

Analyzing the Relationship Between Football Players' Passing Network Structure and Tactical Performance Using Graph Convolutional Networks

Yongkang Guan, Weijia Xue

School of Physical Education and Health, Guilin Institute of Information Technology (GIIT), Guilin 541004, Guangxi Zhuang Autonomous Region, China

Corresponding author: Weijia Xue (e-mail: 18677305290@163.com)

ABSTRACT Current football passing-network analysis methods lack dynamic and interpretable modeling of the relationship between network structure and tactical performance. Based on the GCN framework, this paper constructs a T-GAT model by integrating a dynamic attention mechanism and GRU temporal modeling. The match is represented as a directed attributed passing graph over time windows. An edge-aware multi-head GAT encoder generates window-level graph representations, which are then fed into a GRU with time decay and retrospective attention to form hybrid temporal representations under unequal time intervals. The model jointly predicts window-level ΔxG , ΔxT , Progressive Passes, and Pass Network Penetration Rate, while extracting edge attention for visualization and interpretation of key passing links. Experimental results show that T-GAT achieves ΔxG RMSE values of only 0.06 and 0.09 in the 0–15 and 76–90 minute test periods, respectively. Structural correlation analysis indicates that passing-network density is strongly correlated with ΔxG (Pearson $r = 0.84$, $p < 0.001$), while average node output strength is correlated with ΔxT ($r = 0.92$, $p < 0.001$). The findings show that high connectivity and decentralized control improve tactical performance and provide reference for networked propagation and signal-flow analysis.

INDEX TERMS passing network, T-GAT, temporal graph attention, tactical performance, propagation analysis

I. INTRODUCTION

A football match is a highly coordinated and time-varying system, where the strengths and weaknesses of a team's tactics are often reflected in the coordination between players and the structure of the passing network [1,2]. In recent years, big data and location tracking technologies have made it possible to measure individual passes, player positions, and kinematic information, providing a data foundation for quantifying football tactics from a network perspective. Correspondingly, in the industrial sector, the rise of IoT sensors and advanced monitoring systems has enabled the measurement of individual machine states, material flow positions, and operational kinetics, providing an analogous data foundation for quantifying production efficiency and quality control strategies within the industry [3-5]. Currently, most tactical analysis methods rely on static statistical indicators, such as number of passes, completion rate, and degree centrality, or on tactical evaluations based on manual experience. These methods struggle to simultaneously capture the structure of the passing network, the temporal evolution of the pass chain, and the dynamic impact of different pass types and on-field positions on

tactical effectiveness.

A major challenge is to build a model that can both represent complex graph structures and reflect temporal dynamics, thereby enabling the analysis of the relationship between football players' passing networks and tactical performance and prediction of team tactical performance.

In studying the relationship between passing networks and team tactics and match performance, scholars have approached this from multiple perspectives. Early and recent social network and statistical methods have often used metrics such as centrality, clustering coefficient, and network density to characterize team passing structures and assess coordination and competitiveness. Zhou W et al. proposed a weighted passing channel based on the on-field coordinates and used the average clustering coefficient and the ratio of betweenness centrality to verify the effectiveness of the team coordination index [6]. Ievoli R et al. combined the passing network with the Bayesian hierarchical model to prove that the passing characteristics have significant explanatory power for the game results [7]. Leela J et al. used the passing matrix and centrality to characterize the distribution and efficiency of passes under formations,

and used the traveling salesman problem to find the least connected structure, providing a reference for tactics [8]. Overall, passing network indicators have been proven to be effective in explaining team behavior and predicting game results, and have developed frameworks such as spatial weighting, hierarchical modeling, time series segmentation, and topological comparison. Existing research still relies on static or coarse-grained segmented networks, making it difficult to continuously characterize tactical evolution. Furthermore, research remains limited in modeling edge attributes and high-order interactions, introducing temporal attention mechanisms for graph structures, and jointly predicting and explaining structural representations and multi-dimensional performance. This paper aims to systematically quantify and explain how passing network structure influences tactical performance in football matches, providing actionable metrics and interpretable models for coaches and analysts. This paper first constructs a game as a directed, attributed passing graph based on time windows. In the graph encoder, an edge-aware GAT is used to calculate the dynamic importance between nodes. The window-level graph representation is then serialized and fed into a Gated Recurrent Unit (GRU) with time decay for temporal modeling. Finally, a shared projection and four independent regression heads are used to perform multi-task regression on ΔxG , ΔxT , PP, and PNR. Results show that proposed T-GAT (Temporal Graph Attention) model achieved a Root Mean Square Error (RMSE) of 0.07 for ΔxG , 0.14 for ΔxT , 0.28 for PP, and 0.06 for PNR on the test set, significantly outperforming all baselines. Relationship analysis revealed a Pearson correlation coefficient of 0.84 between the passing network density and ΔxG , and a correlation of -0.96 between the average shortest path length and ΔxT . Partial correlation tests showed a partial correlation of 0.94 between network density and ΔxG , and a partial correlation of 0.96 between average output strength and PNR (all $p < 0.001$). Attention visualization further identifies key passing links across multiple formations (with some edge weights reaching as high as 1.00), and demonstrates the model's robustness and deployment feasibility over time.

II. PASSING NETWORK DYNAMIC MODELING ARCHITECTURE

A. CONSTRUCTION OF A FOOTBALL PLAYER PASSING NETWORK STRUCTURE DIAGRAM

Let the game timeline be a continuous interval, divided into t time windows based on fixed duration or event-driven behavior. The set of passing events within window t is denoted by $P_t = \{p_k \mid \tau_{p_k} \in [T_t^{\text{start}}, T_t^{\text{end}}]\}$, p_k represents a single passing event, and τ_{p_k} represents the absolute time at which the event occurred. Let V be the unique identifier set of all football players on the field. A mask is used to identify whether a football player is effectively on the field within a window. If the football player effectively touches the ball in $[T_t^{\text{start}}, T_t^{\text{end}}]$ or the playing time is greater than 1 second, then the value is 1; otherwise, the value is 0. In each window,

a set of directed edges is defined, and the edge weight of each directed edge is defined as the weighted accumulation of passes within the window, as shown in formula (1).

$$E_t = \{(i \rightarrow j) \mid \exists p \in P_t \text{ s.t. } \text{sender}(p) = i, \text{receiver}(p) = j\},$$

$$w_{ij}^t = \sum_{p \in P_t(i \rightarrow j)} \gamma(\tau_{p,t}) \cdot s_p \cdot \delta(d_p) \quad (1)$$

In formula (1), s_p indicates whether the pass was successful, ranging from 0 to 1; d_p represents the Euclidean distance of the pass; $\delta(d_p)$ represents the distance decay function, with the constant d_0 set to 15 m to reduce the weight of extremely long passes; and $\gamma(\tau_{p,t})$ represents the time decay factor. For each edge, a fixed-dimensional edge feature vector is constructed, as shown in formula (2) [9,10].

$$e_{ij}^t = [\tilde{c}_{ij}^t, \tilde{s}_{ij}^t, \tilde{d}_{ij}^t, f_{zd,ij}^t, \tilde{\theta}_{ij}^t, q_{ij}^t] \quad (2)$$

In formula (2), $\tilde{c}_{ij}^t$ represents the normalized number of passes, $\tilde{s}_{ij}^t$ represents the pass success rate, and $\tilde{d}_{ij}^t$ represents the average Euclidean distance. $f_{zd,ij}^t$ represents the variance, $\tilde{\theta}_{ij}^t$ represents the angle mean, and q_{ij}^t represents the pass type distribution vector (short pass/long pass/through pass/overhead pass), with each component representing the frequency of occurrence of the type.

For each node (football player), the node vector is calculated in the window, as shown in formula (3) [11,12].

$$x_i^t = [\tilde{p}_i^t, \tilde{v}_i^t, r_i, \tilde{tr}_i^t, \tilde{\beta}_i^t] \quad (3)$$

In formula (3), $\tilde{p}_i^t$ represents the normalized mean coordinate, $\tilde{v}_i^t$ represents the average velocity estimate, r_i represents the one-hot encoding of the position role, $\tilde{tr}_i^t$ represents the touch rate, and $\tilde{\beta}_i^t$ represents the duration of play. The edge weights are now incorporated into the directed weighted adjacency matrix to construct a symmetric normalized matrix. For highly sparse or noisy edges, thresholding is first applied, and the edge weights are then smoothed using an exponential moving average to enhance temporal stability. The mean and standard deviation of all node and edge feature dimensions were calculated on the training set and z-score normalized to obtain $\tilde{x}$.

The constructed passing network structure is shown in Figure 1.

