



HAL
open science

Team Dynamics in DotA2 through Attention Mechanism

Alexis Mortelier, Sébastien Bougleux, François Rioult

► **To cite this version:**

Alexis Mortelier, Sébastien Bougleux, François Rioult. Team Dynamics in DotA2 through Attention Mechanism. Machine Learning for Sports Analytics, Sep 2024, Vilnius, Lithuania. hal-04695698

HAL Id: hal-04695698

<https://hal.science/hal-04695698>

Submitted on 12 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Team Dynamics in DotA2 through Attention Mechanism

Alexis Mortelier¹, Sébastien Bougleux¹, and François Rioult¹

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

Abstract. We analyse team dynamics in the popular e-sports game DotA2 using an approach that combines convolutional and LSTM networks with a *feature-temporal* attention mechanism. Our goal is to identify strategic behaviours that lead to successful goals, such as scoring kills, during World Championship matches. Each team’s formation is represented by a polygon, which feeds an RNN that learns *kill* events from this polygon under its area, diameter, and moments around the centroid. By exploiting the attention mechanism, our network highlights the most relevant features at each time step, providing insights into strategic team movements and formations. Our results demonstrate the effectiveness of our approach in capturing critical dynamics that influence the outcome of engagements in DotA2.

Keywords: DotA2 · e-sport · deep learning · CNN-LSTM · attention · multivariate time series

1 Introduction

The e-sport phenomenon has emerged as a new era of sport, highlighting the opportunities it offers for sports scientists [13]. The way in which traditional sports and e-sport interact, particularly with regard to audiences, sponsors, revenues, and other economic and socio-cultural aspects, implies that e-sport fits into the wider landscape of professional sport and can benefit from the experience and infrastructure of traditional sports [5].

We are here mostly interested in analysing collective behavior by the way of a team of players. Data analysis can indeed be used to study the performance of the team as a whole, as well as that of its opponents. Analyzing historical match data can provide valuable insights into game trends, player and team performance, strategies used, and so on.

In DotA2, the *events* that punctuate the match occur when the heroes embodied by the players are temporarily neutralized, referred to as death then rebirth or *lifeline*. These events determine the score, which can be analyzed using machine learning techniques to determine the reasons behind the occurrence of an event, in terms, for example, of the geometric configuration of the set of teammates. This involves analyzing the geometric configurations that a network detects as emblematic of an event occurrence.

Attention mechanisms in recurrent networks are recognized for their ability to enrich the contextualization of a decoder, by learning a complementary path of temporal



Fig. 1: DotA2 map.

dependence. Our aim is to show that the analysis of attention weights offers a strategic perspective for the study of DotA2 matches.

The interpretation of the attention mechanism to obtain explanations for the decisions made by the network is a matter of debate. The interested reader may refer to [3], which gives a broad synthesis on the subject. Thus, attention would not provide an explanation and the use of Total Variation Distance (TVD), Jensen-Shannon Divergence (JSD) or Kullback Leibler divergence should be preferred [15] as attention is not correlated with more traditional gradient-based measures of attribute importance [33]. Experiments show that mixing attention weights does not affect the final decision. This statement is qualified by [35], which indicates that attention provides *one* explanation, but not *the only one*. The use of attention as a flux could also lead to the emergence of explanations [1].

Here, we are content to draw out *interpretations* of the weights of the attention layer of an RNN (Recurrent Neural Network), trained to predict the occurrence of a death event from the geometric features of the players' polygon. After a quick presentation of DotA2 in section 2, section 3 relates the state of the art, section 4 describes our process. The article concludes with the presentation of experiments in section 5.

2 What is DotA2 ?

DotA2, or Defense of the Ancients 2, is a MOBA (Multiplayer Online Battle Arena) video game developed by Valve Corporation. Two teams of five players compete to destroy the opposing team's Ancient, located in their base. Figure 1 shows a representation of the map of DotA2. Each player chooses a hero from a vast selection. Each hero has unique skills and different roles (such as support, carry, tank, etc.). The map is symmetrical, with three main lanes (top, mid, bottom) bordered by jungles and a river

linking the elders of each team. The heroes compete in three different lines, based on the principle of invasion sports like rugby and American football.

