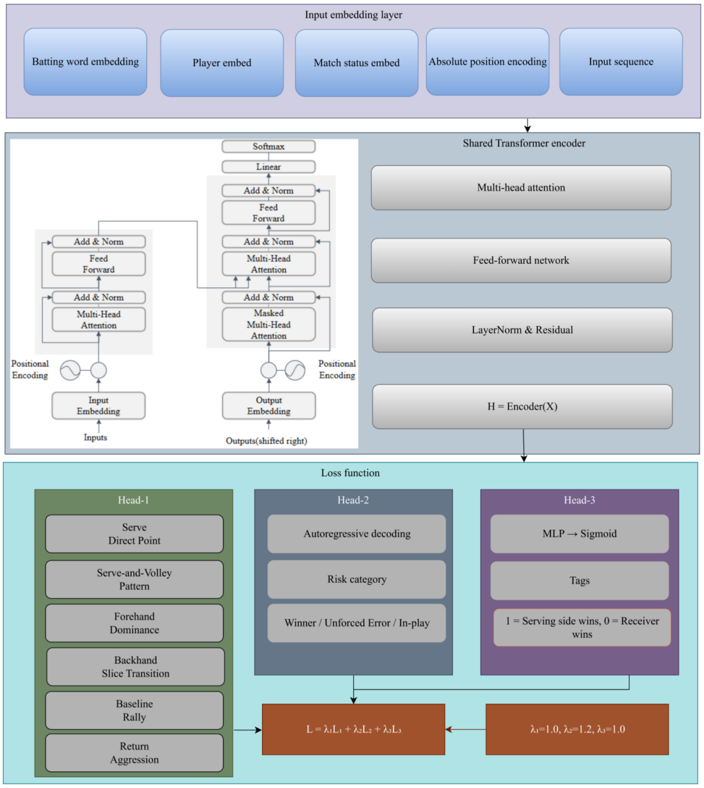

FIGURE 2. Multi-task Transformer model structure

$$z_1 = \mathrm{Dropout}(\mathrm{GELU}(h_0 W_{c1} + b_{c1})) \tag{8}$$

$$o_1 = z_1 W_{c2} + b_{c2} \tag{9}$$

In formulas (8)-(9), $W_{c1} \in \mathbb{R}^{256\times 64}$, $W_{c2} \in \mathbb{R}^{64\times 6}$.

Six types of technical and tactical behaviors are defined (manually annotated based on MCP event logic rules):

Serve Direct Point: A serve directly results in a winner;

Serve-and-Volley Pattern: A volley or high pressure is the second shot after the serve;

Forehand Dominance: A series of forehand diagonals followed by a sudden change to a straight line winner;

Backhand Slice Transition: Using a backhand slice to force a shallow return, followed by a net or attack to score;

Baseline Rally: A rally lasting 5 shots or more with no obvious offensive ending;

Return Aggression: A serve directly results in a winner or forces a forced error.

The output probability distribution is:

$$p_1 = \mathrm{softmax}(o_1) \in \mathbb{R}^6 \tag{10}$$

Head-2 (Hit Intention Prediction) is an autoregressive sequence generation task. The goal is to predict three key attributes of the t+1 shot given the previous t shots (t≥1): shot type (12 categories); shot direction (3 categories: Cross, DTL, Middle); and finish risk (3 categories: Winner, Unforced Error, In-play).

For each time step $t \in \{1, 2, \dots, L-1\}$, h_t is input and output through three independent MLP heads:

$$o_{\mathrm{type}}^{(t)} = \mathrm{MLP}_{\mathrm{type}}(h_t) \in \mathbb{R}^{12} \tag{11}$$

$$o_{\mathrm{dir}}^{(t)} = \mathrm{MLP}_{\mathrm{dir}}(h_t) \in \mathbb{R}^{3} \tag{12}$$

$$o_{\mathrm{risk}}^{(t)} = \mathrm{MLP}_{\mathrm{risk}}(h_t) \in \mathbb{R}^{3} \tag{13}$$

A teacher forcing strategy is used during training: the input at step t is always the real historical sequence, rather than a model-generated result, ensuring training stability and effective gradient propagation. This design enables the model to dynamically capture tactical evolution, such as the intention chain of "serve → receive shallow serve → high pressure at net."

Head-3 (Round Outcome Prediction) supports early prediction: at any K-th shot ($1 \le K \le L$), the final winner of the point is predicted based on the sequence up to shot K (server wins = 1, receiver wins = 0). Specifically, the input sequence is truncated to $X^{(K)} = [x_0, \dots, x_K]$, passed through an encoder to obtain $h_0^{(K)}$, and then passed through a single-layer MLP and Sigmoid to output the win rate:

$$\hat{y}_E^{(K)} = \sigma(h_0^{(K)} w_H + b_H) \in [0, 1] \tag{14}$$

E. CONTEXT FUSION MECHANISM BASED ON EMBEDDING AND SUMMATION STRATEGY

To incorporate non-sequential context into the model, it fuses four embeddings using element-by-element vector summation. The word embeddings (s, d, o, p) (type, direction, effect, depth) are one-hot encoded and then concatenated into $e_{\text{word}} \in \mathbb{R}^{12+3+3+3} = \mathbb{R}^{21}$, which is then linearly projected to $\mathbb{R}^{256}$.

The player embeddings for server ID i_s and receiver ID i_r are obtained by looking up the tables separately, $e_s, e_r \in \mathbb{R}^{64}$, which are then concatenated and projected onto $\mathbb{R}^{256}$, denoted as e_{player}. The match status embeddings (break point, game point, deuce) are one-hot encoded to $e_{\text{state}} \in \mathbb{R}^3$. The absolute position encoding [29,30] uses the original Transformer sinusoidal encoding:

$$\mathrm{PE}(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right) \tag{15}$$

$$\mathrm{PE}(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \tag{16}$$

In formulas (15)-(16), $t \in \{0, 1, \ldots, L\}$ is the position index and $i \in \{0, \cdots, d/2-1\}$.

