Automated Detection of Shot Events in Game Phases Using GNSS Data from a Single Team

Dermot Sheridan[1][0000-0003-3687-1943], Valerio Antonini[1][0009-0004-2607-9623], Michael Scriney[1,2][0000-0001-6813-2630], and Mark Roantree[1,2][0000-0002-1329-2570]

[1] School of Computing, Dublin City University, Dublin 9, D09 V209 Dublin, Ireland
[2] Insight SFI Research Centre for Data Analytics, School of Computing, Dublin City University, Dublin 9, Ireland
dermot.sheridan36@mail.dcu.ie, valerio.antonini3@mail.dcu.ie, michael.scriney@dcu.ie, mark.roantree@dcu.ie

Abstract. Recent advancements in wearable Global Navigation Satellite Systems (GNSS) devices and optical tracking technologies have significantly expanded the availability of spatiotemporal data in sports analytics. Combined with the common practice of manually recording game event logs, these sources provide unprecedented levels of data that describe in detail what happens during a game. However, the manual collection and tagging of events within a game can be costly and time-consuming, requiring domain experts to repeatedly view and annotate footage to identify discrete events. This study presents a novel method for automatically detecting shooting events using spatiotemporal metrics using GNSS tracking data from a single team to achieve event identification.

Keywords: Spatiotemporal Data · GNSS Tracking · Shooting Events Classification · Sports Analytics

1 Introduction

In recent years, the availability of tracking data has significantly increased due to the use of wearable GNSS devices and advancements in optical object tracking technology [1]. The unprecedented increase in data availability has significantly enhanced data analytics in sports, enabling machine learning (ML) algorithms to evaluate player and team performance more effectively, providing deeper insights and improving overall performance [2]. This increase availability of data has also led to a rise in spatial sports research within invasion team sports. In this domain, researchers commonly utilise two types of spatiotemporal data: player tracking data and event log data [3]. A recent framework was outlined to help categorise these diverse research methods and models used in team sports analysis. One category highlighted is data mining, which involves using representations and structures for more advanced analysis. The data collected can serve as inputs for sophisticated algorithms and probabilistic methods to predict and automatically label events more quickly [4].

Gaelic Football, the most popular Irish national sport, has been described as a hybrid of soccer and Australian Rules and is played with a round ball on a rectangular pitch [5]. Elite inter-county competition typically begins in January with the National League and concludes in September after the All-Ireland Championship [6]. The sport maintains an amateur status; however, the training regimes of elite players mirror those of professional athletes. A typical week for an elite player includes one to two resistance training sessions, two field sessions, and a match or training at the weekend [7]. Performance analysis has shown that winning teams in elite Gaelic football often demonstrate superior offensive capabilities, including higher total scores, shot efficiency, and counterattacking ability ([6], [8]). Strong defensive performance, particularly in generating turnovers, is crucial for success [6].

The intense competition in sports has led to the recruitment of dedicated sports scientists within team staff to optimise player and team performance throughout the season [9]. This professional approach has resulted in a significant increase in the use of player-tracking technologies at club and elite inter-county levels during both games and training. These technologies have revolutionised training methodologies, providing valuable insights that enhance overall performance [10].

The extensive data generated from GNSS tracking devices offers a rich source of information that details player activity profiles during games. However, detecting specific events still necessitates extensive manual work by human annotators. The potential exists for this data to aid in the automatic tagging of events, streamlining the process and reducing the reliance on manual annotation

Contribution. This paper introduces a novel approach for automatically detecting shot events during phase of game using spatiotemporal data from a single team's GNSS tracking data. Our method effectively identifies shot events even with data from just a few games.

Paper Structure. The remainder of this paper is structured as follows: in §2, we discuss related research in the evolution of spatiotemporal data mining in the area of sports analytics; in §4, we present a 3-step methodology to process event logs and GNSS data to extract game phases and features based on distance and similarity of speed between players; in §5, we present the results of our experiments, while in §6, we discuss our results and implications; and finally in §6, we present our conclusions along with future directions.

2 Related Research

Recent advancements in sports analytics have leveraged spatiotemporal data for analysing player and team performances. Technologies like optical tracking and GNSS are crucial in capturing detailed movement data in team sports such as soccer and Gaelic Football ([11], [12]). Recent studies have focused on automating game event tagging, providing valuable feedback to players and coaches [13].

