TS-CVAE: Accurate Player Movement Predictions by Leveraging Team and Opponent Dynamics for Doubles Badminton

Anonymous submission

Abstract

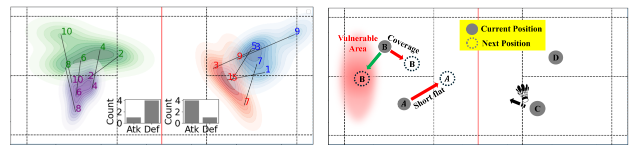
With the advancement of AI, strategic analysis in team sports has become increasingly valuable. In doubles badminton, the challenge lies in accurately predicting player movements in a fast-paced, dynamic environment. To tackle this, we introduce the novel TS-CVAE specifically designed for doubles badminton. TS-CVAE incorporates Team GAT, which leverages team influence graph over a few strokes to capture rapidly changing team strategies, and Opponent GAT, which holistically analyzes interactions between opposing players. Additionally, a new data augmentation module, SOSA, enhances the understanding of player positioning and strategy by incorporating singles data. Experimental results on real doubles badminton data show that TS-CVAE outperforms state-of-the-art sport forecasting models across multiple evaluation metrics. Visualized results also confirm that TS-CVAE's predictions closely align with the ground truth.

1 Introduction

In the recent decade, AI models, particularly for predicting movements, have brought significant advantages to team sports. AI is a game changer that not only effectively alleviates human effort in analysis but also offers the opportunity to uncover complex moving patterns, which may provide coaches and players valuable insights to refine techniques and tactics (Chen et al. 2023; Sung et al. 2024).

Team sports can be categorized into net sports (e.g., badminton and tennis) and full-court games (e.g., basketball and football), with moving patterns and rules significantly differentiate them. In full-court sports, every player can move and hustle freely across the whole field, and there are usually specific zones to score. In net sports, teams are divided by a net, and the scoring zones are typically anywhere on the opponent's side. A line of research (Graber and Schwing 2020; Li et al. 2021) has proposed AI models to analyze multi-player dynamics in full-court games. However, net sports, including badminton, have not yet drawn enough attention.

This paper focuses on doubles badminton. Although there is emerging literature on singles badminton (Wang et al. 2021, 2022; Chang, Wang, and Peng 2023; Wang et al. 2024), generalizing it to doubles scenario is not trivial since the concept of *teammates* does not exist in singles. Applying singles models to doubles scenario may overlook the importance of team chemistry and strategy during learning and



(a) Player position distribution. (b) TS-CVAE usage example.

Figure 1: Examples of doubles badminton scenarios.

forecasting. Few studies were specifically proposed for doubles badminton, and those that did (Sung et al. 2024) actually failed to finely formulate team strategy and opponent competitions, which are crucial factors in doubles games.

To illustrate an example of team strategy in doubles, Figure 1(a) presents a density map estimating player positions during a rally using Kernel Density Estimation (KDE) (Parzen 1962). KDE estimates the probability density function of a random variable, with darker areas indicating higher frequencies of player positions. Contours represent constant density levels, and colors distinguish every player, with numbers denoting the order of strokes. Adjacent histograms show the count of offensive and defensive plays. The figure reveals that different strategies result in distinct player movements and positions. Defensive strategies position players on the left court (green and purple) in parallel to cover a wider area, while offensive strategies position players on the right court (red and blue) vertically to the net to maximize their opportunities and effectiveness in attacking. In contrast, players in singles games do not have such tactical considerations and positioning with teammates.

Given the input of four-player movements in doubles, the goal of this paper is to forecast the movements of all four players, shot types, and players executing each shot. The problem also comprehensively incorporates team strategies, competitive reactions against opponents, and sequential relationships involving both teammates and opponents. Two key factors need to be addressed:

1. **Team Strategy.** Since the pace of badminton is fast, doubles players must rapidly pursue the shuttle while simultaneously covering their teammates, leading to significant changes in player locations and team strategies from stroke to stroke. Failing to consider positional impacts from teammates may result in inaccurate predictions of both players' locations.
2. **Opponent Influence.** Opponent reactions strongly affect

team decision-making. For instance, teams may capitalize on taking front-and-back positioning to execute many smashes and drives when their opponent makes a mistake by raising the shuttle high. However, analyzing the tactical influence among all strokes against each other, which may form a complete influence graph, could be challenging due to scalability issues.

To address these factors, we propose a novel model, namely *Team Strategy-aware CVAE* (*TS-CVAE*), built with Conditional Variational Autoencoder (CVAE) due to its effectiveness in generating additional samples to improve sport trajectory predictions (Graber and Schwing 2020; Li et al. 2021; Sung et al. 2024). The followings distinguish TS-CVAE from literature. For the team strategy, we introduce a new structure, named *team influence graph*, to capture the rapid and dynamic interactions between teammates over a few strokes. Compared to analyzing long stroke sequences, focusing on fewer strokes more accurately reflects the quick reactions required in doubles badminton, as demonstrated by our experiments. Using these team influence graphs, *Team GAT* is employed to compute graph attentions, which quantify and interpret the chemistry between teammates.