Figure 1 shows the passing network structure of 11 football players. Node size reflects the number of touches, demonstrating the involvement of different players in play-making. Edge width and color correspond to pass weight and attention weight, emphasizing the importance of key passing channels and tactics. The arrows reveal the direction of the passing flow, resulting in an orderly distribution of the overall network extending from the core midfield to the forwards. As shown in Figure 1, the midfield (CM, LM, RM) nodes are relatively large and have multiple thick lines connecting them to the forwards and full-backs, indicating their core role in passing, controlling, and dispatching. The

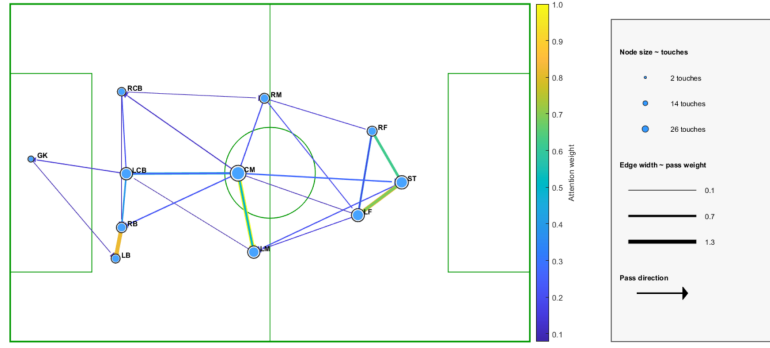


FIGURE 1. Passing Network Structure

passing weights of the wingers (LB, RB, LF, RF) are relatively concentrated, reflecting the tactical characteristics of wide-area advancement.

B. T-GAT ARCHITECTURE

1) GAT Model

Linear Projection and Edge-Node Joint Features

First, linear projection is performed on nodes and edges to form a representation of the same scale, as shown in formula (4).

$$\tilde{h}_i^{(l),t} = W_h^{(l)} h_i^{(l),t}, \quad \tilde{e}_{ij}^{(l),t} = W_e^{(l)} e_{ij}^{(l),t} \quad (4)$$

In formula (4), $h_i^{(l),t}$ represents the node's feature in the window, derived from $\tilde{x}$. $W_h^{(l)}$ and $W_e^{(l)}$ represent the weight matrices of linear transformation and edge transformation, respectively.

For each directed neighbor pair, a joint representation is constructed for attention calculation and message generation, as shown in formula (5).

$$u_{ij}^{(l),t} = [\tilde{h}_i^{(l),t} \parallel \tilde{h}_j^{(l),t} \parallel \tilde{e}_{ij}^{(l),t}] \quad (5)$$

In formula (5), $\parallel$ represents vector concatenation, and $u_{ij}^{(l),t}$ represents the construction of a joint representation for each directed neighbor pair.

Edge-aware multi-head attention weights

An unnormalized score is calculated for each attention head, and to ensure that attention is only calculated on true neighbors, an adjacency mask and a node existence mask are introduced, as shown in formula (6) [13].

$$\begin{aligned} \tilde{\alpha}_{ij}^{(l,k),t} &= \text{LeakyReLU} \left((a^{(l,k)})^\top u_{ij}^{(l),t} \right), \\ \alpha_{ij}^{(l,k),t} &= \tilde{\alpha}_{ij}^{(l,k),t} \cdot I\{A_{ij}^{t,\text{final}} > \epsilon\} \cdot m_i^t \cdot m_j^t \end{aligned} \quad (6)$$

In formula (6), k represents the attention head, $I\{A_{ij}^{t,\text{final}} > \epsilon\}$ represents the adjacency mask, $A_{ij}^{t,\text{final}}$ represents the weighted adjacency matrix after spatial completion and exponential smoothing, and m_i^t and m_j^t represent the node existence mask matrix. Then, softmax normalization is performed on the incoming edge of the target node to obtain the attention coefficient, as shown in formula (7).

$$\alpha_{ij}^{(l,k),t} = \frac{\exp(\tilde{\alpha}_{ij}^{(l,k),t})}{\sum_{u: A_{uj}^{t,\text{final}} > \epsilon} \exp(\tilde{\alpha}_{uj}^{(l,k),t}) + \zeta} \quad (7)$$

In formula (7), ζ represents a numerical stability term with a value of 10^{-12} . $\alpha_{ij}^{(l,k),t}$ represents the normalized attention coefficient.

Message Generation and Multi-Head Aggregation

In each head, a message vector is generated through message transformation using the edge-node joint information, as shown in formula (8).

$$\text{mess}_{ij}^{(l,k),t} = W_{\text{mess}}^{(l,k)} [\tilde{h}_i^{(l),t} \parallel \tilde{e}_{ij}^{(l),t}] \quad (8)$$

In formula (8), $\text{mess}_{ij}^{(l,k),t}$ represents the message vector, and $W_{\text{mess}}^{(l,k)}$ represents the weight matrix of the message transformation.

For the target node, the head aggregation is a weighted sum, and the multi-head output is combined by splicing, as shown in formula (9).

$$\begin{aligned} z_j^{(l,k),t} &= \sum_{i: A_{ij}^{t,\text{final}} > \epsilon} \alpha_{ij}^{(l,k),t} \text{mess}_{ij}^{(l,k),t}, \\ z_j^{(l),t} &= \parallel_{k=1}^K z_j^{(l,k),t} \end{aligned} \quad (9)$$

In formula (9), $z_j^{(l),t}$ represents the merged features, and $z_j^{(l,k),t}$ represents the weighted sum features. To ensure training stability and gradient flow, nonlinearity, residual, and layer normalization are added to each layer, as shown in formula (10).

$$h_i^{(l+1),t} = \text{LayerNorm} \left(h_i^{(l),t} + \text{Dropout} \left(\text{ReLU}(z_i^{(l),t}) \right) \right) \quad (10)$$

In formula (10), $h_i^{(l+1),t}$ represents the result after non-linearity, residual, and layer normalization. At the end of the GAT encoder, attention-based readout is used to map node representations to graph-level vectors, as shown in formula (11).

$$\begin{aligned} l_i^t &= \frac{\exp\left(o^\top \tanh\left(\pi h_i^{(L),t}\right)\right)}{\sum_j \exp\left(o^\top \tanh\left(\pi h_j^{(L),t}\right)\right)}, \\ g_t &= \sum_i l_i^t h_i^{(L),t} \end{aligned} \quad (11)$$

In formula (11), π and o represent learnable parameters, and g_t represents the graph-level representation.

2) GRU Time Series Modeling

To handle unequal delays between windows and prevent old information from excessively influencing the current representation, an exponential decay is applied to the hidden state at the previous moment to obtain $\hat{h}_{t-1}^{(l)}$. For the layer l , the gated update formula is as follows. The update gate formula is shown in formula (12) [14,15].

$$z_t^{(l)} = \sigma\left(\text{LayerNorm}\left(W_z^{(l)} g_t + U_z^{(l)} \hat{h}_{t-1}^{(l)} + b_z^{(l)}\right)\right) \quad (12)$$

In formula (12), σ represents the element-by-element sigmoid function, $W_z^{(l)}$ and $U_z^{(l)}$ represent the corresponding weight matrices, and $b_z^{(l)}$ represents the bias term. The reset gate is shown in formula (13).

$$r_t^{(l)} = \sigma\left(\text{LayerNorm}\left(W_{re}^{(l)} g_t + U_{re}^{(l)} \hat{h}_{t-1}^{(l)} + b_{re}^{(l)}\right)\right) \quad (13)$$

The candidate hidden states are shown in formula (14).

$$\tilde{\pi}_t^{(l)} = \tanh\left(\text{LayerNorm}\left(W_h^{(l)} g_t + U_h^{(l)} \left(r_t^{(l)} \odot \hat{h}_{t-1}^{(l)}\right) + b_h^{(l)}\right)\right) \quad (14)$$

The hidden state is updated as shown in formula (15) [16].

$$\dot{h}_t^{(l)} = \left(1 - z_t^{(l)}\right) \odot \hat{h}_{t-1}^{(l)} + z_t^{(l)} \odot \tilde{\pi}_t^{(l)} \quad (15)$$

In formula (15), $\dot{h}_t^{(l)}$ represents the hidden state. For a multi-layer GRU, $\dot{h}_t^{(l)}$ is calculated layer by layer, and residual projection is introduced between layers to expand the receptive field while avoiding gradient decay. To meet the requirements of multi-task and different time granularity, temporal representation and hybrid representation with retrospective are defined. The temporal representation directly takes the top hidden state as the immediate representation of window t . The hybrid representation with retrospective uses weighted attention fusion of the most recent $\widehat{W}$ steps of hidden states to preserve short-term cumulative effects, as shown in formula (16).

$$\begin{aligned} \varsigma_t^{\text{inst}} &= \dot{h}_t^{(L_r)}, \\ \varrho_{u,t} &= \frac{\exp\left(\varrho^\top \tanh\left(Q \dot{h}_u^{(L_r)} + R \dot{h}_t^{(L_r)}\right)\right)}{\sum_{q=t-\widehat{W}+1}^t \exp(\cdot)}, \\ \varsigma_t^{\text{mix}} &= \sum_{u=t-\widehat{W}+1}^t \varrho_{u,t} \dot{h}_u^{(L_r)} \end{aligned} \quad (16)$$

In formula (16), Q , R , ϱ represent learnable parameters, and $\widehat{W}$ represents the window length. $\varsigma_t^{\text{inst}}$ is the temporal representation, and ς_t^{mix} is the hybrid representation with retrospective. The two are combined into the ς_t temporal representation for regression.