The shift from manual event tagging to ML has utilised rule-based systems and advanced neural networks. One study demonstrated the use of algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests to detect football events from positional data [14]. Another study used a data-driven approach to identify offensive tactics, highlighting the importance of algorithm choice for strategic insights [15]. A rule-based algorithm was proposed for identifying basic soccer events, achieving high accuracy but noting limitations in capturing complex game dynamics [16]. Further advancement was made with a deterministic decision tree-based algorithm, achieving over 90% detection accuracy but relying on detailed data, which may not always be available [17].

GNSS data and heatmaps were employed in the 6MapNet model to identify players based on movement styles, demonstrating GNSS data's potential for detailed analysis [18]. However, a gap remains in using GNSS data from a single team for event labeling. Our research addresses this by proposing a novel method for automatically labeling shot events using players' spatiotemporal data. This approach is particularly relevant for Gaelic football, where data collection challenges exist due to the sport's amateur status. We aim to analyse player movement during the phase of play and identify a shot event.

This novel method not only fills a significant gap in the literature but also provides a framework for applying advanced ML techniques to GNSS data, offering practical solutions for amateur sports teams to gain deeper insights into performance and strategies without extensive resources.

3 Data

This study utilises two distinct datasets: GNSS data and event log data. The GNSS data were collected on the same team during 11 competitive Gaelic Football inter-county games across the 2019-2021 seasons. Athletes of the analyzed team wore micro 10Hz GNSS sensor devices (STATSports Apex 10 Hz, Northern Ireland, UK), which recorded ten observations per second for each of the following variables: latitude, longitude, speed, 3D acceleration, and 3D gyroscope (which provide information about the angular velocity). Data were filtered and pre-processed using STATSports software (version 4.5.19). Previous research [19] has demonstrated that the error margin of the 10 Hz Apex unit, which is between 1-2% of the measured distances, is negligible.

Event logs are generated from video analysis, and teams use these computerised notational analysis statistics to enhance their performance and seek to analyse their own and their opponents' gameplay systematically [6]. Event logs are not continuous; they are only recorded when an event occurs. game footage from internal team video recordings and external media broadcasters was imported and coded using a custom-built tagging panel. All games were coded using Nacsport Scout Plus software (Scout Plus V 6.5.0, New Assistant for Coach Sport, S.L). During an analysis of a game, the coder had the ability to pause and rewind the video to watch events more than once; it was also possible to use a slow-motion feature to allow for a clearer view of fast dynamic events. Following data validation, the coding events were then exported into Microsoft Excel (Microsoft, USA) to facilitate analysis.

4 Methodology

This paper focuses on classifying game phases using features extracted from GNSS data of one of the two playing teams. Initially, we utilise event log data to isolate game phases, which we define as intervals starting when a team gains possession of the ball and ending when the same team loses possession or attempts a shot. We conduct two distinct experiments. In the first experiment, we identify game phases that conclude with a shot by the analysed team (Team A) using GNSS data from that team. In the second experiment, we identify game phases that end with a shot by the opposing team (Team B) using GNSS data from Team A. The objective of this study is to explore the feasibility of classifying different game phases based on GNSS data from one team, without relying on event log data or data from the opposing team.

The methodology for this study follows a sequence of three steps:

1. Phases Extraction from Event Logs Data. The first step involves extracting game phases from the event log data. The event log data describes a series of sequential events occurring during the game, such as ‘Throw in’, ‘Attack Team B’, ‘TB Shot’, ‘TA Kickout’, and ‘TA Possession from KO’, along with their corresponding timestamps. We isolate phases of ball possession for the two teams, which start when one of the two teams gains the ball (it happens after a Throw in, Kickout, Turnover, etc.) and end when the team loses control of the ball or makes a shot. We record the moment when the ball is gained (start time) and when the ball is lost or a shot is made (end time). This step takes as input a data frame of sequential events and transforms it into a data frame of game phases describing which team is in possession of the ball, along with the corresponding timestamp. Each phase is labeled based on the target of the analysis. When classifying shots during the attacking phase, phases ending with a "TA Shot" are labeled as 1, while all other phases are labeled as 0. On the contrary, when classifying shots performed by the opponent team, phases ending with a "TB Shot" are labeled as 1. The resulting data frame for the game's phases will display the corresponding events' start and end times. This process is repeated for each of the analysed games. A sample of the resulting data frame of the phases of game is shown in Table 1.