For the opponent influence, we propose the novel *Opponent GAT* to explore interactions among opposing players. To avoid the scalability issues, Opponent GAT first conducts Temporal Convolution Network (TCN) (Bai, Kolter, and Koltun 2018) to extract temporal features of each individual player. Then, a Graph Attention Network (GAT) is applied to derive the impacts among the four sets of features representing each player. By modeling teammate and opponent dynamics, TS-CVAE can potentially be generalized to volleyball, tennis, table tennis, and other multi-player net sports. Additionally, we propose a new data augmentation module, *Singles Location and Shot Type Auto-encoder* (*SOSA*), to extract similar knowledge from relatively abundant singles data and thereby remedy the data scarcity issue of doubles.

Figure 1(b) illustrates a scenario of how coaches and players can exploit TS-CVAE. Solid and hollow nodes respectively show the current and the next positions of players. Player C is hitting a short flat to the left. Red arrows depict the actual reactions of the left team, where player A intercepts the shuttle and player B moves toward the middle-front to cover. On the other hand, the green arrow shows TS-CVAE's predictions, in which player B moves back instead. Unlike the reality, the predicted strategy could be better since it does not expose a huge red vulnerable area at the back court by considering teammate movements. Therefore, TS-CVAE helps identify discrepancies, allowing the left team to review and adjust their strategies. For the right team, recognizing the opponents' potential misalignment provides insights for designing counter-strategies.

Experimental results and visualizations on real doubles badminton data show that TS-CVAE outperforms state-of-the-art sports forecasting models. Our codes can be found on the anonymous link https://tinyurl.com/drj7pkmu for reproducibility. Our contributions are concluded as follows:

- This paper identifies fine-grained dynamics of teammates and opponents in doubles badminton games, respectively. A new model, TS-CVAE, is proposed to address them.
- Novel modules, Team GAT and Opponent GAT, analyze team strategies and opponent influences. A new data augmentation module, SOSA, alleviates data scarcity.
- Experimental results on real datasets manifest that TS-CVAE effectively predicts player locations, shot types, and hitting players, outperforming baselines by at least 5.65% in terms of ADE. Visualized results confirm that TS-CVAE's predictions closely align with real data.

2 Related Works

Ball Sport AI. AI models have been playing a crucial role in ball sports, including net sports and full-court sports (Yanai et al. 2022; Chen et al. 2023), recently. In regards of badminton games, most literature (Wang et al. 2021, 2022; Chang, Wang, and Peng 2023; Wang et al. 2024) focused on singles badminton. However, generalizing these models to doubles badminton is not trivial since there exists no concept of teammates in singles. Little research studied doubles badminton. MoCVAE (Sung et al. 2024) presented a CVAE-based architecture to analyze multiple objectives jointly, but failed to formulate teammate interactions and opponent competitions. Hence, there is a need for novel models that utilize team strategies inherent in doubles badminton.

Generative Models. CVAE models generating diverse possible movements are commonly used to improve trajectory forecasting (Graber and Schwing 2020; Li et al. 2021) and sport dynamics (Sung et al. 2024). dNRI (Graber and Schwing 2020) inferred interpretable entity relations with deep encoder and decoder. Grin (Li et al. 2021) employed additional graph neural networks to formulate agent movements. Graph-TERN (Bae and Jeon 2023) used control point prediction and multi-relational weighted graphs to improve trajectory prediction. Nevertheless, these models did not incorporate the complex dynamics in doubles badminton.

3 TS-CVAE

The proposed Team Strategy-aware CVAE (TS-CVAE) for movement predictions in doubles badminton is illustrated in Figure 2. The left grey shaded area presents SOSA, introducing extra features as a data augmentation module. The right shows a new CVAE structure that incorporates Team GAT and Opponent GAT to finely analyze complex dynamics within and across teams, respectively. Preliminaries and notations are defined in Section 3.1. TS-CVAE is introduced in Sections 3.2 and 3.3. Section 3.4 proposes SOSA.

3.1 Preliminaries and Notations

Let a rally be defined as a sequence $R = \langle (l^t, s^t, p^t) \rangle_{t=1}^{T}$. The numbering of strokes and the maximum number are denoted as t and T, respectively. Triplet (l^t, s^t, p^t) records the condition at stroke t. $l^t \in \mathbb{R}^{2\times4}$ denotes the two-dimensional locations of the four players. $s^t \in \mathbb{R}^{L_s}$ is the one-hot encoding of the shot type occurring at stroke t, where L_s denotes the number of shot types, which is 16 in this paper. $p^t \in \mathbb{R}^4$ is the one-hot encoding of the player hitting the ball. A training instance divides rallies into observation rally sequences $R_{obs} = \langle (l^t, s^t, p^t) \rangle_{t=1}^{T_{obs}}$ and prediction

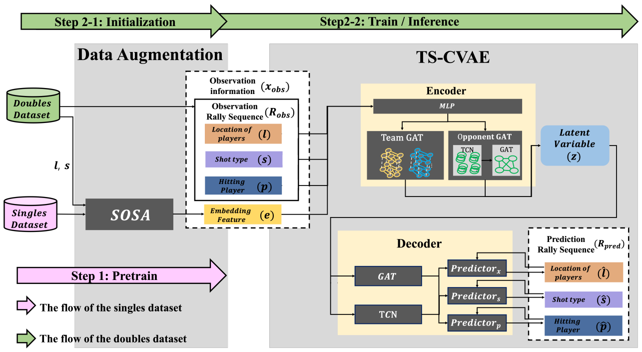


Figure 2: Illustration of TS-CVAE.

rally sequences $R_{pred} = \langle (l^t, s^t, p^t) \rangle_{t=T_{obs}+1}^{T}$, where T_{obs} denotes the number of strokes for observations. Given the observation rally sequence R_{obs}, TS-CVAE aims to forecast the predicted rally sequence R_{pred}.