The T-GAT architecture is shown in Figure 2.

In Figure 2, events and trajectories are first synchronized, cleaned, and windowed using sliding/event-driven methods. A directed, attributed pass graph is constructed for each window, and node (position, speed, touch rate, role) and edge (number of passes, success rate, distance, type) features are calculated. Simultaneously, the adjacency matrix is smoothed and spatially completed to mitigate sparsity.

To address the innovation of this novel backbone network, this study introduces a temporal modeling mechanism combining edge-aware multi-head attention and time-decaying GRU on the basis of traditional graph convolutional and graph attention networks. This enables dynamic capture of the joint node-edge representation and retrospective fusion of short-term lag information in the passing network. Compared with existing mainstream architectures (such as U-Net or ResNet variants), this network can not only handle graphstructured data, but also features a customized design specifically for the directed weighting characteristics and multi-task tactical prediction of football passing networks.

C. MULTI-TASK TACTICAL PERFORMANCE REGRESSION HEAD DESIGN

1) Task Structure Design

This paper selects four tactical performance metrics: ΔxG_t , the net xG increment (expected goals increment) within window t ; ΔxT_t , the net expected threat increment (xT increment, used to measure the contribution of passes/movement to offensive threat) within window t ; PPT, the number of progressive passes into the penalty area within window t ; and PNRt, the passing network penetration rate within window t [17,18]. ΔxG and ΔxT directly reflect the contribution of the passing chain to shots and threats, while PP measures network penetration and the ability to pressure the opponent's penalty area. PNR maps the graph structure (paths, important edges) to quantified network penetration, directly reflecting the relationship between the passing network structure and tactical performance [19,20].

In this paper, an independent lightweight MLP (Multi-layer Perceptron) regressor is designed for each regression task, as shown in formula (17).

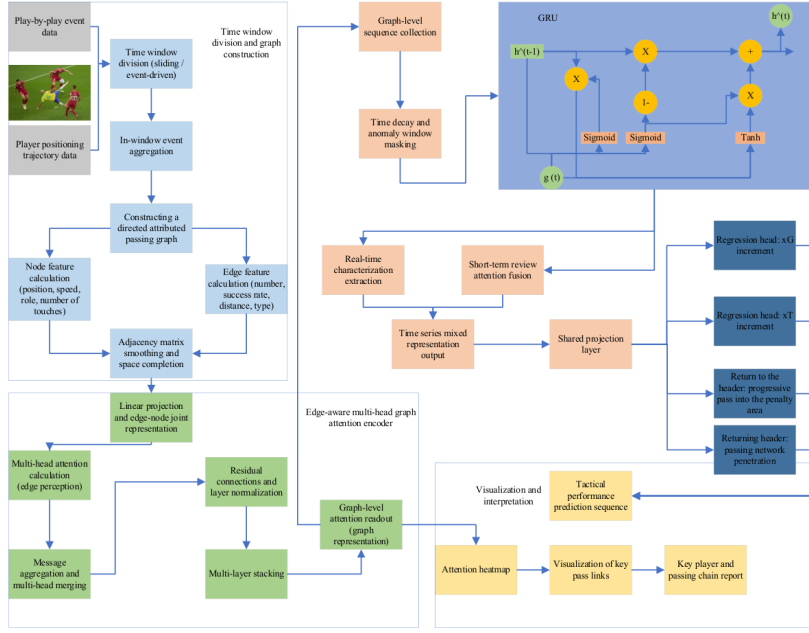


FIGURE 2. T-GAT Architecture

$$\hat{y}_t^{(rw)} = f^{(rw)}(c_t) = W_2^{(rw)} \sigma \left(W_1^{(rw)} c_t + b_1^{(rw)} \right) + b_2^{(rw)} \quad (17)$$

In formula (17), $\hat{y}_t^{(rw)}$ represents the predicted value of task rw , $W_1^{(rw)}$ and $W_2^{(rw)}$ represent learnable weights, and $b_1^{(rw)}$ and $b_2^{(rw)}$ represent biases. To eliminate the magnitude differences between different tasks, the mean and standard deviation of each task are calculated on the training set before training, and the target is z-score normalized to obtain $\tilde{y}_t^{(rw)}$. The training model uses $\tilde{y}$ as the regression target, and after prediction, it is inversely transformed to obtain the original scale value when necessary.

To enhance task-specific performance and preserve shared information, a shared projection layer is first set, as shown in formula (18).

$$\tilde{z}_t = \text{ReLU} (W_{\zeta} c_t + b_{\zeta}) \quad (18)$$

2) Definition of Pass Network Penetration

The window-level PNR is defined as the normalized weighted sum of the source end, as shown in formula (19).

$$PNR_t = \frac{1}{\phi_t} \sum_i w_i^t \left(\sum_{b=1}^{B_{\max}} L_J^b ((A^{t, \text{final}})^b \phi_i) \right) \quad (19)$$

In formula (19), ϕ_t represents the normalization factor, w_i^t represents the node importance weight, and L_J^b represents the path length attenuation factor, which controls the multi-hop contribution. $B_{\max}$ represents the maximum number of path steps, and ϕ_i represents the indicator function vector for the final set of receiving players in the third zone/penalty zone.

3) Overall Loss Function Design and Optimization

The overall optimization objective must simultaneously cover multi-task regression accuracy, time-series smoothness, graph-structure constraints, and parameter regularization.

A mean squared error loss is used for each task, combined with a task uncertainty weighting mechanism. The regression loss is defined as shown in formula (20).

$$Loss_{\text{reg}} = \sum_{rw=1}^{RW} \left(\frac{1}{2\sigma_{rw}^2} \cdot \frac{1}{N_T} \sum_{t=1}^T M_t \left(\tilde{y}_t^{(rw)} - \hat{y}_t^{(rw)} \right)^2 + \log \sigma_{rw} \right) \quad (20)$$

In formula (20), M_t represents the effective mask of the time window, and $\tilde{y}_t^{(rw)}$ represents the true value. To reduce the jitter of the prediction results of short time windows, a smoothing constraint is introduced between adjacent windows, as shown in formula (21).

$$Y S_{\text{smooth}} = \frac{1}{N_T} \sum_{t=2}^T M_t M_{t-1} \left\| \hat{y}_t^{(rw)} - \hat{y}_{t-1}^{(rw)} \right\|_2^2 \quad (21)$$

In formula (21), $\hat{y}_t^{(rw)}$ represents the multi-task prediction vector. To ensure that the model prediction is consistent with the statistical characteristics of the passing network structure, a graph structure regularization is defined as shown in formula (22).

$$Y S_{\text{graph}} = \frac{1}{T} \sum_{t=1}^T \left\| \hat{y}_t^{(\text{PNR})} - \chi(A^t, H^t) \right\|_2^2 \quad (22)$$

In formula (22), $\hat{y}_t^{(\text{PNR})}$ represents the predicted PNR value, A_t represents the adjacency matrix of the time window, H_t represents the node representation output by GAT, and χ represents the structural permeation function based on the graph path.

The overall loss function is shown in formula (23).

$$Loss = Loss_{\text{reg}} + \lambda_{\text{smooth}} Y S_{\text{smooth}} + \lambda_{\text{graph}} Y S_{\text{graph}} \quad (23)$$

In formula (23), λ_{smooth} and λ_{graph} represent the corresponding weight coefficients, which are used to control the contribution of each regularization to the overall goal. The experiment uses AdamW for optimization, and the gradient is clipped and constrained to avoid gradient explosion [21–23]. In addition, a cosine annealing strategy is used for learning rate scheduling. The comprehensive indicator RMSE is monitored on the validation set. If there is no improvement after consecutive N_{stop} epochs, training is terminated [24].