The final input is:

$$x_t = W_Q e^{(t)}_{\text{word}} + W_R(e_s \oplus e_r) + W_K e_{\text{state}} + \mathrm{PE}(t) \tag{17}$$

In formula (17), $W_Q \in \mathbb{R}^{256\times21}$, $W_R \in \mathbb{R}^{256\times128}$, $W_K \in \mathbb{R}^{256\times3}$ are learnable projection matrices, and $\oplus$ represents concatenation. This summation mechanism ensures that all contextual information is aligned in the sequence dimension and processed uniformly by the Transformer.

F. MULTI-TASK LOSS FUNCTION

The total loss is the weighted sum of the three task losses:

$$\mathcal{L} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 + \lambda_3\mathcal{L}_3 \tag{18}$$

After ablation experiment and verification set performance optimization, the final weight is set as ($\lambda_1 = 1.0$, $\lambda_2 = 1.2$, $\lambda_3 = 1.0$). Because intention prediction tasks involve multi-attribute generation and have higher requirements for sequence modeling ability, and the gradient amplitude is relatively small, its weight is slightly increased to promote the learning balance between tasks and avoid the dominant task suppressing auxiliary signals.

Technical and tactical classification loss (cross entropy):

$$\mathcal{L}_1 = -\sum_{c=1}^{6} y_{1,c} \log \hat{y}_{1,c} \tag{19}$$

In formula (19), c is the specific technical and tactical category. y_1 is the one-hot label.

Intent prediction loss (multi-attribute cross entropy sum):

$$\begin{aligned}\mathcal{L}_2 = \frac{1}{N}\sum_{t=1}^{L-1} &\left[\mathrm{CE}(y^{(t)}_{\text{type}}, \hat{y}^{(t)}_{\text{type}}) + \mathrm{CE}(y^{(t)}_{\text{dir}}, \hat{y}^{(t)}_{\text{dir}})\right. \\ &\left. + \mathrm{CE}(y^{(t)}_{\text{risk}}, \hat{y}^{(t)}_{\text{risk}})\right]\end{aligned} \tag{20}$$

In formula (20), $N = L - 1$ is the number of valid prediction steps.

Round result loss (binary cross entropy):

$$\mathcal{L}_3 = -\left[y_3 \log \hat{y}^{(K)}_3 + (1 - y_3)\log(1 - \hat{y}^{(K)}_3)\right] \tag{21}$$

The model hyperparameters and training configuration are shown in Table 2.

TABLE 2. Model hyperparameters and training configuration

Component	Value	Instructions
Maximum length of input sequence L	30	Covering 98.7% of rounds
Embedding dimension	256	—
Encoder layers	6	—
Attention head count h	8	—
Dimension of hidden layers in feedforward networks	1024	—
Dropout	0.1	All sub layers
Activation function	GELU	—
Optimizer	AdamW	$\beta_1 = 0.9$, $\beta_2 = 0.999$
Learning rate	3×10^{-4}	Cosine decay after 1000 steps of linear preheating
Batch size	256 (by tennis round)	—
Training rounds	50	Early Stop Patience=8
Loss weight	$\lambda_1 = 1.0$, $\lambda_2 = 1.2$, $\lambda_3 = 1.0$	Intention prediction task has slightly higher weight

III. EXPERIMENTAL DESIGN

A. EXPERIMENTAL SETUP

1) Dataset and Preprocessing

To ensure the rigor and generalizability of the experiment, this study systematically preprocesses and partitions the original Match Charting Project dataset [31]. Complete rounds (rallies) are screened and the following invalid samples are removed:

(1) rounds with a rally_count of 0 or missing;

(2) abnormal records with a point_victor field other than 1 or 2;

(3) sequences containing undefined shot types (such as null values or "?").

This process ensured that all retained rounds had clear start (serve) and end (score) events, and that the attributes of each shot are complete and parseable.

A match-level partitioning strategy is used: the data is divided into a training set (70%), a validation set (15%), and a test set (15%) based on the unique match ID, totaling 5,050/1,082/1,082 matches. This strategy strictly ensured that the same player did not appear in multiple subsets, effectively preventing the model from overfitting by memorizing player IDs, thereby truly reflecting its generalization ability to unknown opponents. Dataset statistics are shown in Table 3.

TABLE 3. Dataset statistics

Statistical dimension	Training set	Validation set	Testing set	Total
Match sessions	5,050	1,082	1,082	7,214
Complete number of rounds	377,238	80,837	80,837	538,912
Average turn length (in beats)	6.42	6.38	6.41	6.4
Number of players (unique ID)	1,842	398	396	2,126

2) Evaluation Metrics and Baseline Models

To comprehensively evaluate the model's performance in understanding tennis technical and tactical behaviors, this study designed differentiated evaluation metrics for three key tasks and selected six representative sequence modeling methods as baselines to validate the superiority of our Transformer architecture.

Task 1 (Technical and Tactical Behavior Classification):

Due to the imbalance of the six categories of technical and tactical behaviors, Macro-F1 takes the arithmetic average of the F1 values of each category to avoid the majority class dominating the evaluation results:

$$\text{Macro-F1} = \frac{1}{C}\sum_{c=1}^{C}\frac{2\cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \quad C=6 \tag{22}$$

Task 2 (Hit Intention Prediction) evaluates the prediction performance of three sub-attributes: Type Accuracy (Top-1 accuracy across 12 technical labels); Direction Accuracy (Top-1 accuracy across three direction labels); and Risk F1 (Macro-F1) (Measures the overall performance of three risk categories: Winner, Unforced Error, and In-play).