3.2 Encoder and Prior

The movement of players in doubles badminton includes various twists and turns, making it challenging to predict. Inspired by (Sung et al. 2024), TS-CVAE aims to improve accuracy by employing a CVAE architecture, which generates diverse player movements under various conditions. Nevertheless, TS-CVAE introduces two novel modules, *Team GAT* and *Opponent GAT*, to learn fine-grained behavior of teammates and opponents, respectively. These modules distinguish TS-CVAE from literature.

As shown by the green arrows at the top of Figure 2, the input includes two parts: original observation rally sequence R_{obs} in doubles and its processed version (e), enriched by SOSA with singles knowledge. An encoder derives the latent variable z with Team GAT and Opponent GAT. A decoder further predicts all player locations, shot types, and hitting players based on the latent variable z.

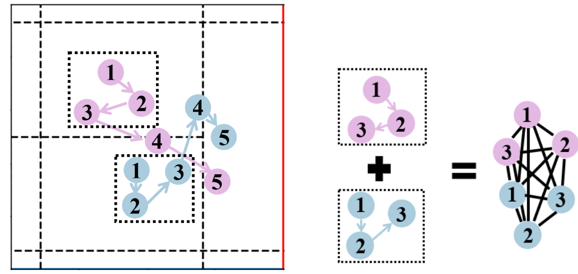
Encoder The goal of the encoder is to extract the posterior distribution of a latent variable z^E by breaking down team strategies and opponent competitions with Team GAT and Opponent GAT in doubles badminton. The input includes both observation and prediction rallies, which is a common setting in CVAE-based models (Graber and Schwing 2020; Li et al. 2021), and is further processed via an MLP layer:

$$Comb^E = MLP(\chi_{obs}),$$

where χ_{obs} denotes observed information by concatenating R_{obs} with additional features e (detail in Section 3.4). $Comb^E$ represents the output of the MLP layer.

Team GAT is designed to capture temporal relations and interactive dynamics between teammates. Specifically, teammates are supposed to cover each other or set up attacks via their positioning. As a result, team strategies significantly influence their movements and positions from stroke to stroke. Team GAT first constructs *team influence graphs* for each team to represent their moving patterns, as an example illustrated in Figure 3.

In Figure 3(a), the color of nodes and edges identifies each player. Nodes represent player locations and numbers correspond to the stroke sequence. Edges connect consecutive



(a) Original stroke sequence. (b) Team influence graph.

Figure 3: Example of constructing team influence graphs.

player locations of a player. A black dashed box represents a sliding window of length three, which focuses on the previous stroke, the current stroke, and the next stroke. For instance, the box of the blue player including her strokes one, two, and three observes the context of the second stroke with its preceding and following strokes. The window length is set to three because it is a sweet spot of capturing fast changing doubles game, which will be examined in experiments.

Figure 3(b) further shows the process of merging the nodes in the two dashed boxes of teammates into a team influence graph. Each edge in the graph is weighted by a stroke-wide influence, reflecting how each stroke impacts each other, thereby interpreting the team strategy. The influence is derived by graph attentions with GAT and $Comb^E$:

$$TeamGATout^E = GAT^E_{team}(Comb^E),$$

where $TeamGATout^E$ denotes the output of Team GAT. Hence, stronger influences results in greater attention.

On the other hand, it is essential for players to consider their opponents' movements and positions while playing to maximize scoring effectiveness. To account for opponents' strategies, an Opponent GAT module is introduced to learn counteracting behaviors. Opponent GAT first appllies a TCN to each player to capture stroke-by-stroke dynamics:

$$OppoTCNout^E = TCN^E_{oppo}(Comb^E),$$

where $OppTCNout^E$ denotes the output of TCN, which polishes $Comb^E$ with temporal influence of each individual. TCN is chosen since its non-recursive structure is less likely to encounter vanishing or exploding gradients during training (Bai, Kolter, and Koltun 2018).

Then, to capture the collective and competitive dynamics among all players, a GAT layer is used after the TCN layer:

$$OppoGATout^E = GAT^E_{oppo}(OppoTCNout^E),$$

where $OppoGATout^E$ denotes the output of the Opponent GAT. This GAT layer attends the sequential embedding $OppoTCNout^E$ based on the interactions between players, modeled as a complete graph. An alternative approach would be using a complete bipartite graph, where teammates are disconnected but opponents remain connected. However, since not only individual players but also team may be affected by opponents, we instead connect the teammates, making it a complete graph.

Note that Team GAT and Opponent GAT conduct different structures for specific reasons. Team GAT focuses on capturing rapidly changing tactical decisions within a team, using complete graphs with fewer strokes to emphasize short-term impacts. In contrast, Opponent GAT aims to understand the broader influences disrupting interactions between players, requiring analysis across all strokes. Attending complete graphs with all strokes and players could overwhelm the process like Team GAT, so Opponent GAT uses a variation to avoid significant computational overhead.