III. EXPERIMENTAL PLAN

A. EXPERIMENTAL DATA AND PREPROCESSING

The experimental data for this study comes from the StatsBomb dataset, which contains publicly available match event data and match metadata. The data fields include individual pass events (pass start and end locations, timestamps, pass results, and pass labels such as short/long/through), shot events, player and team identifiers, substitution information, and match phase information. To ensure the generalizability and sample richness of the method, the sample covers multiple seasons and different leagues, comprising over 1,200 matches, hundreds of thousands of passing events, and touches by approximately 10,000 players. The experiment first performed time alignment and consistency checks on the raw event data, removing game records with significant omissions or abnormal timestamps. On-field coordinates were then unified into a unified coordinate system and jitter smoothed to reduce positioning noise. Matches are segmented into event-driven windows or fixed-length sliding windows using event time. Initial node and edge statistics (including number of touches, number of passes, success rate, average pass distance, and type distribution) are constructed for pass events within each time window. Presence masks are applied to mitigate graph fragmentation in cases where players play for short periods or when substitutes are sparse, and spatial adjacency completion is performed on low-frequency nodes. Node and edge feature vectors (such as position, velocity estimates, player one-hot, and pass attributes) are calculated and normalized. Empirical decay or normalization is applied to pass distance and frequency to mitigate the influence of extreme values. The data was split into 70% training, 15% validation, and 15% test sets based on the competition to avoid time leakage. Finally, all features were normalized on the training set, and the preprocessing pipeline was saved to ensure reproducibility and consistency across experiments.

B. SOLUTION DESIGN

The experimental design was conducted in two steps to ensure comparability and robustness of the conclusions. First, data preparation uses a training-validation-test set divided by game/season. Event-driven or fixed sliding windows are used to construct the input graph sequence and generate four regression labels (ΔxG , ΔxT , PP, and PNR). All models use the same preprocessing pipeline and normalization parameters. During training, a fixed random seed is used and repeated several times (10 times) for reporting the mean and standard deviation. Evaluation metrics include RMSE, MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). During testing, statistical reports are also provided on operational efficiency (inference time, number of parameters, and video memory usage). Hyperparameters are adjusted using grid search and Bayesian optimization on the validation set. Training utilizes AdamW optimization, gradient clipping, and early stopping. Paired T-tests are performed on the final model on the test set to confirm the statistical significance of performance differences.

The comparative experiments are as follows: the main comparison group includes the T-GAT model (the method in this paper), ST-GCN (Spatial Temporal Graph Convolutional Networks), GAT, GCN, GraphSAGE (Graph Sample and Aggregate), and BiLSTM (Bidirectional Long Short-Term Memory). Each baseline is implemented with as equivalent capacity/parameter count as possible (same number of layers, similar hidden dimensions) to ensure fairness. When analyzing the relationship between a football player's passing network structure and tactical performance, interpretability was assessed by visualizing the distribution of attention edge weights and key pass links. Pearson and Spearman rank correlation coefficients were also analyzed, along with partial correlation coefficients. The correlation between the passing network structure and tactical performance at different ball possession rates was further analyzed.

Table 1 shows the experimental parameters.

IV. RELATIONSHIP BETWEEN FOOTBALL PLAYER PASSING NETWORK STRUCTURE AND TACTICAL PERFORMANCE

A. COMPARISON OF OVERALL TACTICAL PERFORMANCE REGRESSION PERFORMANCE

This paper compares the overall regression performance of various models on the test set for four tactical regression metrics (ΔxG , ΔxT , PP, and PNR). The RMSE, MAE, and MAPE (%) are reported. The results are shown in Figure 3. All results are the average of multiple (10) replicates, using the same preprocessing and training/validation splits.

In Figures 3 (a) and 3 (b), proposed T-GAT model achieves optimal results in both tactical performance prediction for ΔxG and ΔxT . The RMSE, MAE, and MAPE for ΔxG are 0.07, 0.05, and 16.45%, respectively, significantly improving over ST-GCN (0.09, 0.07, and 21.30%) and GCN (0.12, 0.09, and 27.80%).

TABLE 1. Experimental Parameters

Parameters	Experimental setup	Parameter adjustment range	Parameters	Experimental setup	Parameter adjustment range
Initial learning rate	1.00E-03	1e-4 — 5e-3	Attention head dropout	0.2	0.0 — 0.5
Weight decay (L2)	1.00E-05	0 — 1e-3	Inter-layer dropout	0.3	0.0 — 0.5
Batch size	16	8 — 64	Number of GRU stacking layers	2	1 — 3
Maximum training epochs	100	50 — 300	GRU hidden dimension	256	64 — 512
Number of early stopping epochs for validation invariance	10	5 — 20	GRU time constant initial value	10.0 s	1—30 s
Gradient clipping threshold (L2 norm)	1	0.5 — 5.0	Input time window sequence length	20 windows	10 — 40
Number of GAT heads	4	2 — 8	Per time window duration	10 s	5 s — 30 s / event-driven
Number of GAT layers	2	1 — 4	Shared projection layer dimension	128	64 — 256
Output dimension per head	64 (per head) → 256 total	32–128 per head	Regression head hidden dimension	128	32 — 256
Edge feature projection dimension	32	16 — 128	Temporal smoothing regularization coefficient	1.00E-02	1e-4 — 1e-1
Node projection dimension	128	64 — 256	Graph structure consistency regularization coefficient	1.00E-03	1e-5 — 1e-2

In Figures 3(c) and 3(d), the T-GAT model continues to demonstrate leading performance in both PP and PNR metrics. For PP, the T-GAT model's RMSE, MAE, and MAPE are 0.28, 0.18, and 13.80%, respectively, significantly lower than ST-GCN (0.39, 0.27, 21.40%), GCN (0.56, 0.40, 33.25%), and BiLSTM (0.62, 0.44, 37.75%).

To explore the significance of tactical performance prediction, a two-sample Welch t-test was used for analysis. The steps are as follows:

(1) Each model on the test set was trained and evaluated 10 times under the same data partitioning and the same random seed, and the RMSE (ΔxG and PNR) was recorded each time.

(2) The two-sample Welch t-test (without assuming equal variances) was used for the comparative test. The test statistic was based on the mean difference between two independent samples; the significance level was 0.05.

(3) For each pair of "this model T-GAT model vs. comparison model", the mean and sample standard deviation (mean $\pm$ std), t-value, two-sided p-value, and 95% confidence interval (confidence interval of the mean difference) were calculated.

(4) If $p < 0.05$, the difference is considered significant; for ease of reading, $p < 0.001$ is marked as "Significant ($p < 0.001$)".

The hypothesis is as follows:

Null hypothesis H_0 : There is no significant difference in the average RMSE between the T-GAT model and the comparison model.

Alternative hypothesis H_A : There is a significant difference in the mean values of the two models. The T-test results are shown in Table 2.

The t-test results in Table 2 show that the T-GAT model significantly outperforms all compared models in both ΔxG and PNR, with the differences being highly significant ($p < 0.001$). For ΔxG , the RMSE for the T-GAT model is 0.07 ± 0.01 , while those for ST-GCN, GAT, GCN, GraphSAGE, and BiLSTM are 0.09 ± 0.01 , 0.10 ± 0.02 , 0.12 ± 0.03 , 0.11 ± 0.02 , and 0.13 ± 0.05 , respectively. The corresponding t-value reaches a maximum of -18.46 (BiLSTM), and the 95% confidence interval for the mean difference ranges from (-0.07, -0.05) to (-0.03, -0.01). In terms of pass-to-noise ratio (PNR), the T-GAT model achieved an RMSE of 0.06 ± 0.01 , while ST-GCN, GAT, GCN, GraphSAGE, and BiLSTM achieved RMSEs of 0.08 ± 0.01 , 0.10 ± 0.01 ,

0.13 ± 0.02 , 0.11 ± 0.02 , and 0.15 ± 0.03 , respectively. The corresponding t-value reached a maximum of -31.82 (BiLSTM), with the minimum confidence interval difference reaching (-0.02, -0.02) and the maximum reaching (-0.10, -0.08).

B. TIME SERIES PERFORMANCE ANALYSIS

To verify the time-varying nature and time series analysis performance of the T-GAT model, this paper analyzes the prediction performance of each indicator for the game time periods 0-15, 16-30, 31-45, 46-60, 61-75, and 76-90. The results are shown in Figure 4.