Task 3 (Round Outcome Prediction): Predict the winner (server/receiver) at the K-th shot of the round. Accuracy@K and AUC (Area Under the Curve) are reported.

$$\text{Accuracy@K} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\hat{y}_i^{(K)} = y_i) \tag{23}$$

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x))\,dx \tag{24}$$

In formulas (23)-(24), I is the indicator function, TPR and FPR are the true positive rate and false positive rate, respectively. AUC measures the model's discriminative ability at different thresholds and is robust to class imbalance.

In the round result prediction task, the symbol '@k' indicates that the model predicts the final outcome (server wins or receiver wins) based on the current hit sequence when the round is up to the k-th hit. For example, AUC@5 Represents the area under the curve of the model to predict the outcome of the whole round after the fifth stroke. This indicator is used to evaluate the discrimination ability of the model in the early stage of the round.

To ensure fair comparison and academic rigor, all baselines use the same input representation and training/validation/testing split as the model in this paper. Specifically:

LSTM: 2-layer stacked LSTM, hidden dimension 256, used to model short-term temporal dependencies;

BiLSTM (Bidirectional Long Short-Term Memory) [32,33]: bidirectional LSTM, forward and backward hidden state splicing, enhanced context perception; GRU [34]: gated recurrent unit, simplified structure but retaining long-term memory capacity; TCN (Temporal Convolutional Network) [35]: 4-layer causal convolution, expansion factor exponential growth (1, 2, 4, 8), receptive field covering 30 beats; QRNN (Quasi-Recurrent Neural

Network): convolution + pooling simulated cycle, support parallel training, hidden dimension 256; CNN (Convolutional Neural Network)-LSTM [36,37]: 2-layer CNN (kernel=3) extracts local batting patterns, followed by 2-layer LSTM to model temporal sequences.

Considering the potential demand of tactical decision support for real-time, this paper measured the reasoning performance of each model on NVIDIA A10 GPU. Taking the typical round sequence with a length of 10 as the input (batch size=1), the average delay of single forward reasoning in this transformer model is 4.2 MS (± 0.3 ms), and the peak memory consumption is 186 MB. In contrast, the delay of bilstm is 2.1 MS (98 MB), the TCN is 1.9 MS (76 MB), and the cnn-lstm is 2.8 MS (112 MB). Although the computational overhead of transformer is slightly high, its delay is still far lower than the video frame interval or the tactical decision window, so it fully meets the application requirements of sideline coaches or post game analysts in real-time or near real-time scenarios. It is worth noting that the model has not been quantified or optimized by distillation; If it needs to be deployed on mobile or edge devices with limited resources, knowledge distillation can be used to compress the transformer into a lightweight LSTM teacher student architecture, which can retain more than 90% of the performance while reducing the delay to less than 1.5 Ms. The current implementation is more suitable for tactical analysis workstation or cloud reasoning service with GPU support.

IV. RESULTS AND ANALYSIS

A. OVERALL PERFORMANCE COMPARISON

To systematically verify the comprehensive advantages of this Transformer model in understanding technical and tactical behavior, it compares its performance on three tasks on the test set with that of six baselines. This comprehensively reflects the model's leading position in behavior recognition, intention prediction, and outcome assessment. The overall performance comparison is shown in Table 4.

TABLE 4. Overall performance comparison

Model	Techniques and Tactics (Macro-F1)	Type Accuracy	Direction Accuracy	Risk F1	AUC@5
LSTM	0.821	0.835	0.827	0.802	0.842
TCN	0.834	0.842	0.836	0.811	0.848
BiLSTM	0.841	0.848	0.84	0.818	0.853
QRNN	0.837	0.845	0.838	0.815	0.850
CNN-LSTM	0.846	0.852	0.845	0.823	0.857
GRU	0.828	0.839	0.831	0.807	0.845
Transformer	0.935	0.944	0.941	0.912	0.923

The Transformer model in this paper achieves a robust lead across all metrics. Its Macro-F1 for technical and tactical behavior classification reaches a high of 0.935, a 0.089 improvement over the best baseline (CNN- LSTM, 0.846). This demonstrates that the model can accurately identify fine-grained, complex behaviors such as "forehand dominance" and "backhand slice transition." In the intention prediction task, the accuracy of shot type and direction reaches 0.944 and 0.941, respectively, and the F1 for end risk is 0.912, far exceeding the baseline (which reached a maximum of 0.823). This demonstrates that the model is highly reliable in predicting the technical choices and tactical risks of the next shot. In terms of early prediction of the outcome of a round, the AUC@5 reaches 0.923, surpassing the BiLSTM (0.853), demonstrating that it can confidently predict the direction of the round within five shots. The performance advantage comes from the fact that the Transformer's global self-attention mechanism can effectively model non-adjacent but tactically closely related hitting events, while RNN/CNN models are limited by local receptive fields or sequential dependency assumptions and find it difficult to capture such long-range, nonlinear tactical coupling patterns. In addition to the significant improvement of performance indicators, the advantage of transformer model fundamentally stems from its strong ability to capture the long-range dependence in tennis hitting sequence through its self attention mechanism. Unlike the rnn/cnn model, which is limited by local receptive fields or order dependent assumptions, the self attention mechanism allows the model to directly focus on and integrate the information of any historical shot in the sequence, regardless of its distance, when calculating the representation of the current shot. In tennis tactics, the final result of a round is often determined by multiple non adjacent strokes (for example, the opening serve creates an opportunity for the winning point of the fifth racket). This "technology tactics result" causal chain usually spans multiple shots, and the self attention mechanism can dynamically learn the tactical correlation weights between these non adjacent shots, so as to more completely model the evolution process of the whole tactical layout. Therefore, the mechanism is assumed to be able to more effectively identify and utilize these long-range dependencies that are essential for understanding tactical logic.

B. TECHNICAL AND TACTICAL PATTERN RECOGNITION ANALYSIS

Boxplots are used to display the distribution of bout lengths for different skill and tactic categories. This aims to verify the rationality of the category definitions and reveal the temporal dynamics of various tactics, providing a statistical basis for behavior recognition. The average bout lengths for different skill and tactic categories are shown in Figure 3.