Team coordination and strategy are crucial : players must plan their attacks, defend their towers, and collectively take on objectives. DotA2 is very popular in esports competitions, with tournaments like The International ¹ offering millions of dollars in prizes. DotA2 combines real-time action, strategy and role-playing, requiring both individual skills and strong team collaboration to succeed.

3 Related work

The literature survey shows numerous contributions to sports analysis using RNNs with attention mechanisms, possibly on geometric features.

The popularity of video games in general and e-sports in particular has led to a number of articles studying the mechanics of these games, using a data science approach. MOBA games are very relevant for comparative studies because they offer a complex but relatively formatted game, both in terms of the playground and the way the games are played. Themes range from pick recommendation [32,28], studies of individual behaviour [26,4] to game summaries [2,23], outcome prediction [6], the use of geometric features and RNN [20].

The chronological nature of a video game match and the popularity of neural network methods are leading to a significant use of recurrent neural network techniques, for example to predict the outcome of the match [22,27,16]. This type of network can help to understand the strategic preparation phases of the game (pick/ban of heroes) [37]. As the game progresses, an important task is the recommendation of equipment items, supported by an RNN in [36] that models player interest from historical match sequences. A RNN can also predict a player's performance based on previous performances [34].

The theme of geometric features on trajectories is not new. The analysis of a collective practice on a given field naturally lends itself to studies of the trajectories and geometry of player organizations. Thus, for football, a work similar to the one presented in this article is proposed in [10], with the aim of predicting the outcome of the match. For DotA2, the geometric features allow to correctly predict the outcome of a match and this relevance is confirmed on 6 billion matches of League Of Legends (another instance of MOBA) via a RandomForest system with an F-measure of 92%. Regularities in the trajectories, discriminating in winning of the match, could also be extracted [6]. Behavior on a MOBA court is indeed crucial and differences between levels of play are revealed by positional features: change of zone, distance between players, which allows a grouping of matches according to the similarity of movements [9].

Using RNNs carefully to learn models from trajectories is a natural task for time series data from sporting activities. In particular, in the sports domain, there are many approaches proposing models to predict or complete trajectories [31,39,19,11], model sequences to predict the outcome of an event (e.g. a basketball shot [25]) or the outcome of a match [38]. RNNs are also used to analyze complex spatio-temporal dependencies in order to recognize individual events or detect group activities [18,12,24].

¹ see <https://www.dota2.com/esports/> for this year's edition.

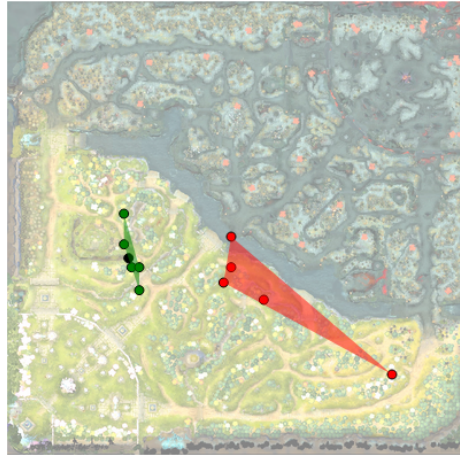


Fig. 2: Polygon visualization (one for each team).

Attention mechanism in RNN is useful for capturing temporal dependencies. So it's not surprising to find many uses for attention in sports data analysis, for example for movement prediction [29] or detecting similarity between movements [17].

Attention contributes to the design of the network to be learned, a classic task in machine learning. In this article, however, we interpret attention to analyze sports situations. This approach, also found in [40] to identify relevant portions of trajectories or understand differences between classes and determine correlations between variables and labels, is unusual in the literature.

4 Method

We start by explaining how the spatial configuration of the players is considered by the neural network. Its architecture is then described, with a focus on the attention layer and its analysis for tactical purposes.

4.1 Polygon features

DotA2 provides a record of game activity in the form of a *replay* file, which can be retrieved after each game. Hero movement data is then available, and death events for each hero are recorded. For the 258 matches of one of the game's major tournaments, The International 2018, we therefore have the temporal evolution of players' positions and their lifelines.