In Figure 4, T-GAT maintains the lowest error across all time periods and metrics, and exhibits a smaller increase over time. Taking ΔxG as an example, T-GAT's RMSE for the 0-15, 16-30, 31-45, 46-60, 61-75, and 76-90 time periods is 0.06, 0.07, 0.07, 0.08, 0.08, and 0.09, respectively. For GCN, the RMSE for the corresponding time periods is 0.11, 0.12, 0.12, 0.13, 0.12, and 0.12. ST-GCN and GAT's RMSE ranges are approximately 0.08–0.11 and 0.09–0.11, respectively, both exceeding T-GAT. For ΔxT , T-GAT's RMSE ranged from 0.12–0.15, significantly outperforming BiLSTM (0.25–0.27), GCN (0.22–0.25), and GAT (0.18–0.21). For the performance metric, T-GAT achieved a performance of 0.25–0.30, while GCN, GAT, and BiLSTM achieved performances of approximately 0.53–0.58, 0.42–0.46, and 0.60–0.64, respectively. The difference is particularly significant. For performance metric (PNR), T-GAT's RMSE was only 0.05–0.07, while GCN's was 0.12–0.14, and GAT's was 0.09–0.11. The MAE also shows a consistent trend. T-GAT achieves the lowest mean absolute error across all metrics and time periods, with minimal fluctuation. For ΔxG , T-GAT's MAE for the corresponding time periods is 0.05, 0.05, 0.05, 0.06, 0.06, and 0.07, while GCN maintains a constant MAE of approximately 0.09 and GAT ranges from 0.07 to 0.08. For ΔxT , T-GAT achieves a MAE of 0.09–0.12 across all time periods, significantly outperforming GCN's MAE of 0.16–0.19 and BiLSTM's MAE of 0.19–0.20. For PP MAE, T-GAT maintains a stable MAE of 0.19–0.23, while GCN/GAT/BiLSTM achieve MAEs of 0.39–0.44, 0.30–0.35, and 0.43–0.48, respectively, showing a significant difference. T-GAT maintains a PNR MAE of 0.04–0.05, significantly lower than GCN's 0.09–0.11.

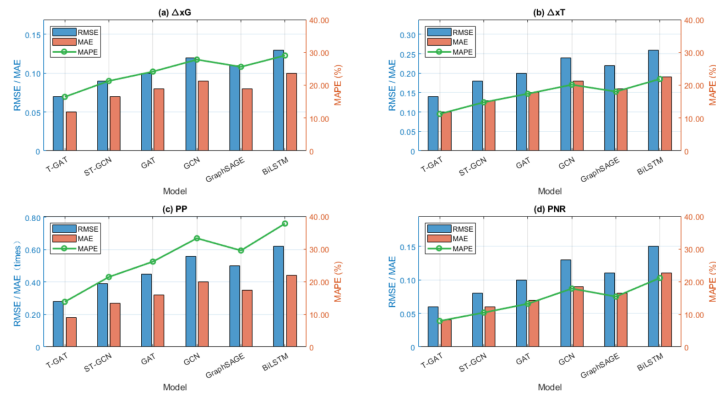


FIGURE 3. Overall Tactical Performance Regression Performance Comparison: (a) ΔxG ; (b) ΔxT ; (c) PP; (d) PNR

TABLE 2. T-test results

Indicator	Model	Presented model T-GAT (mean $\pm$ std)	Compared model (mean $\pm$ std)	t-value	p-value	Significance	95% Confidence interval
ΔxG	ST-GCN	0.07 $\pm$ 0.01	0.09 $\pm$ 0.01	-8.10	<0.001	significant (p<0.001)	(-0.03, -0.01)
	GAT	0.07 $\pm$ 0.01	0.10 $\pm$ 0.02	-11.03	<0.001	significant (p<0.001)	(-0.04, -0.02)
	GCN	0.07 $\pm$ 0.01	0.12 $\pm$ 0.03	-16.72	<0.001	significant (p<0.001)	(-0.06, -0.04)
	GraphSAGE	0.07 $\pm$ 0.01	0.11 $\pm$ 0.02	-14.71	<0.001	significant (p<0.001)	(-0.05, -0.03)
	BiLSTM	0.07 $\pm$ 0.01	0.13 $\pm$ 0.05	-18.46	<0.001	significant (p<0.001)	(-0.07, -0.05)
PNR	ST-GCN	0.06 $\pm$ 0.01	0.08 $\pm$ 0.01	-9.88	<0.001	significant (p<0.001)	(-0.02, -0.02)
	GAT	0.06 $\pm$ 0.01	0.10 $\pm$ 0.01	-17.54	<0.001	significant (p<0.001)	(-0.05, -0.04)
	GCN	0.06 $\pm$ 0.01	0.13 $\pm$ 0.02	-27.45	<0.001	significant (p<0.001)	(-0.08, -0.06)
	GraphSAGE	0.06 $\pm$ 0.01	0.11 $\pm$ 0.02	-21.93	<0.001	significant (p<0.001)	(-0.06, -0.05)
	BiLSTM	0.06 $\pm$ 0.01	0.15 $\pm$ 0.03	-31.82	<0.001	significant (p<0.001)	(-0.10, -0.08)

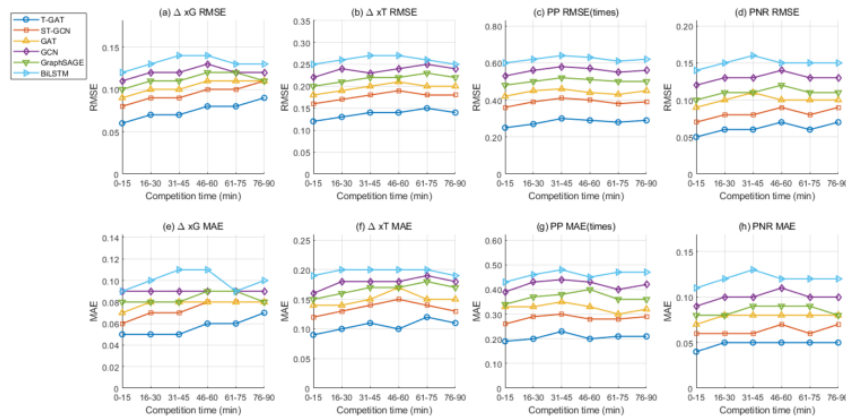


FIGURE 4. Timing Performance Analysis: (a) ΔxG RMSE; (b) ΔxT RMSE; (c) PP RMSE; (d) PNR RMSE; (e) ΔxG MAE; (f) ΔxT MAE; (g) PP MAE; (h) PNR MAE

C. INTERPRETABILITY AND ATTENTION ANALYSIS

To explore the interpretability of the T-GAT model, the paper used key pass link visualization (attention edge weights) to identify key players and key links. The key pass link visualization (attention edge weights) is shown in Figure 5. In Figure 5, the formations include 4-3-3, 4-4-2, 4-5-1, and 3-5-2.

To elevate attention weighting from visualization to actionable tactical insights for coaches/analysts, each directed edge was first normalized and aggregated over time (average attention, peak, and persistence), and then sorted by “inten-

sity $\times$ frequency” to obtain candidate key pathways. Next, lagged correlation and partial correlation analyses (after controlling for confounding variables such as possession rate, position, and overall network density) were used to examine the statistical correlation between high-weighted edges and subsequent $\Delta xG/\Delta xT$ increments. A virtual ablation/perturbation experiment was conducted (resetting the weight of a certain edge to zero or amplifying it by a certain proportion), and the marginal contribution of that edge was quantified by the change in $\Delta xG/\Delta xT$ in the model’s predicted output, resulting in a quantifiable “contribution

score". Finally, the results were packaged into a coach-friendly output, including a Top-K key passing channel list labeled by on-field position, a scenario-based heatmap (such as a prompt like "Increasing the frequency of short passes from CM to LF near the opponent's penalty area can increase the average predicted xG within 10 seconds"), and targeted training suggestions (such as increasing short pass penetration training or adjusting the connection of key players). This process ensures the interpretability of attention weighting and provides verifiable and actionable tactical suggestions, making it easy for coaches to use directly in pre-match/half-match adjustments.

In the visualization of key pass links in Figure 5, it can observe significant differences in the core players and high-weight links in the passing network across different formations. In the 4-3-3 formation, the reverse link from forward to goalkeeper (e.g., 9 (LF) $\rightarrow$ 10 (ST)) has an attention weight as high as 0.85, and 10 (ST) $\rightarrow$ 11 (RF) reaches 1.00, demonstrating the importance of close coordination in the frontcourt for the finishing phase. In the 4-4-2 formation, the interaction between midfielders and wingers is particularly prominent. For example, the attention for the link 4 (RCB) $\rightarrow$ 5 (RB) reaches a high of 0.95, and for 6 (LM) $\rightarrow$ 3 (LCB) it reaches 0.37, demonstrating the crucial role of midfield support in a two-striker system.

D. OPERATIONAL EFFICIENCY ANALYSIS

To evaluate the practical feasibility of each model in real-world deployments, this section compares training time, single-sample inference latency, model parameter count, and graphics memory usage using the same hardware and data/preprocessing conditions. Training time represents the total training time. Inference latency represents the average single-window latency at a batch size of 1; parameter count is measured in megabytes. Graphics Processing Unit (GPU) memory usage is the peak GB used during training.

All inference speed and runtime efficiency tests were performed on the same hardware environment. The experimental platform included an NVIDIA RTX 4090 GPU (24 GB VRAM), an Intel Core i9-13900K CPU (3.0 GHz, 32 cores), 64 GB DDR5 memory, and an Ubuntu 22.04 operating system. Inference latency was measured using the average latency of a single window with a batch size of 1 (averaged over 1000 runs), and was repeatedly tested to ensure stability. Table 3 shows the results of the performance analysis.