Figure 3 reveals significant differences in round lengths across various skill and tactical behaviors, strongly validating the rationality and discriminability of our proposed skill and tactical category definitions. The median length of Serve Direct Point is only 1 beat, with a very narrow interquartile range, consistent with its "one-beat" nature. Meanwhile, the median length of Baseline Rally is a whopping 9 beats, longer than other categories, fully reflecting its seesaw nature. This distribution not only provides a learnable temporal prior for the model but also demonstrates the feasibility of using round length to aid skill and tactical classification, supporting the effectiveness of our proposed multi-dimensional feature fusion modeling.

A comparison of BiLSTM, QRNN, CNN-LSTM, and Transformer models is shown in Figure 4, along with a heatmap of the confusion matrix for skill and tactical behavior classification.

The confusion matrix for tactical behavior classification shown in Figure 4 shows significant confusion between categories among BiLSTM, QRNN, and CNN-LSTM. For example, the BiLSTM (Figure 4a) misclassified 1,250 instances of "Serve Direct Point" as "Serve-and-Volley Pattern." Furthermore, 61 instances of "Baseline Rally" are misclassified as "Return Aggression," demonstrating the model's difficulty distinguishing between passive stalemate and active attack during long rounds. The QRNN (Figure 4b) exhibited confusion between "Backhand Slice Transition" and "Baseline Rally" (612 misclassifications), indicating that its convolutional structure is insufficiently sensitive to changes in tactical rhythm. The CNN-LSTM (Figure 4c) improved performance somewhat through local feature extraction, but still misclassified 241 instances of "Forehand Dominance" as "Return Aggression," indicating its inability to effectively model the continuity of tactical intent across shots. These confusions essentially stem from the local receptive field limitations and sequential dependency assumptions of the RNN/CNN architecture, which make it difficult to capture the tactical coupling between non-adjacent shots, resulting in blurred boundary judgments on complex technical and tactical behaviors.

In contrast, the Transformer model (Figure 4d) demonstrates superior category discrimination capabilities, significantly reducing confusion. It misclassifies "Serve Direct Point" only 1,232 times, correctly identifies "Baseline Rally" 5,359 times, and misclassifies "Return Aggression" only 54 times, demonstrating that the model can accurately distinguish between passive rallies and active finishes. Particularly notable, confusion between "Backhand Slice Transition"

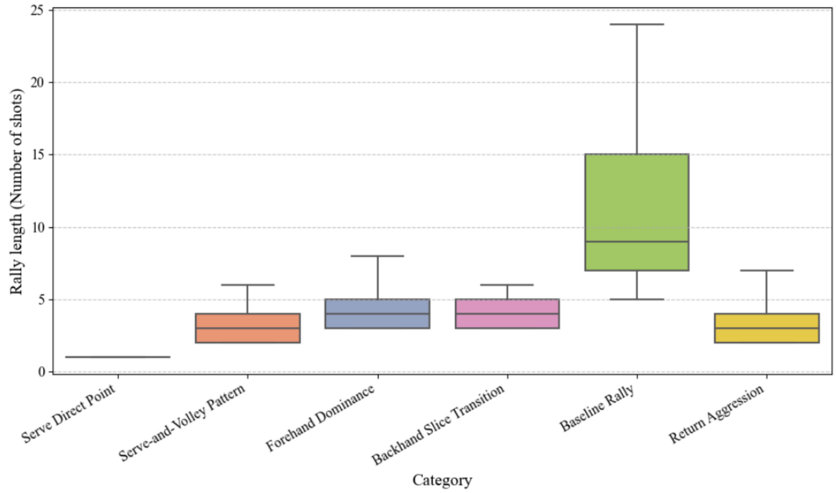

FIGURE 3. Boxplot of average round lengths for different skill and tactical categories

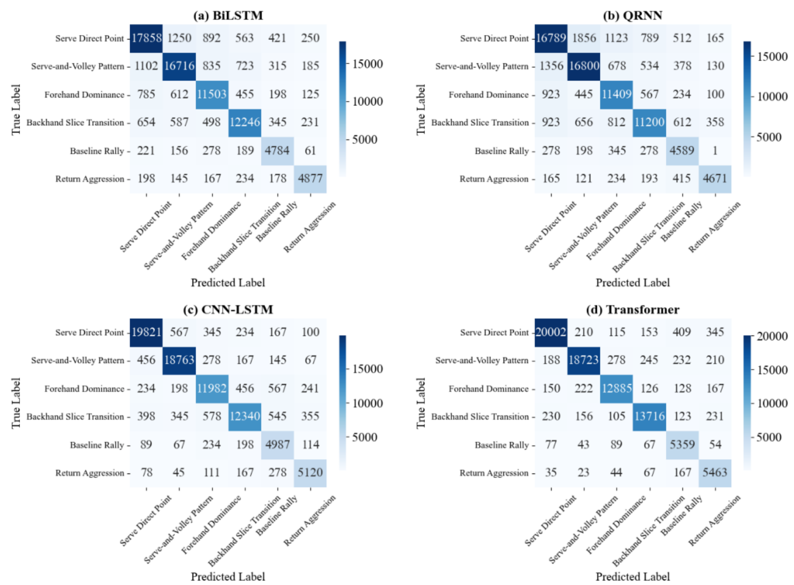

FIGURE 4. Confusion Matrix Heatmap for Technical and Tactical Behavior Classification. (a) BiLSTM Heatmap; (b) QRNN Heatmap; (c) CNN-LSTM Heatmap; (d) Transformer Heatmap

and "Baseline Rally" is reduced to 123 times, validating the self-attention mechanism's ability to sensitively capture turning points in tactical intent. This performance advantage stems from the Transformer's ability to model global dependencies: the attention weights between any two shots can dynamically learn tactical associations, thereby constructing more robust technical and tactical representations. Furthermore, multi-task learning further enforces behavioral semantic consistency, giving the model a stronger discrimination boundary across fine-grained categories, fully demonstrating its superiority in recognizing complex tennis technical and tactical behaviors.