Table 1. Sample data frame illustrating the phases of the game. Columns include *gameID* (ID of the corresponding game), *Team* (team in possession of the ball), *Start Event* (event initiating the phase), *End Event* (event concluding the phase), *Start Time* (game time in seconds at the beginning of the phase), and *End Time* (game time in seconds at the end of the phase).

GameID	Team	Start Event	End Event	Start Time	End Time
1	A	Team A Pos frm TO	TA Shot(P)	5	62
1	B	TB Kickout	TB Kickout	63	71
1	A	TA Pos frm KO	TA Shot(P)	72	98
1	B	Team B Pos frm TO	Attack TB	99	149

2. Features Extraction from GNSS Data. After identifying the game phases and their respective starting and ending timestamp, we created a data frame containing features extracted from GNSS data. The 10Hz GNSS device collects data at 10 observations per second, including latitude, longitude, speed, 3D acceleration, and 3D gyroscope. Given a game phase with a defined starting and ending timestamp, we computed metrics such as the average and standard deviation of distances, and the average and standard deviation of the absolute differences in speed, 3D acceleration, and 3D gyroscope between each pair of players in that interval of time. This process is repeated for each phase of the game. Metrics were collected for every player on the team to ensure consistency, with non-participating players' values recorded as 0. These metrics indicate team movement and coordination during the phase. Each metric corresponds to a column in the final data frame, resulting in columns for the average and standard deviation of distance, speed, 3D acceleration, and 3D gyroscope between each pair of players. This step is summarised in Figure 1. The input GNSS data is transformed into a data frame of similarity scores (average and standard deviation) for each variable and for each couple of players Table 2. The resulting data frame contains 441 rows (one for each game phase) and 13,921 columns. Due to varying player participation, some features may be sparse, necessitating a feature selection step to eliminate noise and identify the most relevant features for each ML model tested.

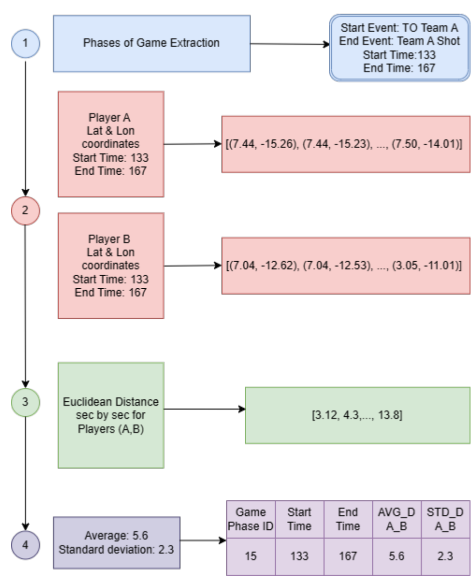

Fig. 1. Computation of average and standard deviation of distance between a pair of players for a selected game phase. This process is repeated for each pair of players and for each extracted game phase.

Table 2. Resulting data frame after feature extraction from GNSS data. Each row represents a game phase, with columns for game ID, start/end time, distances, speed, and acceleration (x) metrics between player pairs (only shown between player 1 and player 2). Other features that are not shown include additional acceleration (y and z) and gyroscope (x, y, and z) differences for each player pair.

EventID	Players Pair	Avg Dist	Std Dist	AvG D Sp	Std D Sp	Avg D ACx	Std D ACx	Avg D ACy	Std D ACy	Avg D ACz	Std D ACz
1	(1,2)	19.3	2.3	2.9	2.3	1.1	0.9	0.8	0.4	1.3	0.3
1	(1,3)	32.1	11.3	3.1	1.9	0.5	1.3	0.6	0.3	1.1	0.2
1	(1,4)	5.7	5.1	1.3	0.5	0.4	0.1	1.4	0.6	1.0	0.1
1	(1,5)	8.9	2.3	4.4	1.7	0.6	0.4	1.7	0.8	0.9	0.3
1	(1,6)	22.1	6.3	3.3	1.6	1.1	0.3	2.1	0.9	0.8	0.3
1	(1,7)	31.1	8.1	2.9	1.4	0.9	0.5	2.0	0.8	0.8	0.3