One objective would be to train a network to predict the lifelines of all the players based on their movements, on the assumption that this network will carry knowledge about the relationships between field occupation and lifelines. A single network would be common to all matches and would require comparable data. If fed with all ten trajectories, it would treat the first player's trajectory identically across matches. However,

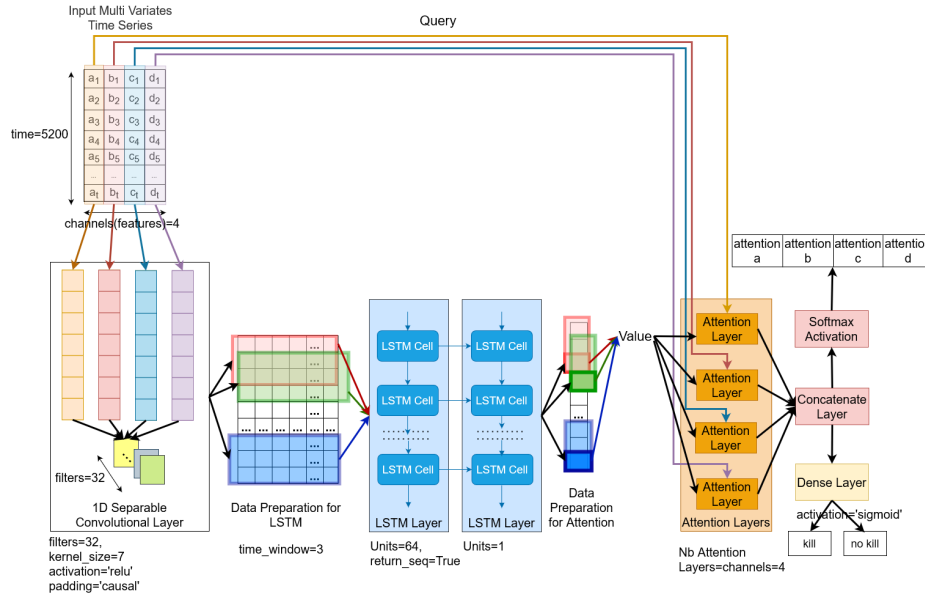


Fig. 3: Network architecture.

the first player varies in playing style, hero, and team composition from match to match. The order of players also depends on their position on the pitch – in DotA2, the *Radiant* team is on the lower half, including the first five players, while the *Dire* team is on the upper half, without exchanging positions at half-time.

In order to feed the network with comparable feature from match to match, we therefore consider, for each team and each time step in the sequence, the features of the polygon defined by the convex hull of all the players in a team, as illustrated in Figure 2, rather than individual trajectories. For each polygon, we calculated :

- *area*: the polygon surface;
- *diameter*: the maximum distance between two players;
- *grouping*: average of distances to the barycenter;
- *inertia*: second-order moment of the distances to the barycenter.

These simple measurements are a fine reflection of the spatial configuration of the players and their ability to carry out actions in a grouping configuration. For example, a high *diameter* indicates high dispersion (at least one player is far from the others), while *grouping* capacity is a clue of the average distance from the barycenter. *Inertia* is relative to the amplitude of variation in *grouping* capacity. We have already shown that these geometric features are sufficient to estimate the complexity of this game [21]: individual trajectories do not need to be observed.

4.2 Network architecture

Figure 3 shows a representation of the neural network architecture. Given a match, the input consists of a sequence of t values per polygonal feature for one team, represented by a matrix of size $t \times f$, with f the number of input features (4 in our case). The objective is to predict whether all the players on the opposing team are alive (*no kill*) or not (*kill*). This is a binary classification problem. In order to decide, the network first extract spatio-temporal information from the input features by a 1D depthwise separable convolution layer [7] followed by two LSTM layers [14], reducing the input sequence of $f = 4$ features to an hidden sequence of 1 feature per time step. Then, $f = 4$ independent dot-attention mechanisms [30] are applied on this sequence (*key* and *value*) to focus on information important for the final prediction, according to each input feature (a *query*) at each time. The resulting 4 feature sequences, one per input feature, are concatenated, and fed to a dense layer with bias and a sigmoid activation function, which produces the binary prediction for each time step of the input sequence. We briefly describe each main component of the network.