Finally, assume that the latent variable z^E follows a multivariate normal distribution. The posterior is designed using the reparameterization trick (Kingma and Welling 2014), ensuring differentiability during optimization:

$$\mu_{z^E}, \sigma_{z^E} = FC(TeamGATout^E, OppoGATout^E),$$

where FC denotes a fully connected layer. μ_{z^E} and σ_{z^E} are the mean and the standard deviation of the posterior distribution $\mathcal{N}(z^E|\mu_{z^E}, \sigma_{z^E})$, respectively. The posterior distribution is derived from the outputs of Team GAT and Opponent GAT, which depict the interactive dynamics.

Prior The prior aims to approximate z^E, derived by the encoder, by generating another latent variable, denoted by z^P, based solely on the observation rally sequences, excluding the prediction rally sequences. Let $\mathcal{N}(z^P|\mu_{z^P}, \sigma_{z^P})$ denote the prior distribution, where μ_{z^P} and σ_{z^P} are the mean and the standard deviation. To obtain μ_{z^P} and σ_{z^P}, following CVAE-based models, the prior conducts the same structure as the encoder to maintain implementation consistency.

3.3 Decoder and Loss Function

Decoder The decoder aims at forecasting future strokes by decoding latent variable z^P with TCN and GAT for time domain and interactive dynamics features, respectively. Existing literature of doubles badminton conducted a vanilla decoder (Sung et al. 2024), failing to retrieve these essential factors. The bottom of Figure 2 illustrates our novel design of the decoder, which is formulated as follows:

$$\begin{aligned} DecTCNout &= TCN^D(FC^D(z^P)) \\ DecGATout &= GAT^D(FC^D(z^P)) \\ Dout &= Concat(DecTCNout, DecGATout), \end{aligned}$$

where FC^D denotes a fully connected layer. $DecTCNout$ and $DecGATout$ denote the decoded results of TCN and GAT, respectively. TS-CVAE employs TCN and GAT separately rather than in a cascade to improve efficiency while reducing dependency on processing order. Altogether, $Dout$ concatenates both decoded features for further predictions.

Afterwards, the decoder predicts all four player locations $\hat{l}^{t+1}$, the shot type $\hat{s}^{t+1}$, and the hitting players $\hat{p}^{t+1}$ of next stroke $t+1$ based on $Dout$ as follows:

$$\begin{aligned} \hat{l}^{t+1} &= FC_l(Concat(Dout^d, \hat{l}^t)) \\ \hat{s}^{t+1} &= FC_s(Concat(Dout^d, \hat{s}^t)) \\ \hat{p}^{t+1} &= FC_p(Concat(Dout^d, \hat{p}^t)), \end{aligned}$$

where FC are fully connected layers and subscripts l, s, and p denote player locations, shot types, and hitting players respectively. In addition to $Dout$ containing information from observation rally sequences, these predictors also consider the preceding prediction results (i.e., $\hat{l}^t$, $\hat{s}^t$, and $\hat{p}^t$) to align with the sequential nature inherent in doubles badminton.

To sum up, introducing Team GAT and Opponent GAT enables the encoder to finely read team strategies and opponent impacts hidden in doubles badminton. Their distinct structures offer multiple perspectives of strategies. Finally, TCN and GAT in the decoder provide a sophisticated understanding in temporal features and interactive dynamics, respectively. These innovations distinguish TS-CVAE.

Loss Function The loss function includes two parts. First, the Evidence Lower Bound (ELBO) (Kingma and Welling 2014) is used to maximize the negative log-likelihood while minimizing the Kullback-Leibler (KL) divergence between the posterior and the prior distribution:

$$\begin{aligned} ELBO = &E[log(p(R_{pred}|z^E, \chi_{obs})] \\ &- KL[q(z^E|\chi_{obs}, R_{pred})||p(z^P|\chi_{obs})]. \end{aligned}$$

Maximizing the negative log-likelihood ensures that the predictions derived from the latent variable (z^E) closely match the original input data. Minimizing the KL divergence aligns the prior distribution with the posterior distribution.

For the second part, cross-entropy $\mathcal{L}_{shot\_type}$ and $\mathcal{L}_{hitting\_player}$ respectively evaluate the shot type and the hitting player predictions:

$$\begin{aligned} \mathcal{L}_{shot\_type} &= -s \cdot log(\hat{s}) \\ \mathcal{L}_{hitting\_player} &= -p \cdot log(\hat{p}). \end{aligned}$$

Altogether, our goal is to minimize the total loss $\mathcal{L}_{total}$:

$$\mathcal{L}_{total} = min(-ELBO + \mathcal{L}_{shot\_type} + \mathcal{L}_{hitting\_player}).$$

Therefore, minimizing $\mathcal{L}_{total}$ enables TS-CVAE to learn the team strategies and opponent competitions in doubles badminton through a CVAE-based structure.

The inference time of TS-CVAE is about 2 seconds on average, which is sufficient in real-world usage. Note that in doubles badminton, coaches are not permitted to communicate with players during games until technical breaks at every 11 points or the end of a set, making real-time predictions unnecessary. Instead, TS-CVAE prioritizes refining prediction accuracy by using the available time effectively.