C. DYNAMIC CHARACTERISTICS OF HITTING INTENTION PREDICTION

The goal of revealing the dynamic evolution of batting intention prediction performance over the course of a round is to identify key junctures in tactical decision-making and verify the model's predictive sensitivity to batting behavior at different stages. The results are shown in Figure 5.

The accuracy of shot type and direction predictions jumped significantly in the third and fifth shots, reaching 0.894/0.872 and 0.918/0.896, respectively, before gradually declining. This phenomenon has a clear tactical explanation: the third shot is the server's first active attack after serving (e.g., serve $\rightarrow$ receive $\rightarrow$ forehand pressure on the third shot), or the receiver's critical moment for attacking, where tactical intent is highly clear. The fifth shot corresponds to

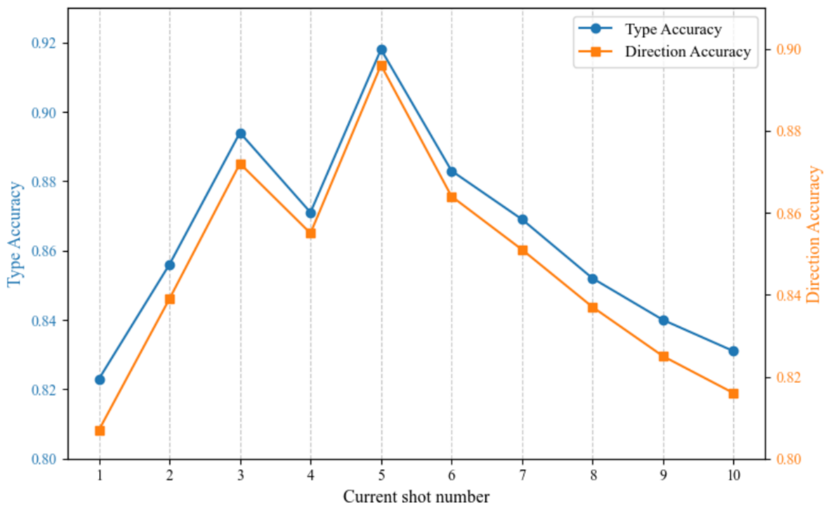

FIGURE 5. Changes as the round progresses

the "turning shot" in a multi-shot rally, such as by repeatedly landing shots to induce the opponent to return shallowly and suddenly in a straight line. At this time, shot selection is more constrained by the previous layout and the pattern is more predictable. However, the first shot is more difficult to predict because it relies solely on serve information. After the fifth shot, the round enters a highly competitive, random stalemate phase, where intent uncertainty increases, leading to a decline in accuracy. The Transformer, with its global attention mechanism, effectively captures these stage-by-stage tactical patterns, achieving high-precision predictions at key decision points, providing reliable support for re-altime tactical assistance. Figure 6 shows the distribution of attention weights when the model predicts the next shot in five typical rounds, revealing how the Transformer dynamically focuses on historical key shots based on tactical logic, thereby verifying its ability to model the dependencies between tennis techniques and tactics.

In Rounds 1 and 4, the model assigned high attention weights of 0.78–0.82 to "Serve-Slice-Wide," indicating that serve-receive decisions are highly dependent on the spin and placement of the serve, consistent with the tactical common sense that "receive reactions are dominated by serve quality." In Round 2, attention is focused on both "Serve-Slice-Wide" (0.35) and "Forehand-Cross" (0.51), indicating that the choice of the third shot depends not only on the player's serve intention but also significantly on the quality of the opponent's return, highlighting the interactive nature of two-player confrontation. In Round 3, the model assigned the highest attention weight (0.52) to "Backhand-Down-the-line," indicating that in mid-game rallies, the direction and depth of the last one or two shots are the core factors in determining the next shot's strategy. Most notably, in Round 5, the model focused 0.53% of its attention on "Forehand-Down-the-line-Winner" when predicting the winning point. This is the key shot in the strategic setup—repeated placement of the ball induces the opponent to return shallowly, thus creating a finishing opportunity. Overall, the attention distribution is not uniform or random, but precisely aligned with the causal chain and decision nodes in tennis tactics, fully demonstrating that our model can not only predict behavior with high accuracy but also has explainable tactical reasoning capabilities.

D. EARLY PREDICTION OF ROUND RESULTS AND GAME SITUATION ANALYSIS

The model's ability to discriminate the final scorer in the early stages of a round (3, 5, and 7 shots) is evaluated using the Winner Prediction Accuracy @ K curve, verifying its tactical situational awareness and forwardlooking analysis. Figure 7 shows the early predictions for round outcomes.

The Transformer model in this paper outperforms the baseline at all prediction time points. In particular, its accuracy reaches 0.864 when only three shots are played, an improvement of 0.057 over the best baseline (CNN-LSTM, 0.807). This demonstrates that the model can extract key tactical signals (such as serve quality, return intention, and third-shot aggression) from very early shot sequences. By the fifth shot, the accuracy jumps to 0.923, indicating that the outcome of most rounds is already determined within the first five shots. The Transformer effectively integrates the server's style, score pressure, and shot patterns to achieve highconfidence predictions. In contrast, RNN/CNN models are limited in their local modeling capabilities and struggle to capture long-range tactical dependencies (the three-shot chain of "serve $\rightarrow$ shallow return $\rightarrow$ net") when early information is limited, resulting in significantly lagging performance. This result not only validates the Transformer's superiority in tennis tactical analysis but also provides a reliable technical foundation for real-time decision-making support.

Figure 8 shows a comparison of ROC (Receiver Operating Characteristic)-AUC under different match scenarios.