The 1D depthwise separable convolution layer applies a separate 1D temporal convolution per input feature sequence with a kernel of size 7, followed by 32 independent pointwise convolutions (1×1 convolutions) to learn correlations between features, and the ReLu activation function. The output of this layer is a $t \times 32$ feature matrix.

Then, both short and long term feature dependencies are captured by two LSTM layers. To this end, the sequence of 32 features is decomposed into overlying sub-sequences of 60 time steps, giving $t/60$ sub-sequences of 32 features. Each sub-sequence is fed to a first LSTM layer composed of 60 cells (one per time steps), a cell being composed of 64 units to produce a 60×64 intermediate feature matrix. The $t/60$ intermediate feature matrices are then fed to a second LSTM layer, producing $t/60$ feature matrices of size 60×1 . Then, the last time step of the $t/60$ sub-sequences are concatenated into a feature sequence of length t ($t \times 1$ matrix), denoted H , which is used by the attention mechanism.

Attention allows to relate the input sequence to output sequence H of the LSTM, capturing the specific influences of each feature. To this end, the dot-attention mechanism calculates similarity scores between the input sequence (*query*) and H (*keys*), with a dot-product followed by a softmax, to weight H (*value*), for the four attention block independently. This enables the network to focus on the most relevant parts of the sequence for each feature. The concatenation of all attention layers is normalized with a softmax and leads to a *spatio-temporal attention*, offering a detailed view of the importance of each feature at each time step, enabling a fine-grained and nuanced analysis of the impact of each feature on the overall prediction.

5 Experiments

5.1 Protocol

The dataset was constructed from the *replays* of the 258 DotA2 matches played during The International 2018¹. It contains a temporal sequence of the 4 polygon features

Table 1: Network performances for the test set.

	with attention	without attention
accuracy	0.87	0.74
precision	0.86	0.74
recall	0.87	0.73
f1 score	0.86	0.73

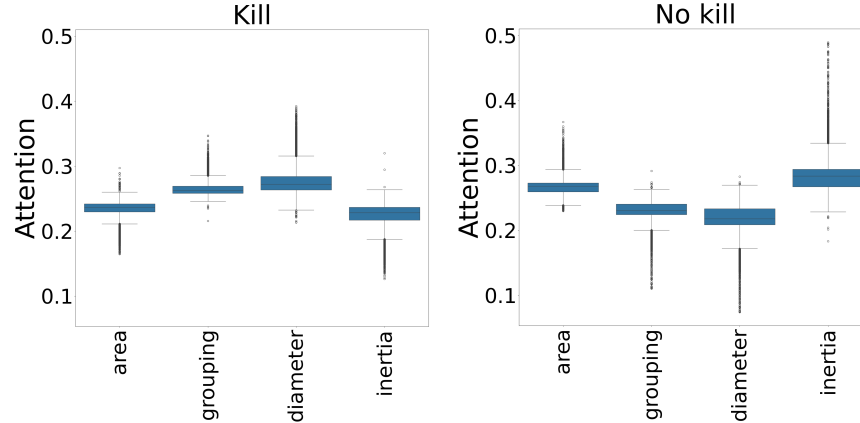


Fig. 4: Feature importance with Attention.

described in Section 4.1 for each match and each team, i.e., 516 examples. As all sequences do not have the same length, we standardized them to 5 200 time steps (maximum duration of a match in the dataset) by 0-padding, so that they can be processed in a batch. A time step corresponds to 1 second.

The 516 examples were divided randomly (with `random_state = 69`) into training (60%), validation (20%) and test (20%) sets. For each example, the lifelines of the opposing players provide, at each time step, the ground truth class *kill* (at least 1 player from the opposing team is dead) or *no kill* (all players from the opposing team are alive). As the *no kill* class has a large majority (*kill*: 33% and *no kill*: 67% in the training set), we considered a *cost-sensitive learning* technique which weights each class in the loss function. Class weights were calculated from the inverse of the class distributions present in the training set (*kill*: 1.52, *no kill*: 0.74), and applied during model training. The network was trained according to the cost-sensitive binary cross-entropy loss, with the Adam optimizer, 0.001 learning rate and a batch size of 32.