3.4 Singles Location and Shot Type Auto-encoder

In concern of data scarcity in doubles badminton, borrowing similar knowledge from relatively abundant singles badminton datasets (Chang, Wang, and Peng 2023; Wang et al. 2022; Huang et al. 2022) could be a feasible solution. Existing literature pretrained and initialized player location embeddings based on singles knowledge (Sung et al. 2024). However, we argue that this is not the best approach.

Specifically, we propose SOSA as a pretrained data augmentation module that integrates shot type information with player locations. The rationale is twofold. First, shot types are strongly correlated with player locations (Sung et al. 2024). For instance, a player positioned at the back court is more likely to opt for a smash than a short flat shot. Pretraining shot types alongside player locations may leverage

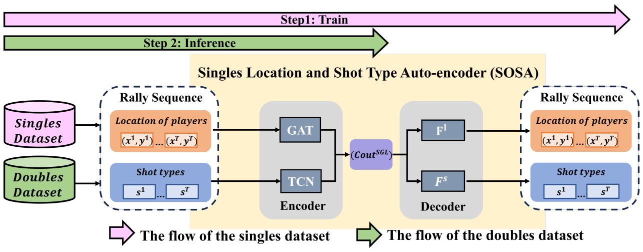


Figure 4: Illustration of SOSA.

these correlations to improve accuracy. Second, concatenating these augmented features with original doubles features is less likely to eliminate the latter in deep models, and thereby leads to better experimental results.

Figure 4 illustrates the auto-encoder structure of SOSA. The encoder takes player locations and shot types from singles data as input, and further exploits GAT and TCN to derive positional influence and temporal features, respectively:

$$Cout^{SGL} = Concat(GAT^{SGL}(l), TCN^{SGL}(s)).$$

Specifically, GAT attends a complete graph constructed by connecting the location of every stroke of a player within a rally. TCN analyses the temporal relations in shot types within a rally. The output of the encoder, $Cout^{SGL}$, concatenates the results of GAT and TCN and is then transformed into a low-dimensional vector with a fully connected layer.

The goal of the decoder is to rebuild the initial input from the encoded low-dimensional features. Two fully connected layers, F^l and F^s, are employed to reconstruct player locations and shot types, respectively:

$$\hat{l} = F^l(Cout^{SGL})$$
$$\hat{s} = F^s(Cout^{SGL}),$$

where the predicted player locations and shot types are denoted by $\hat{l}$ and $\hat{s}$, respectively. The auto-encoder loss function, AE_{loss}, for SOSA is designed as follows:

$$AE_{loss} = MSE(l, \hat{l}) + \lambda \cdot CE(s, \hat{s}),$$

where the first term conducts mean square error, MSE, to evaluate the predicted player locations $\hat{l}$, and the second term uses cross-entropy, CE, to evaluate the predicted shot types $\hat{s}$. Hyper parameter λ controls the weight to balance these two heterogeneous losses.

After training with singles data, only the encoder is retained for further use. Specifically, as shown in the middle of Figure 2, when doubles data is used as input, $e = Cout^{SGL}$ offers extra knowledge to TS-CVAE. As e is appended (rather than replacing) to R_{obs}, the original doubles features are preserved and remain visible during training.

In summary, we propose the novel SOSA to transfer knowledge of both shot types and player locations from singles to doubles with an auto-encoder structure. SOSA generates additional informative features to enhance TS-CVAE through data augmentation techniques.

4 Experiment

4.1 Experiment Setting

To evaluate TS-CVAE, we compare with state-of-the-art CVAE-based models for doubles badminton and full-court sports (e.g., basketball) models:

1. *MoCVAE* (Sung et al. 2024) utilized shot type and the hitter to predict the movement of players in doubles badminton without separating teams and opponents.
2. *dNRI* (Graber and Schwing 2020) uncovered the dynamic relations between players with a recursive model.
3. *Grin* (Li et al. 2021) utilized graph neural networks to incorporate interactive dynamics and agent intention.

To understand each baseline's behavior, we consider three observation and prediction stroke combinations, which are (11,1), (10,2), and (8,4), respectively. The default combination is (8,4) if not specified. Note that original dNRI and Grin did not predict shot types and hitters. To maintain a fair setting, we concatenate this two information as the input for them. Furthermore, they all use an MLP and two individual softmax layers to predict the shot type and the hitter.

To pretrain SOSA, 170 singles matches including 448 sets and 11,284 rallies are used (Huang et al. 2022). For the main task, 25 doubles matches including 58 sets and 2230 rallies are used. Shot types, positions of each player, and hitting players are labeled by (Huang et al. 2023). The number of shot type classes is 16. The player locations adopt a min-max normalization. The dataset is separated based on rallies, where 80% and 20% are for training and testing, respectively.

In TS-CVAE, the dimensions of the hidden states and the latent variables are 8 and 16, respectively. The learning rate is 10^{-3} and the batch size is 32. λ is set to 1. For SOSA, the learning rate is set to 5×10^{-4} and the batch size is 1024. All baselines are trained on a machine of AMD Ryzen 7 5700X 8-Core CPU and an Nvidia RTX 3060 Ti GPU.