3. Feature Selection and Modelling. Our approach utilises an automated feature selection method to identify the most significant features for predictive modeling from a data frame containing thousands of features. We use the Recursive Feature Elimination with Cross-Validation (RFECV) method, which iteratively removes the least important features and evaluates model performance using k-fold (k = 3) cross-validation [20]. K-fold cross-validation is a method used to assess the performance of a ML model by dividing the data into k equal subsets. The model is trained on k-1 subsets and validated on the remaining subset, and this process is repeated k times, with each subset used exactly once as the validation set. The RFECV method reduces the risk of overfitting and provides a reliable estimate of the model's generalisability. RFECV involves training a model on the full set of features, ranking them based on importance scores, removing the least important feature, and evaluating the model. This cycle repeats until only one feature remains, and the set of features achieving the best cross-validation performance is selected. The outcome of the recursive feature elimination process is a refined and optimised set of features tailored by each ML model during the training phase. In the testing phase, each model is evaluated using this optimised feature set to assess its performance. The motivation for using RFECV in our study is supported by several key factors and references. Firstly, RFECV is known for its ability to effectively handle high-dimensional datasets by identifying the most relevant features, thereby improving model performance and interpretability [21]. RFECV method helps in mitigating the risk of overfitting by incorporating cross-validation, which ensures that the selected features generalize well to unseen data [22]. Lastly, RFECV has been widely used in various domains, including malicious intrusion detection in computer networks [23], diagnosis of Alzheimer's disease [24], and fraud detection [25]; demonstrating its versatility and robustness in feature selection. By using RFECV, we aim to leverage its strengths to enhance the predictive power of our models, ensuring that the features selected contribute significantly to the performance and reliability of the predictions.

5 Results

In this section, we present the performance evaluation of ML models on two distinct classification tasks: identifying game phases culminating in a shooting event for Team A (Task A) and for the opposing Team B (Task B). Due to the relative rarity of shooting events compared to other events, both datasets are inherently imbalanced. To address this issue, we employed undersampling of the majority class in both tasks, training the models on datasets balanced with an equal number of shooting and non-shooting events.

This paper aims to demonstrate the effectiveness of ML models in recognising partial game phases, potentially leading to the future automation of event labeling based solely on spatiotemporal data from a single team. Both tasks utilise GNSS data collected on the same team during the first halves of 11 GF games across the years 2019, 2020, and 2021. The analysis includes data from all players participating in these games. The ML models' performances are evaluated using the following metrics [26]:

- **Accuracy** measures the proportion of correct predictions (both true positives TP and true negatives TN) out of the total number of predictions, which includes true positives, true negatives, false positives (FP), and false negatives (FN).
- **Precision** is the proportion of true positive results out of all positive predictions made by the model.
- **Recall** (or True Positive Rate) is the proportion of true positive results out of all actual positive cases.
- **F1 Score** is the harmonic mean of precision and recall. It balances the two metrics and provides a single measure of a model's performance, especially useful when the class distribution is imbalanced.
- **AUC** (Area Under the ROC Curve) measures the ability of the classifier to distinguish between positive and negative classes. It is a single scalar value that summarizes the performance of a classifier across all possible classification thresholds.

5.1 Classifying shooting events for Team A

In the first experiment, we evaluated the performance of four ML models in classifying shooting events performed by Team A: Random Forest, Decision Tree, XGBoost, and Logistic Regression. The results for each model are presented in Table 3. The Random Forest model demonstrated the highest performance, with an F1 score of 0.69. Decision Tree and Logistic Regression models showed lower performance. Achieving a high F1 score indicates that the model not only accurately predicts shooting events (high precision) but also effectively identifies most of the actual shooting events (high recall). Random Forest, which selected 11,136 features, and Decision Tree, which selected 13,786 features, highlight the benefits of feature-rich models. However, the XGBoost model achieved similar performance with a significantly smaller set of features. This indicates that while a high number of features can enhance performance, the efficiency and power of the algorithm are also crucial. These findings highlight the importance of balancing feature selection with model complexity to achieve optimal performance.

5.2 Classifying shooting events for opposing team (Team B)

Table 4 displays the performances of the ML models in classifying shooting events for the opposing team (Team B). Recall, as this is the opposing team, these

Table 3. Performance metrics of ML models in classifying Team A shooting events. Precision and recall are weighted by the number of true instances per class and averaged.