The results are presented in the next section for the test set (used as a batch).

5.2 Results

First, we are interested in controlling whether or not the attention mechanism improves the learning process. Table 1 shows the results of the two CNN-LSTM networks with and without the attention mechanism for the test set. The network with attention out-

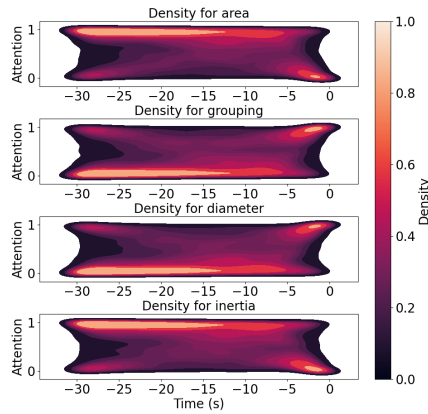


Fig. 5: Attention distribution (KDE plots) – the horizontal axis indicates the time before the kill occurs.

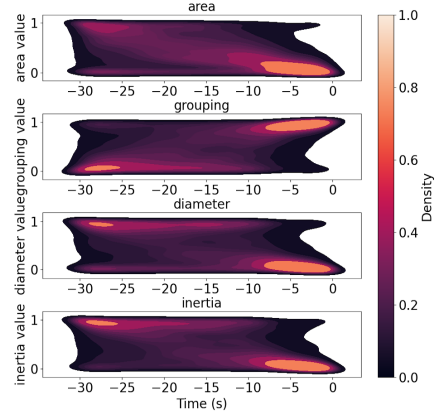


Fig. 6: Feature value distribution (KDE plots) – the horizontal axis indicates the time before the kill occurs.

performs the network without attention, with scores around 0.86 while the network without attention peaks at 0.74. On the one hand, these results demonstrate the robustness of the CNN-LSTM network for processing the dataset, but more importantly they confirm the contribution of the attention mechanism. The attention layer is therefore useful for learning, and its interpretation is capable of providing information specifying the relationship between the input features and the prediction of an event.

Figure 4 shows the distribution of attention between polygon features, depending on whether the event to be predicted is *kill* or *no kill*. During a *kill* event, *diameter* and *grouping* show higher attention values with a wide distribution. *area* and *inertia* register a lower value. For *no kill* events, the trend is reversed. To predict a *kill* event, attention is drawn to the features *grouping* and *diameter*, which are clues of the shape of the polygon, depending on whether it is more or less stretched. To predict a *no kill* event, the network is more attentive to the polygon’s inertia, relative to the variation in its shape.

Then, we analyzed the distribution of the attention weights in a 30-second interval before the number of live players on the opposing team decreased. Figure 5 shows that a change in attention occurs around 15 s before a death event (confirming [8] in establishing 15 s as enabling the best detection rate of important events). In addition, there is a clear variation pattern in the attention, with *grouping* and *diameter* attention increasing as one gets closer to the death event. Figure 6 shows that feature values change over time. In addition, we can clearly see that there is a tendency for the team to group together to make a *kill*.

We can observe these trends from a more general point of view with Figure 7, which shows high attention as *area* increases with high *grouping*. This is indicative of the need for the team to group together to make a kill, surrounding the target within reach of teammates to help each other.

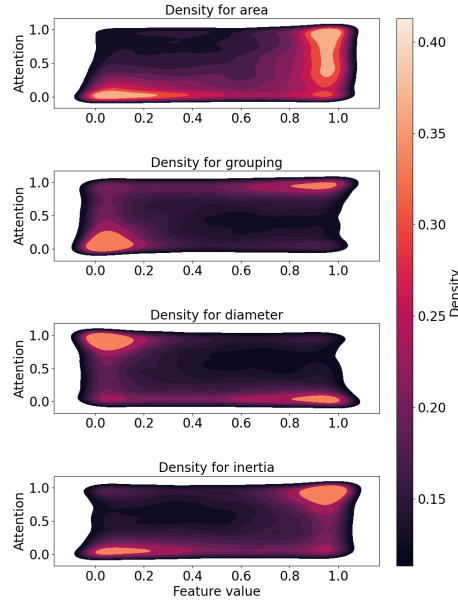


Fig. 7: Correlation distribution between Attention and Feature value (KDE plots).