TABLE 3. Performance Analysis Results

Model	Training time (Hours)	Inference latency (ms/window)	Model parameters (M)	GPU memory usage (GB)
T-GAT	4.30	6.40	3.80	4.20
ST-GCN	9.45	12.35	6.12	6.72
GAT	7.80	10.50	5.50	5.90
GCN	4.95	7.20	3.95	4.45
GraphSAGE	5.60	8.90	4.20	4.95
BiLSTM	3.25	5.85	2.70	3.10

Table 3 shows significant differences in operational effi-

ciency among different models. Overall, BiLSTM achieved the best performance in terms of training time (3.25 hours), inference latency (5.85 ms/window), number of parameters (2.70 MB), and video memory usage (3.10 GB), but its architecture failed to effectively model graph structural relationships. Proposed T-GAT maintains strong modeling capabilities while also balancing efficiency. Its training time is only 4.30 hours, its inference latency is 6.40 ms/window, its parameter count is 3.80 MB, and its GPU memory usage is 4.20 GB. Its efficiency is close to that of GCN, but significantly outperforms ST-GCN and GAT.

E. PEARSON AND SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN PASSING NETWORK STRUCTURAL CHARACTERISTICS AND TACTICAL PERFORMANCE

This experiment examined the relationship between passing network structural characteristics, including average node strength, degree centrality, average shortest path length, global clustering coefficient, network density, and maximum eigenvector centrality, and the tactical performance indicators ΔxG , ΔxT , PP, and PNR. Pearson and Spearman rank correlation coefficients were calculated, and the results are shown in Figure 6. In Figure 6, *** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$.

The Pearson correlation coefficient results (Figure 6 (a)) show that different passing network structure characteristics are significantly correlated with tactical performance. Specifically, network density is significantly positively correlated with ΔxG (0.84***), PP (0.86***), and PNR (0.25***), indicating that a tighter passing network helps improve expected goal difference and offensive efficiency.

The Spearman rank correlation coefficient results (Figure 6(b)) show consistent trends with the Pearson correlation results, but the rank correlation more clearly reflects non-linear relationships. Maximum eigenvector centrality shows significant positive correlations with both PNR (0.92***) and ΔxT (0.67***), indicating that the greater the influence of the core node, the more prominent the team's performance in offensive progression and transitions.

F. PARTIAL CORRELATION VERIFICATION

To eliminate the influence of other structural indicators, the paper sequentially examined average node strength, degree centrality, average shortest path length, global clustering coefficient, network density, and maximum eigenvector centrality. It investigated the partial correlation coefficients of each structural feature with tactical performance (ΔxG , ΔxT , PP, and PNR) while controlling for the other five features. The results are shown in Figure 7.

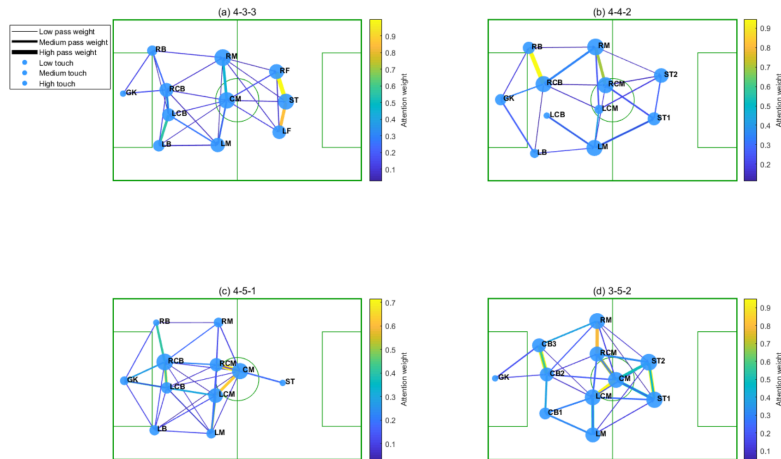


FIGURE 5. Visualization of Key Pass Links Based on Attention Edge Weights: (a) 4-3-3; (b) 4-4-2; (c) 4-5-1; (d) 3-5-2

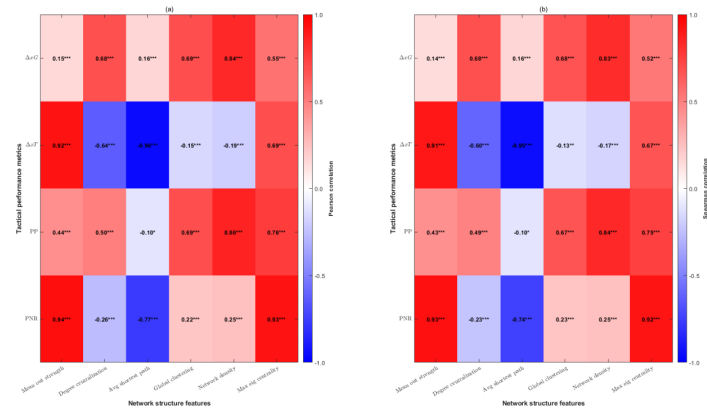


FIGURE 6. Pearson Correlation Coefficient and Spearman Rank Correlation Coefficient between Passing Network Structure and Tactical Performance: (a) Pearson Correlation Coefficient between Passing Network Structure Characteristics and Tactical Performance; (b) Spearman Rank Correlation Coefficient between Passing Network Structure Characteristics and Tactical Performance

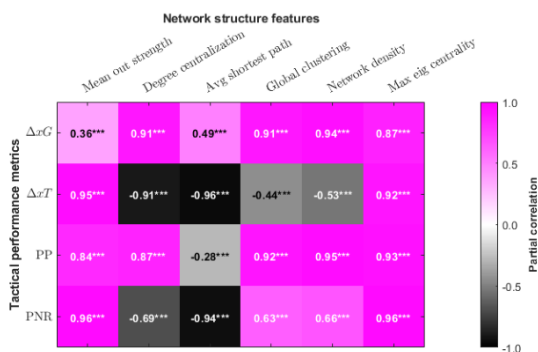


FIGURE 7. Partial Correlation Coefficients

The partial correlation coefficient results (Figure 7) demonstrate that different network structural characteristics

still significantly influence tactical performance even after controlling for other indicators. Average node out-put strength maintains a highly positive correlation with ΔxT (0.95***) and PNR (0.96**), indicating that passing activity directly contributes to offensive progression and transition efficiency.

G. CORRELATION BETWEEN PASSING NETWORK STRUCTURE CHARACTERISTICS AND TACTICAL PERFORMANCE AT DIFFERENT BALL POSSESSION RATES

The experiment categorized the teams' ball possession rates in each time window of the game into three groups: low, medium, and high. The Pearson correlation coefficient was calculated for each group. The results are shown in Figure 8.

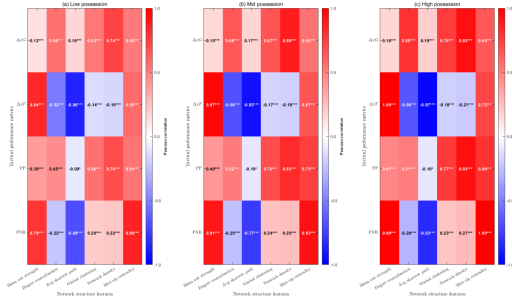


FIGURE 8. Correlation between passing network structural characteristics and tactical performance at different ball possession rates: (a) Low ball possession; (b) Medium ball possession; (c) High ball possession

Under low ball possession (Figure 8a), the relationship between a team's passing network structural characteristics and tactical performance is relatively weak, but certain patterns can still be observed. Network density maintains a strong positive correlation with ΔxG (0.74***) and PP (0.70***), indicating that a more compact passing network can help create offensive opportunities even when possession is limited. Under moderate possession conditions (Figure 8b), the overall correlation increases, indicating that the passing network structure better reflects tactical value under balanced possession conditions. Network density is more closely correlated with ΔxG (0.89***) and PP (0.83***), highlighting the crucial role of tight passing in transitions and chance creation.

Under high possession conditions (Figure 8c), the correlation reaches a strong level, indicating that the passing network structure is almost fully coupled to tactical performance. Average node outgoing strength exhibits a near-perfect positive correlation with ΔxT (1***) and PNR (0.99***). This suggests that teams with frequent and balanced passing in high-possession situations can significantly improve their tactical execution.

H. COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART METHODS

To further validate the advantages of the T-GAT model in predicting tactical performance in football passing networks, this section compares it with four recently proposed state-of-the-art methods: Graphormer, ST-GraphSAGE, DyGNN, and TrajGAT. The experimental procedure remained consistent with the previous experiments, using the same StatsBomb dataset, divided into training, validation, and test sets at 70%/15%/15%. Preprocessing and feature normalization were consistent with T-GAT. For each state-of-the-art model, a publicly available implementation or a reproducible version was used, maintaining similar training parameters (learning rate, batch size, number of training epochs, etc.) as T-GAT, and the training was repeated 10 times, recording the mean performance and standard deviation. Performance was evaluated using four tactical regression metrics: ΔxG , ΔxT , PP, and PNR, and RMSE, MAE, and MAPE were calculated. The results of the comparative analysis with

state-of-the-art methods are shown in Table 4.

As shown in Table 4, the T-GAT model exhibits the best performance across all four tactical regression indices. In ΔxG prediction, T-GAT's RMSE is 0.07 and MAE is 0.05, both lower than DyGNN (0.08, 0.06), and its MAPE is also reduced to 16.45%, demonstrating higher prediction accuracy for offensive threat gain. In the ΔxT index, T-GAT's RMSE is 0.14 and MAE is 0.10, both lower than the second-best performing TrajGAT (0.15, 0.11), indicating more stable modeling of pass progression value. In the PP index, T-GAT's RMSE and MAE are 0.28 and 0.18, respectively, significantly better than Graphormer (0.33, 0.22) and ST-GraphSAGE (0.35, 0.24), with a reduced average error. In terms of PNR (Pass Rate), T-GAT achieved the lowest RMSE and MAE (Mean Effect) values of 0.06 and 0.04 respectively, with a MAPE of 7.95%, demonstrating superior predictive ability for pass consistency and organizational efficiency. Overall, T-GAT outperformed T-GAT in RMSE and MAE, validating its significant advantages in complex spatiotemporal relationship modeling and multi-task prediction. The StatsBomb sample used in this study covers multiple seasons and several top leagues (>1200 matches, hundreds of thousands of passes, $\approx 10,000$ players), thus the conclusions are robust within this data distribution; however, it acknowledges the risk of bias due to uneven league/team distribution, differences in tactical styles, and data collection quality (tracking system and event annotation). This paper emphasizes that structural correlation conclusions should not be directly extrapolated to leagues not included, amateur/youth levels, or scenarios using different data sources. Future work will conduct cross-league external validation, stratified (league/season) performance analysis, and domain adaptation/transfer learning research to quantify and improve the model's generalization ability.

I. ABLATION EXPERIMENT

To verify the contribution of each module in the T-GAT model to tactical performance prediction, an ablation experiment was designed in this section. The steps are as follows:

(1) The complete T-GAT model was used, including edge-aware multi-head attention and GRU time-series modeling, as well as regression head and task-specific attention gating mechanisms.

(2) The following modules were removed and compared: w/o Edge-Att: Edge-aware multi-head attention was removed, and only node attention was used. w/o Temporal-GRU: GRU time-series modeling was removed, and only graph attention was used for independent prediction of each window. w/o Task-Att: Task-specific attention gating was removed, and direct regression was used with shared temporal representation. w/o Both: Both edge-aware attention and GRU time-series modeling were removed simultaneously, and only the basic GAT + MLP was used.

(3) Using the same data partitioning, training strategy, and evaluation metrics as the complete model, the experiment was repeated 10 times, and the average and standard deviation

TABLE 4. Comparative Analysis Results with State-of-the-Art Methods

Indicators	Model	RMSE	MAE	MAPE (%)
ΔxG	T-GAT	0.07 $\pm$ 0.01	0.05 $\pm$ 0.01	16.45
	Graphormer	0.09 $\pm$ 0.01	0.06 $\pm$ 0.01	19.82
	ST-GraphSAGE	0.10 $\pm$ 0.02	0.07 $\pm$ 0.01	21.13
	DyGNN	0.08 $\pm$ 0.01	0.06 $\pm$ 0.01	18.74
	TrajGAT	0.09 $\pm$ 0.01	0.06 $\pm$ 0.01	20.25
ΔxT	T-GAT	0.14 $\pm$ 0.02	0.10 $\pm$ 0.01	11.25
	Graphormer	0.17 $\pm$ 0.02	0.12 $\pm$ 0.01	14.8
	ST-GraphSAGE	0.18 $\pm$ 0.02	0.13 $\pm$ 0.02	15.92
	DyGNN	0.16 $\pm$ 0.02	0.11 $\pm$ 0.01	13.65
	TrajGAT	0.15 $\pm$ 0.02	0.11 $\pm$ 0.01	13.1
PP	T-GAT	0.28 $\pm$ 0.03	0.18 $\pm$ 0.02	13.80
	Graphormer	0.33 $\pm$ 0.03	0.22 $\pm$ 0.02	16.54
	ST-GraphSAGE	0.35 $\pm$ 0.04	0.24 $\pm$ 0.03	18.01
	DyGNN	0.31 $\pm$ 0.03	0.21 $\pm$ 0.02	15.32
	TrajGAT	0.30 $\pm$ 0.03	0.20 $\pm$ 0.02	14.98
PNR	T-GAT	0.06 $\pm$ 0.01	0.04 $\pm$ 0.01	7.95
	Graphormer	0.08 $\pm$ 0.01	0.05 $\pm$ 0.01	9.42
	ST-GraphSAGE	0.09 $\pm$ 0.01	0.06 $\pm$ 0.01	10.31
	DyGNN	0.07 $\pm$ 0.01	0.05 $\pm$ 0.01	8.68
	TrajGAT	0.07 $\pm$ 0.01	0.05 $\pm$ 0.01	8.2

tion were taken. The RMSE, MAE, and MAPE of the four tactical performance metrics were compared. The ablation test results are shown in Table 5.

Based on the ablation experiment results in Table 5, it can be seen that each module in the T-GAT model makes a significant contribution to the prediction of tactical performance. The complete model performs best on the four indicators, with RMSEs of 0.07, 0.14, 0.28, and 0.06 for ΔxG , ΔxT , PP, and PNR, respectively, and MAEs of 0.05, 0.10, 0.18, and 0.04, respectively, corresponding to MAPEs of 16.45%, 11.25%, 13.80%, and 7.95%. After removing edge-aware attention (w/o Edge-Att), the RMSEs of all indicators increase, such as ΔxG RMSE increasing to 0.09 and ΔxT RMSE increasing to 0.17, demonstrating the importance of edge information in graph structure modeling. Removing GRU temporal modeling (w/o Temporal-GRU) increased the RMSE of ΔxG and ΔxT to 0.10 and 0.18, respectively, and also significantly increased the RMSE of PP and PNR, indicating that time series modeling is crucial for prediction accuracy. Removing task-specific attention (w/o Task-Att) led to varying degrees of increase in RMSE, MAE, and MAPE, but the magnitude was smaller than the former two, indicating that this module has an auxiliary role in optimizing task-related representations. When both side attention and GRU were removed simultaneously (w/o Both), all indicators deteriorated significantly, with ΔxG RMSE reaching 0.12, ΔxT RMSE reaching 0.20, PP RMSE reaching 0.38, and PNR RMSE reaching 0.11, and the corresponding MAPE all exceeding 13%–23%, further validating the synergistic effect and overall importance of each module. Overall, edge-aware attention and time-series modeling are the core modules for improving the prediction accuracy of T-GAT, while task-specific attention provides additional optimization.

V. CONCLUSION

The T-GAT framework constructed in this paper divides the game into time windows and constructs a directed attributed passing graph. The window-level graph representation is

serialized and a GRU with time decay and short-term retrospective attention are introduced to produce a hybrid temporal representation. The shared projection is then used in parallel to enter the four-way regression head to predict ΔxG , ΔxT , PP and PNR, forming a complete end-to-end analysis. Experiments show that T-GAT performs well on the test set (ΔxG RMSE = 0.07, ΔxT RMSE = 0.14, PP RMSE = 0.28, PNR RMSE = 0.06), and significantly outperforms the comparison model in statistical tests ($p < 0.001$). Structural correlation analysis reveals a Pearson correlation coefficient of 0.84 between network density and ΔxG , and a Pearson correlation of 0.92 between average node output strength and ΔxT . Partial correlation tests maintain strong correlations even after controlling for other features. However, causal inference is insufficient, and further optimization is needed for real-time deployment. Future work can expand causal testing, model distillation, and on-line updates to enhance explanatory depth and engineering usability. Applying these techniques to industrial settings, such as manufacturing, will allow for the deployment of highly efficient, real-time systems that provide deeper causal insight into production process optimization.

AUTHOR CONTRIBUTIONS

Conceptualization – Yongkang Guan and Weijia Xue; methodology – Yongkang Guan and Weijia Xue; investigation – Yongkang Guan and Weijia Xue; writing-original draft preparation – Yongkang Guan and Weijia Xue. 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.