Following baselines (Graber and Schwing 2020; Li et al. 2021; Sung et al. 2024), same evaluation metrics of player locations are used, specifically final displacement error (FDE) and average displacement error (ADE). ADE provides an overall assessment of movement performance by calculating the average squared error between the actual and predicted locations in x-y coordinates across all strokes. Conversely, FDE evaluates final movement performance by measuring the average squared error at the endpoint of each movement. Additionally, we use cross-entropy to evaluate shot types and player predictions.

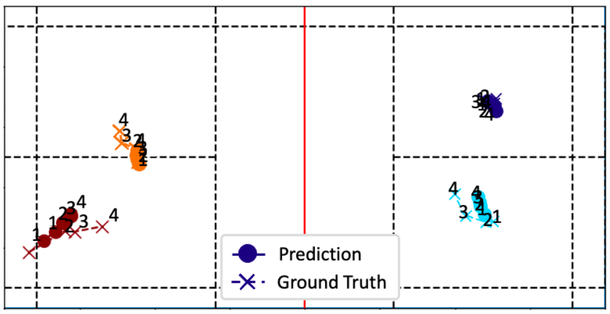
The performance of TS-CVAE against baselines is first studied in Section 4.2. Comprehensive sensitivity tests are presented in Section 4.3. Ablation studies examine effectiveness of each component of TS-CVAE in Section 4.4.

4.2 Performance Evaluation

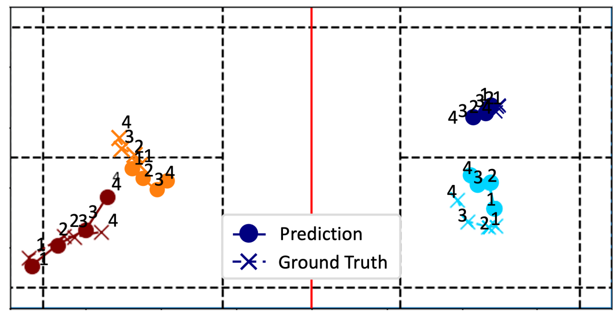
Table 1 compares each baseline in predicting player position, shot type, and hitting player. The results demonstrate that TS-CVAE outperforms all other models across most scenarios. Specifically, for an observation window of 8 strokes and a prediction of 4 strokes, TS-CVAE surpasses all baselines by at least 7.19% and 2.27% in terms of ADE and FDE, respectively. It also exceeds by 14.7% and 4.49% in terms of CE for shot types and hitting players, respectively. These results manifest that TS-CVAE effectively identifies player locations by introducing Team GAT and Opponent

Table 1: Prediction performance of each baseline.

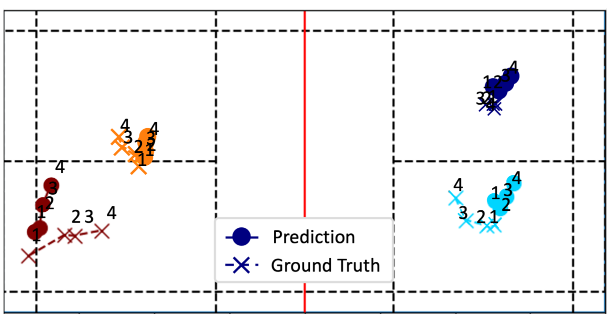
Model	11 strokes/ 1 stroke				10 strokes/ 2 strokes				8 strokes/ 4 strokes			
	ADE	FDE	Shot Type	Hitting Player	ADE	FDE	Shot Type	Hitting Player	ADE	FDE	Shot Type	Hitting Player
dNRI	0.110	0.110	3.175	2.546	0.128	0.158	3.593	2.887	0.160	0.208	3.653	2.978
Grin	**0.108**	**0.108**	3.899	3.955	0.124	0.152	4.339	3.807	0.153	0.176	3.907	3.567
MoCVAE	0.136	0.136	2.586	1.433	0.147	0.181	2.746	1.401	0.289	0.42	3.204	1.448
TS-CVAE	0.111	0.111	**2.437**	**1.348**	**0.117**	**0.140**	**2.676**	**1.389**	**0.142**	**0.172**	**2.733**	**1.383**



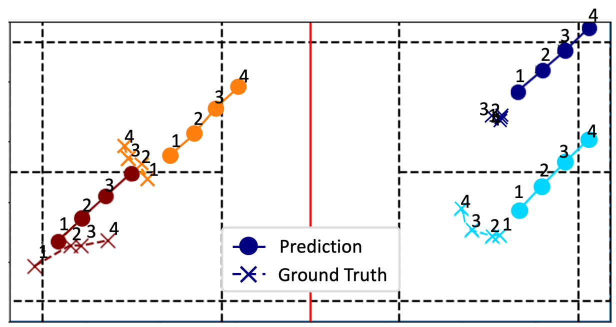
(a) TS-CVAE.



(b) dNRI.



(c) Grin.



(d) MoCVAE.

Figure 5: Visualized player movements of each baseline.