6 Conclusion

Using a CNN-LSTM network with attention not only improved the accuracy of predictions by focusing on the relevant information, it also provided a way for interpretation. By examining the attention weights, it was indeed possible to identify which geometric features influenced most the network. This information highlighted the moments and actions that impacted the number of active players on the opposing team. Integrating an attention mechanism into a CNN-LSTM network therefore offered the opportunity to interpret the dynamics of collective play.

Acknowledgements

The research described here is supported by the French National Research Agency (ANR) and Region Normandie under grant HAISCoDe

References

1. Abnar, S., Zuidema, W.: Quantifying attention flow in transformers. arXiv preprint arXiv:2005.00928 (2020)

2. Ahmad, S., Bryant, A., Kleinman, E., Teng, Z., Nguyen, T.H.D., Seif El-Nasr, M.: Modeling individual and team behavior through spatio-temporal analysis. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. pp. 601–612 (2019)
3. Bibal, A., Cardon, R., Alfter, D., Wilkens, R., Wang, X., François, T., Watrin, P.: Is attention explanation? an introduction to the debate. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 3889–3900 (2022)
4. Block, F., Hodge, V., Hobson, S., Sephton, N., Devlin, S., Ursu, M.F., Drachen, A., Cowling, P.I.: Narrative bytes: Data-driven content production in esports. In: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video. pp. 29–41 (2018)
5. Castronova, E.: The surprising impact of professional sports on esports. *Journal of Sports Economics* **21**(7), 705–717 (2020)
6. Cavadenti, O., Codocedo, V., Boulicaut, J.F., Kaytoue, M.: What did I do Wrong in my MOBA Game?: Mining Patterns Discriminating Deviant Behaviours. In: International Conference on Data Science and Advanced Analytics. Montréal, Canada (Oct 2016), <https://hal.archives-ouvertes.fr/hal-01388321>
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
8. Chu, W.T., Chou, Y.C.: Event detection and highlight detection of broadcasted game videos. In: Computational Models of Social Interactions. pp. 1–8 (2015)
9. Drachen, A., Yancey, M., Maguire, J., Chu, D., Wang, I.Y., Mahlmann, T., Schubert, M., Klabajan, D.: Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2). In: 2014 IEEE Games Media Entertainment. pp. 1–8. IEEE (2014)
10. Frencken, W., Lemmink, K., Delleman, N., Visscher, C.: Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science* **11**(4), 215–223 (2011)
11. Hauri, S., Djuric, N., Radosavljevic, V., Vucetic, S.: Multi-modal trajectory prediction of NBA players. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1640–1649 (2021)
12. Hauri, S., Vucetic, S.: Group activity recognition in basketball tracking data—neural embeddings in team sports (nets). arXiv preprint arXiv:2209.00451 (2022)
13. Himmelstein, M., Shapiro, J.: Esports: A new era of sport with opportunities for sports scientists. *International Journal of Sports Physiology and Performance* **15**(4), 485–487 (2020)
14. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
15. Jain, S., Wallace, B.C.: Attention is not explanation. arXiv preprint arXiv:1902.10186 (2019)
16. Lan, X., Duan, L., Chen, W., Qin, R., Nummenmaa, T., Nummenmaa, J.: A player behavior model for predicting win-loss outcome in MOBA games. In: International Conference on Advanced Data Mining and Applications. pp. 474–488. Springer (2018)
17. Mao, W., Liu, M., Salzmann, M.: History repeats itself: Human motion prediction via motion attention. In: Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16. pp. 474–489. Springer (2020)
18. Mehra, N., Zhong, Y., Tung, F., Bornn, L., Mori, G.: Learning person trajectory representations for team activity analysis. arXiv preprint arXiv:1706.00893 (2017)
19. Monti, A., Bertugli, A., Calderara, S., Cucchiara, R.: Dag-net: Double attentive graph neural network for trajectory forecasting. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 2551–2558. IEEE (2021)
20. Mora-Cantalops, M., Sicilia, M.Á.: MOBA games: A literature review. *Entertainment computing* **26**, 128–138 (2018)