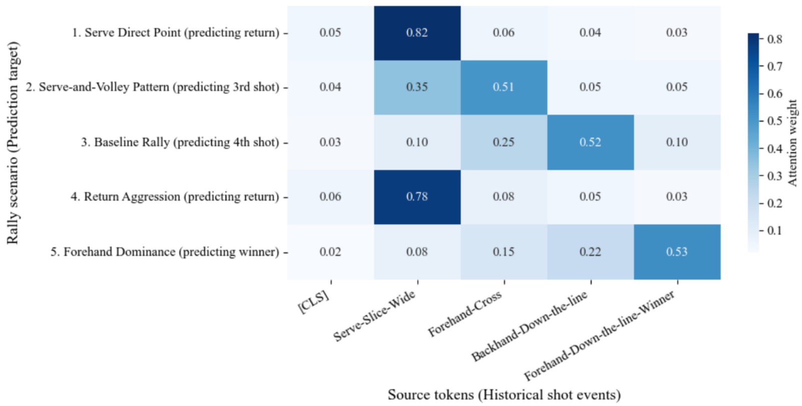

FIGURE 6. Attention weight visualization

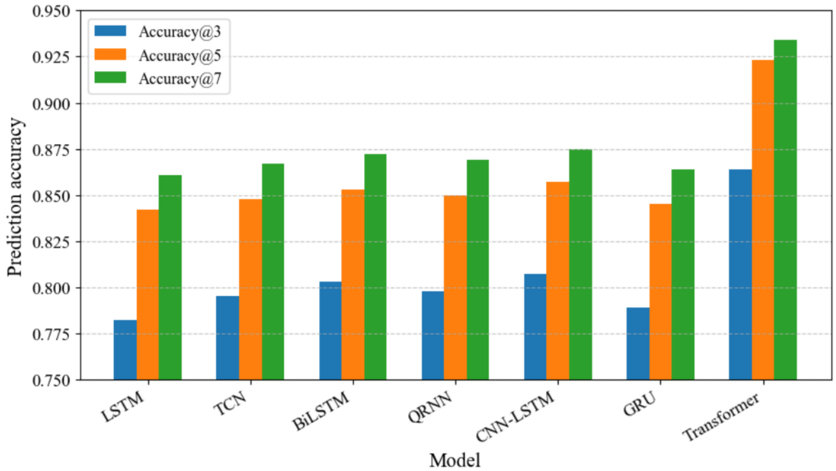

FIGURE 7. Early prediction of round results

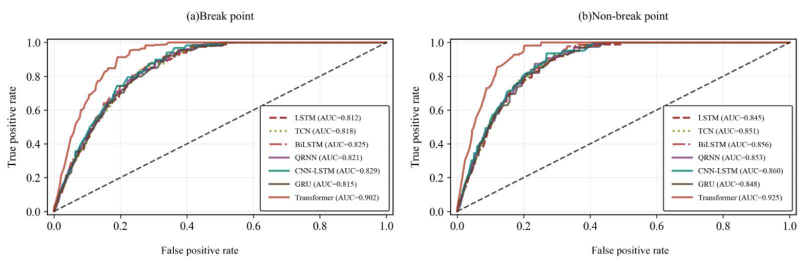

FIGURE 8. ROC-AUC comparison in different match scenarios. (a) Break point scenario; (b) Non-break point scenario

All models performed slightly better in non-break point scenarios due to the more stable selection of conventional tactical choices. However, in the high-pressure break point scenario, decision uncertainty increases significantly, resulting in a generally lower AUC. However, our Transformer model significantly outperformed the baseline in both scenarios, particularly in the break point scenario, achieving an AUC of 0.902, a 0.073 improvement over the best baseline (CNN-LSTM, 0.829). This demonstrates that the Transformer model not only effectively models conventional tactical patterns but also, by integrating score state embeddings with player style representations, accurately captures subtle behavioral deviations in high-pressure matches, demonstrating superior contextual awareness and robust predictive capabilities.

E. ABLATION EXPERIMENT

To verify the contribution of each model component to technical and tactical understanding, Table 5 quantifies the impact of word embedding, player embedding, game status embedding, and multi-task learning on the overall performance through ablation experiments.

TABLE 5. Ablation experiments

Configuration	Techniques and Tactics (Macro-F1)	Risk F1	AUC@5
w/o word embeddings	0.762	0.731	0.784
w/o player embeddings	0.891	0.867	0.882
w/o match state embeddings	0.873	0.845	0.861
w/o multi-task learning	0.884	0.859	0.875
Full model	0.935	0.912	0.923

Note: w/o (without).

Ablation experiments show that each component contributes significantly to model performance. Term embeddings (type, direction, effect, depth) are the cornerstone of technical and tactical modeling. Their absence causes Macro-F1 to plummet to 0.762, demonstrating the indispensability of fine-grained behavioral representation. Removing player embeddings reduces performance by 0.044, validating that the model effectively captures individual style differences by learning player IDs. Missing match state embeddings reduces AUC@5 by 0.062, highlighting the critical impact of high-pressure situations on shot decisions. Removing multi-task learning also results in a significant performance loss, demonstrating the strong semantic connection between technical and tactical recognition, intent prediction, and outcome assessment, and that joint training promotes representation generalization. By integrating multidimensional context and collaboratively optimizing multitask objectives, the complete model achieves the most comprehensive and accurate modeling of tennis technical and tactical behavior.

In order to further explore the role of multi head attention mechanism in the model architecture, we added an ablation experiment to compare the model performance under the configuration of standard 8 head attention (h=8) and single head attention (h=1). The results are shown in Table 6. Experiments show that reducing the number of attention heads from 8 to 1 will lead to the performance degradation of all evaluation indicators, especially in the technical and tactical classification (Macro-F1 decreased by 0.021) and the risk prediction of batting intention (risk F1 decreased by 0.018). This verifies the necessity of multi head design, which enables the model to capture the diversified tactical relationships in the hitting sequence from different representation subspaces in parallel (such as player style, hitting type interaction, spatial position, etc.), while the expression ability of single head attention mechanism is limited, so it is difficult to fully model this complex dependency.