GAT to simultaneously analyze teammate and opponent dynamics. This approach gains key insights into doubles match positioning, further linking those insights to shot type and hitter predictions. While MoCVAE focuses on doubles badminton as well, it does not separate teammates and opponents, resulting in less accurate predictions. On the other hand, in the setting of 11 and 1 strokes, TS-CVAE only achieves comparable ADE and FDE to Grin because predicting one single stroke is relatively easy for full court models. However, TS-CVAE is more comprehensive in all crucial tasks of doubles badminton, as evident by at least 37.5% improvement on the other two predictions.

Figure 5 visualizes four consecutive predicted player locations by observing 8 previous strokes. Different colors indicate different players, and the numbers represent the order of strokes. The visualization result highlights that TS-CVAE in Figure 5(a) predicts player locations more accurately than other models. Specifically, the team on the right (blue and cyan) adopts a defensive strategy, so they remain parallel on their respective halves and hardly move. In contrast, Figure 5(b) shows that dNRI predicts the two players progressively moving closer, which contradicts typical defensive team strategies by reducing their coverage area. GRIN predicts that the red player aggressively moves toward the middle, as shown in Figure 5(c), which causes larger vulnerable areas on both their left-hand and right-hand sides. These observations demonstrate that TS-CVAE, by incorporating Team GAT, better reflects team strategies in practice. Figure 5(d) shows that all four players are predicted to move upper-right by MoCVAE, indicating that using a single MLP layer for decoding fails to correctly decode temporal and interactive features. This result underscores the necessity of employing both TCN and GAT in the decoder of TS-CVAE.

4.3 Sensitivity Tests

Different prediction strokes. Figure 6 shows the ADE of different prediction strokes of all baselines while fixing the observation setting at 10 strokes. The solid line presents the average results, and the shaded area shows the range between the maximum and the minimum. For TS-CVAE, the slope is less steep and has lower ADEs in most of the prediction strokes. Besides, the shaded area is relatively smaller. All these results manifest that TS-CVAE is robust in extending prediction strokes by accurately interpreting and forecasting game situations through the comprehensive knowledge extracted by Team GAT and Opponent GAT. Without factoring in these essentials, ADEs of other baseline models increase faster when prediction strokes increase.

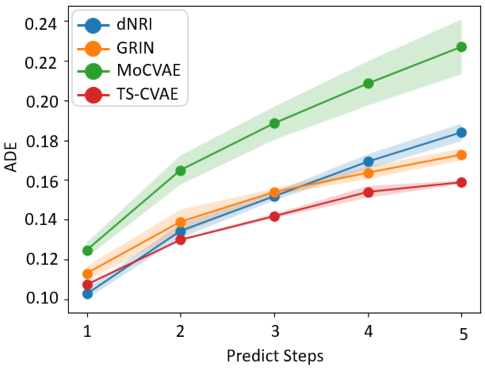
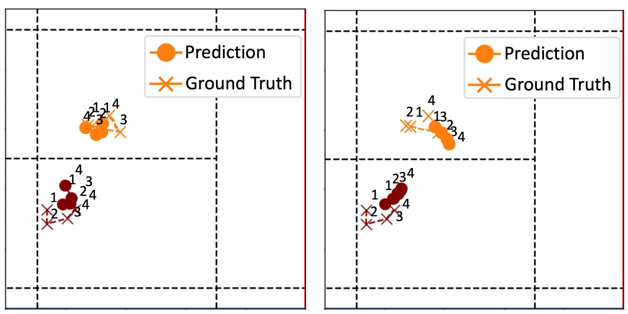


Figure 6: Performance of different prediction strokes.



(a) Using 3 strokes. (b) Using 7 strokes.

Figure 7: Visualization varying observation window length.

Team influence graphs in Team GAT. This subsection examines the sensitivity of the observation strokes for constructing team influence graphs in Team GAT, as shown in Figure 3. Table 2 shows ADE and FDE for different observation strokes. The results manifest that 3-stroke setting provides the best overall performance. The ADE increased by 4.9% for 2-stroke setting, suggesting that it is insufficient to capture team strategies. Although the 5 and 7-stroke settings are supposed to bring in more context, their results show at least a 3.5% degradation in terms of ADE. It is because the pace and content of doubles badminton change rapidly, requiring team strategies to react quickly. Including too many strokes actually brings noise instead of useful information.

Figure 7 respectively visualizes the outcomes of 3 and 7-stroke settings for the same rally. Predictions of both players in Figure 7(a) show more twists and turns, aligning with the ground truth. Figure 7(b) shows that incorporating too many strokes may eliminate twists and turns, causing overly smooth predictions that could mislead coaches and players while designing future strategies.

4.4 Ablation Tests

Five leave-one-out variants of TS-CVAE are compared:

- *W/o TCN in Decoder*: TS-CVAE excluding the TCN layer for decoding sequential patterns in decoder.
- *W/o Team GAT*: TS-CVAE excluding the Team GAT module which captures teammate interactions.
- *W/o Opponent GAT*: TS-CVAE excluding the Opponent GAT module which captures opponent influence.

Table 2: Performance on observation strokes in Team GAT.

strokes	2 strokes	3 strokes	5 strokes	7 strokes
ADE	0.15	**0.143**	0.148	0.149
FDE	0.182	**0.174**	0.18	0.182

Table 3: Ablation study on leave-one-out variants.