Model	Accuracy	F1 Score	AUC	Precision	Recall	Selected Features
Logistic Regression	0.52	0.52	0.53	0.52	0.52	4
Decision Tree	0.52	0.54	0.52	0.52	0.53	13786
Random Forest	**0.63**	**0.69**	0.61	**0.65**	**0.63**	11136
XGBoost	0.61	0.67	**0.67**	0.62	0.61	3

data are based on Team A' spatiotemporal data. In this evaluation, XGBoost outperformed the other models across all metrics (sharing the highest accuracy with Random Forest) while using a smaller set of features. An AUC of 0.78 indicates that there is a 78% chance that the model will correctly distinguish between a randomly chosen positive instance (shooting event) and a randomly chosen negative instance (non-shooting event). In practical terms, this level of performance is useful for sports analysts because it provides a reliable way to identify shooting events from the available GNSS data.

Table 4. Performance metrics of ML models in predicting opponent (Team B) shooting events.

Model	Accuracy	F1 Score	AUC	Precision	Recall	Selected Features
Logistic Regression	0.67	0.68	0.70	0.67	0.67	1695
Decision Tree	0.47	0.45	0.47	0.47	0.47	8294
Random Forest	**0.69**	**0.68**	0.73	0.68	0.68	11325
XGBoost	**0.69**	0.67	**0.78**	**0.69**	**0.69**	15

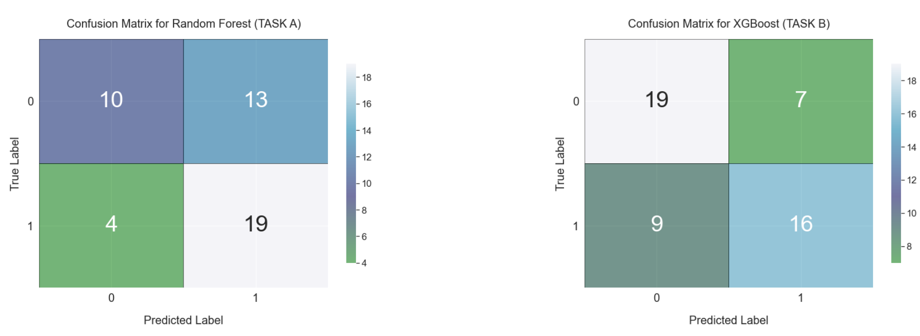

Fig. 2. Confusion matrix for Random Forest predicting Team A (Left) and Team B (Right) shooting events.

6 Discussion

Four ML models were developed to predict shooting events using pair-wise similarity metrics of players' movements. The Random Forest model performed best for predicting Team A shooting events (Task A), while XGBoost perfomed best for the opponent team's events (Task B). In Task A, the Random Forest model achieved a higher F1 score and recall, effectively identifying true positives (actual Team A shooting events) but also producing many false positives, indicating a tendency to overestimate shots (Figure 3 Left). In contrast, the XGBoost model in Task B showed higher accuracy and AUC score, demonstrating superior overall performance and a better ability to distinguish between classes (Figure 3 Right). XGBoost achieved better precision with fewer false positives, though it had more false negatives, reflecting a more balanced performance but missing more actual Team B shots (Figure 3 Right).

Classyfing game phases events from GNSS data can offer valuable insights into how players' spatiotemporal data influences game phases and leads to specific events. Currently, these event labels are manually assigned by watching game videos, a time-consuming and labour-intensive process. By developing ML algorithms that automatically assign labels to events, we can significantly speed up this process and improve efficiency, enabling more timely and accurate analysis of player movements and game dynamics [4].

7 Conclusion

This study represents the first attempt to identify shooting events in Gaelic Football using player tracking data of a single team. We leverage spatiotemporal data (speed, distance, 3D acceleration, and 3D gyroscope) to compute similarity metrics between players during play phases. Our experimental findings demonstrate that integrating similarity scores, automatic feature selection, and ML models enables the identification of phases of play that result in shot events using GNSS tracking data. Future research should explore the use of additional features to identify the movements that positively impact the outcome of a phase of game.

Acknowledgments. This work was funded by Science Foundation Ireland through the Centre for Research Training in Artificial Intelligence (SFI/18/CRT/6223), Machine Learning (SFI/18/CRT/6183) and Insight Centre for Data Analytics (SFI/12/RC/2289_-P2) .

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lutz, J., Memmert, D., Raabe, D., Dornberger, R. and Donath, L., 2020. Wearables for integrative performance and tactic analyses: Opportunities, challenges, and future directions. International journal of environmental research and public health, 17(1), p.59.
2. Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Rodrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S. and Lemmink, K.A.P.M., 2021. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. European Journal of Sport Science, 21(4), pp.481-496.