TABLE 6. Ablation study on the number of attention heads

Configuration	Techniques and Tactics (Macro-F1)	Type Accuracy	Direction Accuracy	Risk F1	AUC@5
Single-Head (h=1)	0.914	0.928	0.925	0.894	0.911
Multi-Head (h=8)	0.935	0.944	0.941	0.912	0.923

V. CONCLUSION

This study proposes a Transformer-based multi-task sequence modeling method specifically for the recognition and prediction of technical and tactical behaviors of tennis players, where the need to process complex event sequences for predictive analysis mirrors the high-precision sequence modeling required for automated defect detection and continuous process control in modern manufacturing. First, by constructing a structured sequence of hitting events and incorporating multidimensional context such as player identity and match status, a fine-grained representation of the coupled technical and tactical patterns is achieved. Second, a shared encoder and task-specific heads are designed to jointly optimize the three tasks of technical and tactical classification, hitting intention prediction, and round outcome assessment, effectively avoiding information loss caused by task fragmentation. Finally, a self-attention mechanism and an embedding summation strategy are applied to enable the model to model global dependencies and contextual awareness. The results demonstrate that the proposed model achieves performance advantages in technical and tactical behavior recognition, dynamic intention prediction, and early round outcome assessment, validating the powerful expressive power of the Transformer in modeling discrete event sequences. This paper achieves accurate recognition of complex tactical patterns without visual constraints and provides interpretable attention weights to reveal the logic of tactical decision-making, offering a scalable, data-driven methodology that parallels the need for interpretable predictive modeling and autonomous quality control decision-making support in complex manufacturing processes.

AUTHOR CONTRIBUTIONS

Conceptualization – Juan Pu and Xianyi Zha; methodology – Juan Pu and Xianyi Zha; investigation – Juan Pu and Xianyi Zha; writing-original draft preparation – Juan Pu and Xianyi Zha. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] I. Dogan, S. Revan, and S. Arıkan, "Analysis of tennis competitions on different court surfaces," Turkish Journal of Sport and Exercise, vol. 23, no. 1, pp. 60-66, 2021, doi: 10.15314/tsed.895726.

[2] E. Baiget, F. Corbi, and J. Lopez, "Influence of anthropometric, ball impact and landing location parameters on serve velocity in elite tennis competition," Biology of Sport, vol. 40, no. 1, pp. 273-281, 2023, doi: 10.5114/biolsport.2023.112095.

[3] L. Wang, "Construction of Evaluation Model of Tennis Skills and Tactic Level and Application of Grey Relational Algorithm," Journal of Sensors, vol. 2022, no. 2, pp. 1-11, 2022, doi: 10.1155/2022/9446175.

[4] E. Sanchez Mencia, J. Campos-Rius, X. González Santamaria, and E. Borrajo Mena, "Tactical skills in tennis: A systematic review," International Journal of Sports Science & Coaching, vol. 19, no. 2, pp. 894-907, 2024, doi: 10.1177/17479541231216268.

[5] N. Eugster and C. Ewerhart, "Rules in International Tennis Tournaments: A Game Theoretic Perspective," Journal of Sports and Games, vol. 6, no. 1, pp. 33-42, 2024, doi: 10.22259/2642-8466.0601004.

[6] S. Liu, "Tennis players' hitting action recognition method based on multimodal data," International Journal of Biometrics, vol. 16, no. 3-4, pp. 317-336, 2024, doi: 10.1504/IJBM.2024.138223.

[7] V. Candila and L. Palazzo, "Neural networks and betting strategies for tennis," Risks, vol. 8, no. 3, pp. 68, 2020, doi: 10.3390/risks8030068.

[8] T. C. Y. Chan, D. S. Fearing, C. Fernandes, and S. Kovalchik, "A Markov process approach to untangling intention versus execution in tennis," Journal of Quantitative Analysis in Sports, vol. 18, no. 2, pp. 127-145, 2022, doi: 10.48550/arXiv.2110.01527.

[9] S. Wu, M. Diao, J. Wang, Z. Song, and C. Zhang, "A Study on the Optimisation of Tennis Players' Match Strategies from the Perspective of Momentum," Applied Sciences, vol. 15, no. 10, pp. 5624, 2025, doi: 10.3390/app15105624.

[10] J. Li, "Machine learning-based analysis of defensive strategies in basketball using player movement data," Scientific Reports; vol, 15, no. 1, Art. no. 13887, 2025, doi: 10.1038/s41598-025-98877-1.

[11] C. Ma, J. Fan, J. Yao, and T. Zhang, "NPU RGBD dataset and a feature-enhanced LSTM-DGCN method for action recognition of basketball players," Applied Sciences, vol. 11, no. 10, pp. 4426, 2021, doi: 10.3390/app11104426.

[12] A. Anderson, J. Rosen, J. Rust, and K. Wong, "Disequilibrium play in tennis," Journal of Political Economy, vol. 133, no. 1, pp. 190-251, 2025, doi: 10.1086/732529.

[13] A. Ganser, B. Hollaus, and S. Stabinger, "Classification of tennis shots with a neural network approach," Sensors, vol. 21, no. 17, pp. 5703, 2021, doi: 10.3390/s21175703.

[14] T. Zou, J. Wei, B. Yu, X. Qiu, H. Zhang, X. Du, et al., "Fast moving table tennis ball tracking algorithm based on graph neural network," Scientific reports, vol. 14, no. 1, Art. no. 29320, 2024, doi: 10.1038/s41598-024-80056-3.

[15] H. Li, S. G. Ali, J. Zhang, B. Sheng, P. Li, Y. Jung, et al., "Video-based table tennis tracking and trajectory prediction using convolutional neural networks," Fractals, vol. 30, no. 05, Art. no. 2240156, 2022, doi: 10.1142/S0218348X22401569.