Task	Movement Forecasting		Shot Type	Hitting Player
	ADE	FDE	Cross Entropy	
w/o TCN in Decoder	0.150	0.182	2.742	1.412
w/o Team GAT	0.144	0.174	2.731	1.387
w/o Opponent GAT	0.143	0.176	2.751	1.386
w/o SOSA	0.144	0.176	2.760	1.391
w/ Initialization	0.146	0.177	2.752	1.39
TS-CVAE	0.142	0.172	2.733	1.383

- *W/o SOSA*: TS-CVAE without data augmentation.
- *W/ Initialization*: TS-CVAE replacing (instead of concatenating) shot type and player location embeddings with the results of SOSA.

Table 3 presents the prediction performance of each variant. Specifically, the removal of the TCN layer from the decoder (*w/o TCN in Decoder*) results in the worst prediction outcomes for player movement, with decreases of 5.33% and 5.49% in terms of ADE and FDE, respectively. The drop in performance indicates that temporal features are crucial in retrieving dynamics on courts. Removing Team GAT from the encoder (*w/o Team GAT*) leads to declines in player movement predictions by 1.41% and 1.16% for ADE and FDE, respectively, highlighting the importance of the Team GAT in capturing team strategies to better understand movement between teammates. On the other hand, *w/o Opponent GAT* increases CE for both shot types and hitters due to losing insight into opponents' positioning, which is supposed to identify vulnerable areas to attack.

Additionally, *w/o SOSA* results in the worst performance for shot type prediction, showing the effectiveness of SOSA in transferring shot type knowledge from singles to doubles. *w/ Initialization* shows worse ADE compared to *w/o SOSA*, suggesting that trajectories are more inaccurately predicted on average when original doubles data is excluded.

5 Conclusion

This paper recognizes various and crucial interactions among players in doubles badminton, specifically team strategies and opponent competitions. Accordingly, we propose the novel TS-CVAE to predict player locations, shot types, and hitting players from stroke to stroke, respectively. TS-CVAE includes two innovative modules: Team GAT and Opponent GAT. Team GAT copes with rapidly varying team strategies based on novel team influence graphs and Opponent GAT holistically analyzes opposing player dynamics. Furthermore, SOSA leverages data augmentation from singles data to improve predictions. Evaluations on real datasets and visualizations manifest that TS-CVAE outperforms state-of-the-art sport forecasting models.

References

Bae, I.; and Jeon, H.-G. 2023. A set of control points conditioned pedestrian trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 6155–6165.

Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Chang, K.-S.; Wang, W.-Y.; and Peng, W.-C. 2023. Where will players move next? dynamic graphs and hierarchical fusion for movement forecasting in badminton. In *AAAI*.

Chen, X.; Wang, W.-Y.; Hu, Z.; Chou, C.; Hoang, L.; Jin, K.; Liu, M.; Brantingham, P. J.; and Wang, W. 2023. Professional Basketball Player Behavior Synthesis via Planning with Diffusion. arXiv:2306.04090.

Graber, C.; and Schwing, A. G. 2020. Dynamic neural relational inference. In *CVPR*.

Huang, Y.-H.; Huang, Y.-C.; Lee, H. S.; İk, T.-U.; and Wang, C.-C. 2022. S2-Labeling: Shot-By-Shot Microscopic Badminton Singles Tactical Dataset. In *2022 23rd Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 1–6.

Huang, Y.-H.; Sung, P.-C.; Huang, Y.-C.; İk, T.-U.; and Huang, J.-L. 2023. Shot-by-shot technical data collection for badminton doubles games. In *APNOMS*.

Kingma, D. P.; and Welling, M. 2014. Auto-encoding variational bayes. In *ICLR*.

Li, L.; Yao, J.; Wenliang, L.; He, T.; Xiao, T.; Yan, J.; Wipf, D.; and Zhang, Z. 2021. Grin: Generative relation and intention network for multi-agent trajectory prediction. In *NeurIPS*.

Parzen, E. 1962. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3): 1065–1076.

Sung, P.-C.; Lai, H.-C.; Jhang, G.-Y.; İk, T.-U.; Wang, C.-C.; and Huang, J.-L. 2024. MoCVAE: Movement Prediction by A Conditional Variational Autoencoder for Doubles Badminton. In *2024 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 40–47. IEEE.

Wang, K.-D.; Wang, W.-Y.; Hsieh, P.-C.; and Peng, W.-C. 2024. Offline Imitation of Badminton Player Behavior via Experiential Contexts and Brownian Motion. arXiv:2403.12406.

Wang, W.-Y.; Chan, T.-F.; Yang, H.-K.; Wang, C.-C.; Fan, Y.-C.; and Peng, W.-C. 2021. Exploring the Long Short-Term Dependencies to Infer Shot Influence in Badminton Matches. In *2021 IEEE International Conference on Data Mining (ICDM)*, 1397–1402.

Wang, W.-Y.; Shuai, H.-H.; Chang, K.-S.; and Peng, W.-C. 2022. Shuttlenet: Position-aware fusion of rally progress and player styles for stroke forecasting in badminton. In *AAAI*.

Yanai, C.; Solomon, A.; Katz, G.; Shapira, B.; and Rokach, L. 2022. Q-ball: Modeling basketball games using deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 8806–8813.