21. Mortelier, A., Rioult, F.: Addressing the complexity of MOBA games through simple geometric clues. In: International Conference on Computational Collective Intelligence. pp. 643–649. Springer (2022)
22. Qi, Z., Shu, X., Tang, J.: Dotanet: Two-stream match-recurrent neural networks for predicting social game result. In: 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM). pp. 1–5 (Sep 2018). <https://doi.org/10.1109/BigMM.2018.8499076>
23. Ringer, C., Nicolaou, M.A.: Deep unsupervised multi-view detection of video game stream highlights. In: Proceedings of the 13th International Conference on the Foundations of Digital Games. pp. 1–6 (2018)
24. Sanford, R., Gorji, S., Hafemann, L.G., Pourbabae, B., Javan, M.: Group activity detection from trajectory and video data in soccer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 898–899 (2020)
25. Shah, R., Romijnders, R.: Applying deep learning to basketball trajectories. arXiv preprint arXiv:1608.03793 (2016)
26. Sifa, R., Drachen, A., Bauckhage, C.: Profiling in games: Understanding behavior from telemetry. *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective* (2018)
27. Silva, A.L.C., Pappa, G.L., Chaimowicz, L.: Continuous outcome prediction of league of legends competitive matches using recurrent neural networks. In: SBC-Proceedings of SBCGames. pp. 2179–2259 (2018)
28. Summerville, A., Cook, M., Steenhuisen, B.: Draft-analysis of the ancients: predicting draft picks in dota 2 using machine learning. In: Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (2016)
29. Tang, Y., Ma, L., Liu, W., Zheng, W.: Long-term human motion prediction by modeling motion context and enhancing motion dynamic. arXiv preprint arXiv:1805.02513 (2018)
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*. vol. 30. Curran Associates, Inc. (2017)
31. Victor, B., Nibali, A., He, Z., Carey, D.L.: Enhancing trajectory prediction using sparse outputs: application to team sports. *Neural Computing and Applications* **33**, 11951–11962 (2021)
32. Wang, W.: Predicting Multiplayer Online Battle Arena (MOBA) Game Outcome Based on Hero Draft Data. Master’s thesis, Dublin, National College of Ireland (December 2016), <http://trap.ncirl.ie/2523/>
33. Wang, Y., Zhang, T., Guo, X., Shen, Z.: Gradient based feature attribution in explainable ai: A technical review (2024)
34. White, A., Romano, D.: Scalable psychological momentum estimation in esports. In: Workshop SUM ’20: State-based User Modelling, The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM ’20). ACM (2020)
35. Wiegrefe, S., Pinter, Y.: Attention is not not explanation. arXiv preprint arXiv:1908.04626 (2019)
36. Yao, Q., Liao, X., Jin, H.: Hierarchical attention based recurrent neural network framework for mobile MOBA game recommender systems. In: 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom). pp. 338–345. IEEE (2018)
37. Yu, C., Zhu, W.n., Sun, Y.m.: E-sports ban/pick prediction based on bi-LSTM meta learning network. In: International Conference on Artificial Intelligence and Security. pp. 97–105. Springer (2019)
38. Zhang, Q., Zhang, X., Hu, H., Li, C., Lin, Y., Ma, R.: Sports match prediction model for training and exercise using attention-based LSTM network. *Digital Communications and Networks* **8**(4), 508–515 (2022)

39. Zhao, Y., Yang, R., Chevalier, G., Shah, R.C., Romijnders, R.: Applying deep bidirectional LSTM and mixture density network for basketball trajectory prediction. *Optik* **158**, 266–272 (2018)
40. Ziyi, Z., Bunker, R., Takeda, K., Fujii, K.: Multi-agent deep-learning based comparative analysis of team sport trajectories. *IEEE Access* (2023)