[16] B. Lim, S. O. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International journal of forecasting, vol. 37, no. 4, pp. 1748-1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[17] L. Huang, F. Mao, K. Zhang, and Z. Li, "Spatial-temporal convolutional transformer network for multivariate time series forecasting," Sensors, vol. 22, no. 3, pp. 841, 2022, doi: 10.3390/s22030841.

[18] L. Zhang, "Sports action recognition algorithm based on multi-modal data recognition," Intelligent Decision Technologies, vol. 18, no. 4, pp. 3243-3257, 2024, doi: 10.3233/IDT-230372.

[19] Y. Gao, J. Zou, and Y. Zhang, "Exploring the potential of deep learning techniques for analyzing athlete movements in competitive athletics sports," Machine Vision and Applications, vol. 36, no. 6, pp. 1-20, 2025, doi: 10.1007/s00138-025-01733-5.

[20] C. N. Chen and W. T. Chu, "How it flies and why it flies? Volleyball trajectory segmentation and classification," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 68, no. 5, pp. 1591-1595, 2021, doi: 10.1109/tcsii.2021.3067027.

[21] X. Hu, W. Zhang, H. Ou, S. Mo, F. Liang, J. Liu, et al., "Enhancing squat movement classification performance with a gated long-short term memory with transformer network model," Sports Biomechanics, vol. 24, no. 6, pp. 1629-1644, 2025, doi: 10.1080/14763141.2024.2315243.

[22] J. Deng, S. Zhang, and J. Ma, "Self-attention-based deep Convolution LSTM framework for sensor-based badminton activity recognition," Sensors, vol. 23, no. 20, pp. 8373, 2023, doi: 10.3390/s23208373.

[23] A. Fayomi, R. Majeed, A. Algarni, S. Akhtar, F. Jamal, and J. A. Nasir, "Forecasting tennis match results using the Bradley - Terry model," International Journal of Photoenergy, vol. 2022, no. 1, Art. no. 1898132, 2022, doi: 10.1155/2022/1898132.

[24] N. Buhamra, A. Groll, and S. Brunner, "Modeling and prediction of tennis matches at Grand Slam tournaments," Journal of Sports Analytics, vol. 10, no. 1, pp. 17-33, 2024, doi: 10.3233/JSA-240670.

[25] H. Cheng, J. Wang, A. Zhao, Y. Zhong, J. Li, and L. Dong, "Joint graph convolution networks and transformer for human pose estimation in sports technique analysis," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 10, Art. no. 101819, 2023, doi: 10.1016/j.jksuci.2023.101819.

[26] Y. Chen, "Hybrid Transformer-LSTM Model for Athlete Performance Prediction in Sports Training Management," Informatica, vol. 49, no. 24, pp. 185-196, 2025, doi: 10.31449/inf.v49i24.8013.

[27] A. Muniasamy, "Revolutionizing health monitoring: Integrating transformer models with multi-head attention for precise human activity recognition using wearable devices," Technology and Health Care, vol. 33, no. 1, pp. 395-409, 2025, doi: 10.3233/THC-241064.

[28] L. Chang, S. Rani, and A. Akbar M, "ChampionNet: a transformer-enhanced neural architecture search framework for athletic performance prediction and training optimization," Discover Computing, vol. 28, no. 1, pp. 71, 2025, doi: 10.1007/s10791-025-09560-y.

[29] X. Chen and Z. Liu, "Hierarchical query design and distributed attention in transformer for player group activity recognition in sports analysis: X," Scientific Reports, vol. 15, no. 1, Art. no. 31571, 2025, doi: 10.1038/s41598-025-16752-5.

[30] X. Wang, Z. Tang, J. Shao, S. Robertson, A. Gomez M, and S. Zhang, "Hooptransformer: advancing NBA offensive play recognition with self-supervised learning from player trajectories," Sports Medicine, vol. 54, no. 10, pp. 2663-2673, 2024, doi: 10.1007/s40279-024-02030-3.

[31] F. Lisi, M. Grigoletto, and M. G. Briglia, "On the distribution of rally length in professional tennis matches," Journal of Sports Analytics, vol. 10, no. 1, pp. 105-121, 2024, doi: 10.3233/JSA-240728.

[32] Z. Quan, J. Li, and W. Song, "An Improved LSTM Based Early Warning Model for Physical Education Network Teaching Achievements," Journal of Information Processing Systems, vol. 20, no. 6, pp. 793-800, 2024, doi: 10.3745/JIPS.04.0328.

[33] K. Wang, D. Zhu, Z. Chang, and Z. Wu, "Research on prediction and evaluation algorithm of sports athletes performance based on neural network," Technology and Health Care, vol. 32, no. 6, pp. 4869-4882, 2024, doi: 10.3233/THC-232000.

[34] B. Toussaint and M. Raison, "Real-time trajectory prediction of a ping-pong ball using a GRU-TAE," Applied Intelligence, vol. 55, no. 5, pp. 339, 2025, doi: 10.1007/s10489-024-06204-4.

[35] V. Bijalwan, A. M. Khan, H. Baek, S. Jeon, and Y. Kim, "Interpretable human activity recognition with temporal convolutional networks and model-agnostic explanations," IEEE Sensors Journal, vol. 24, no. 17, pp. 27607-27617, 2024, doi: 10.1109/JSEN.2024.3418496.

[36] A. Rostamian and J. G. O'Hara, "Event prediction within directional change framework using a CNN-LSTM model," Neural Computing and Applications, vol. 34, no. 20, pp. 17193-17205, 2022, doi: 10.1007/s00521-022-07687-3.

[37] A. S. Lombu, I. V. Paputungan, and C. K. Dewa, "Predicting fantasy premier league points using convolutional neural network and long short term memory," Jurnal Teknik Informatika (JUTIF), vol. 5, no. 1, pp. 263-272, 2024, doi: 10.52436/1.jutif.2024.5.1.1792.