TABLE 5. Ablation Test Results

Model	ΔxG			ΔxT			PP			PNR		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
T-GAT (Complete Model)	0.07 $\pm$ 0.01	0.05 $\pm$ 0.01	16.45	0.14 $\pm$ 0.02	0.10 $\pm$ 0.01	11.25	0.28 $\pm$ 0.03	0.18 $\pm$ 0.02	13.80	0.06 $\pm$ 0.01	0.04 $\pm$ 0.01	7.95
w/o Edge-Att	0.09 $\pm$ 0.02	0.06 $\pm$ 0.01	19.10	0.17 $\pm$ 0.02	0.12 $\pm$ 0.02	13.05	0.32 $\pm$ 0.04	0.21 $\pm$ 0.03	16.30	0.08 $\pm$ 0.01	0.06 $\pm$ 0.01	10.15
w/o Temporal-GRU	0.10 $\pm$ 0.02	0.07 $\pm$ 0.01	20.50	0.18 $\pm$ 0.03	0.13 $\pm$ 0.02	14.25	0.34 $\pm$ 0.04	0.22 $\pm$ 0.03	17.20	0.09 $\pm$ 0.01	0.06 $\pm$ 0.01	10.80
w/o Task-Att	0.08 $\pm$ 0.01	0.06 $\pm$ 0.01	18.20	0.15 $\pm$ 0.02	0.11 $\pm$ 0.01	12.10	0.30 $\pm$ 0.03	0.20 $\pm$ 0.02	15.50	0.07 $\pm$ 0.01	0.05 $\pm$ 0.01	8.90
w/o Both	0.12 $\pm$ 0.03	0.09 $\pm$ 0.02	23.75	0.20 $\pm$ 0.03	0.15 $\pm$ 0.02	16.50	0.38 $\pm$ 0.05	0.26 $\pm$ 0.03	20.00	0.11 $\pm$ 0.02	0.08 $\pm$ 0.01	13.45

REFERENCES

- [1] B. Low, R. Rein, S. Schwab, and D. Memmert, "Defending in 4-4-2 or 5-3-2 formation? Small differences in footballers' collective tactical behaviours," *Journal of Sports Sciences*, vol. 40, no. 3, pp. 351-363, 2022, doi: 10.1080/02640414.2021.1993655.
- [2] N. Coito, K. Davids, H. Folgado, T. Bento, and B. Travassos, "Capturing and quantifying tactical behaviors in small-sided and conditioned games in soccer: a systematic review," *Research quarterly for exercise and sport*, vol. 93, no. 1, pp. 189-203, 2022, doi: 10.1080/02701367.2020.1823307.
- [3] G. Jin, "Player target tracking and detection in football game video using edge computing and deep learning," *The Journal of Supercomputing*, vol. 78, no. 7, pp. 9475-9491, 2022, doi: 10.1007/s11227-021-04274-6.
- [4] M. Callejas-Cuervo, F. Pineda-Rojas, and M. A. Vélez-Guerrero, "Measurement, Control, and Tracking Technologies Used in Soccer Players' Training: A Review," *Journal of Hunan University Natural Sciences*, vol. 50, no. 7, pp. 1-14, 2023, doi: 10.55463/issn.1674-2974.50.7.16.
- [5] P. D'Urso, L. De Giovanni, and T. Swartz, "Big data and data science in sport," *Annals of Operations Research*, vol. 325, no. 1, pp. 1-7, 2023, doi: 10.1007/s10479-023-05344-z.
- [6] W. Zhou, G. Yu, S. You, and Z. Wang, "An improved passing network for evaluating football team performance," *Applied Sciences*, vol. 13, no. 2, pp. 845-857, 2023, doi: 10.3390/app13020845.
- [7] R. Levoli, A. Gardini, and L. Palazzo, "The role of passing network indicators in modeling football outcomes: An application using Bayesian hierarchical models," *ASTA Advances in Statistical Analysis*, vol. 107, no. 1, pp. 153-175, 2023, doi: 10.1007/s10182-021-00411-x.
- [8] J. Leela, K. Rahaman, and D. M. G. Comission, "Analyzing the Efficiency of Passing Networks in Soccer," *Athens Journal of Sports*, vol. 11, no. 1, pp. 37-58, 2024, doi: 10.30958/ajspo.11-1-3.
- [9] D. Raabe, R. Nabben, and D. Memmert, "Graph representations for the analysis of multi-agent spatiotemporal sports data," *Applied Intelligence*, vol. 53, no. 4, pp. 3783-3803, 2023, doi: 10.1007/s10489-022-03631-z.
- [10] C. Yang and O. Kolbinger, "Equal Pass: Comparing Passing Networks of England's Top Women's and Men's Football Leagues," *European Journal of Sport Sciences*, vol. 3, no. 5, pp. 25-32, 2024, doi: 10.24018/ejsport.2024.3.5.201.
- [11] A. Vyas, A. Parnami, and M. R. Prusty, "Graph theory-based mathematical modeling and analysis to predict a football dream team," *Knowledge and Information Systems*, vol. 65, no. 4, pp. 1523-1547, 2023, doi: 10.1007/s10115-023-01849-y.
- [12] R. Alves, T. Sousa, V. Vaz, H. Sarmiento, P. Bradley, and G. Dias, "Analysis of the interaction and offensive network of the Portuguese national team at the 2016 European Football Championship," *Retos*, vol. 47, no. 1, pp. 35-42, 2023, doi: 10.47197/retos.v47.94621.
- [13] J. Yang, H. Ge, and Y. Cui, "An AI framework for counterattack detection and decision-making evaluation in football," *Journal of Big Data*, vol. 12, no. 1, pp. 91-108, 2025, doi: 10.1186/s40537-025-01128-3.
- [14] J. AlMulla, M. T. Islam, H. R. H. Al-Absi, and T. Alam, "SoccerNet: A Gated Recurrent Unit-based model to predict soccer match winners," *PLoS One*, vol. 18, no. 8, pp. 1-19, 2023, doi: 10.1371/journal.pone.0288933.
- [15] S. Zhou, "DeepRace: GRU-Based Sequence Modeling Framework for Marathon Performance Prediction," *Informatica*, vol. 49, no. 28, pp. 153-160, 2025, doi: 10.31449/inf.v49i28.8818.
- [16] C. Yang, P. Jin, and Y. Chen, "Leveraging graph neural networks and gate recurrent units for accurate and transparent prediction of baseball pitching speed," *Scientific Reports*, vol. 15, no. 1, pp. 7745-7756, 2025, doi: 10.1038/s41598-025-88284-x.
- [17] V. Armatas, S. Plakias, S. Drikos, and M. Mitrotasios, "Tactical Indicators and Situational Variables Affecting Goal-Scoring Opportunities in the UEFA Youth League 2023-2024," *Applied Sciences*, vol. 15, no. 8, pp. 4532-4547, 2025, doi: 10.3390/app15084532.
- [18] P. Eusebio, P. Prieto-González, and R. Marcelino, "Unlocking dynamics of goal-scoring: the showdown between direct and indirect transition goals across football leagues," *Biology of Sport*, vol. 42, no. 2, pp. 113-123, 2025, doi: 10.5114/biolsport.2025.142640.
- [19] B. De-la-Cruz-Torres, M. Navarro-Castro, and A. Ruiz-de-Alarcón-Quintero, "An Expected Goals on Target (xGOT) Model: Accounting for Goalkeeper Performance in Football," *Big Data and Cognitive Computing*, vol. 9, no. 3, pp. 64-76, 2025, doi: 10.3390/bdcc9030064.
- [20] J. Ball, M. Huynh, and M. C. Varley, "Comparing player rating systems as a metric for assessing individual performance in soccer," *Journal of Sports Sciences*, vol. 43, no. 7, pp. 676-686, 2025, doi: 10.1080/02640414.2025.2471208.
- [21] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Computing and Applications*, vol. 35, no. 23, pp. 17095-17112, 2023, doi: 10.1007/s00521-023-08568-z.
- [22] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards understanding convergence and generalization of AdamW," *IEEE transactions on pattern analysis and machine intelligence*, vol. 46, no. 9, pp. 6486-6493, 2024, doi: 10.1109/TPAMI.2024.3382294.
- [23] Y. Shao, J. Wang, H. Sun, H. Yu, L. Xing, Q. Zhao, et al., "An improved BGE-Adam optimization algorithm based on entropy weighting and adaptive gradient strategy," *Symmetry*, vol. 16, no. 5, pp. 623-638, 2024, doi: 10.3390/sym16050623.
- [24] C. Zhang, Y. Shao, H. Sun, L. Xing, Q. Zhao, and L. Zhang, "The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms," *Mathematical Biosciences and Engineering*, vol. 21, no. 1, pp. 1270-1285, 2024, doi: 10.3934/mbe.2024054.