3. Memmert, D. and Rein, R., 2018. game analysis, big data and tactics: current trends in elite soccer. German Journal of Sports Medicine/Deutsche Zeitschrift fur Sportmedizin, 69(3).
4. Gudmundsson, J. and Horton, M., 2017. Spatio-temporal analysis of team sports. ACM Computing Surveys (CSUR), 50(2), pp.1-34.
5. Cullen, B.D., Cregg, C.J., Kelly, D.T., Hughes, S.M., Daly, P.G. and Moyna, N.M., 2013. Fitness profiling of elite level adolescent Gaelic football players. The Journal of Strength Conditioning Research, 27(8), pp.2096-2103.
6. Gamble, D., Bradley, J., McCarren, A. and Moyna, N.M., 2019. Team performance indicators which differentiate between winning and losing in elite Gaelic football. International Journal of Performance Analysis in Sport, 19(4), pp.478-490.
7. Beasley, K.J., 2015. Nutrition and Gaelic football: Review, recommendations, and future considerations. International Journal of Sport Nutrition and Exercise Metabolism, 25(1), pp.1-13.
8. Carroll, R., 2013. Team performance indicators in Gaelic football and opposition effects. International Journal of Performance Analysis in Sport, 13(3), pp.703-715.
9. Malone, S., Collins, K., McRobert, A. and Doran, D., 2021. Quantifying the training and match-play external and internal load of elite Gaelic football players. Applied Sciences, 11(4), p.1756.
10. Mangan, S., Ryan, M., Devenney, S., Shovlin, A., McGahan, J., Malone, S., O'Neill, C., Burns, C. and Collins, K., 2017. The relationship between technical performance indicators and running performance in elite Gaelic football. International Journal of Performance Analysis in Sport, 17(5), pp.706-720.
11. Beernaerts, J., De Baets, B., Lenoir, M. and Van de Weghe, N., 2020. Spatial movement pattern recognition in soccer based on relative player movements. PloS one, 15(1), p.e0227746.
12. Sheridan, D., Brady, A.J., Nie, D. and Roantree, M., 2024. Predictive analysis of ratings of perceived exertion in elite Gaelic football. Biology of Sport, 41(4), pp.61-68.
13. Salim, F.A., Postma, D.B., Haider, F., Luz, S., Beijnum, B.J.F.V. and Reidsma, D., 2024. Enhancing volleyball training: empowering athletes and coaches through advanced sensing and analysis. Frontiers in Sports and Active Living, 6, p.1326807.
14. Richly, K., Bothe, M., Rohloff, T. and Schwarz, C., 2016, April. Recognizing compound events in spatio-temporal football data. In International conference on internet of things and big data (Vol. 2, pp. 27-35). SCITEPRESS.
15. Decroos, T., Van Haaren, J. and Davis, J., 2018, July. Automatic discovery of tactics in spatio-temporal soccer match data. In Proceedings of the 24th acm sigkdd international conference on knowledge discovery data mining (pp. 223-232).
16. Khaustov, V. and Mozgovoy, M., 2020. Recognizing events in spatiotemporal soccer data. Applied Sciences, 10(22), p.8046.
17. Vidal-Codina, F., Evans, N., El Fakir, B. and Billingham, J., 2022. Automatic event detection in football using tracking data. Sports Engineering, 25(1), p.18.
18. Kim, H., Kim, J., Chung, D., Lee, J., Yoon, J. and Ko, S.K., 2021, September. 6MapNet: Representing soccer players from tracking data by a triplet network. In International Workshop on Machine Learning and Data Mining for Sports Analytics (pp. 3-14). Cham: Springer International Publishing.
19. Beato, M., Coratella, G., Stiff, A. and Iacono, A.D., 2018. The validity and between-unit variability of GNSS units (STATSports Apex 10 and 18 Hz) for measuring distance and peak speed in team sports. Frontiers in physiology, 9, p.1288.

20. Misra, P. and Yadav, A.S., 2020. Improving the classification accuracy using recursive feature elimination with cross-validation. Int. J. Emerg. Technol, 11(3), pp.659-665.
21. Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46(1), 389-422.
22. Priyatno, A.M. and Widiyaningtyas, T., 2024. A Systematic Literature Review: Recursive Feature Elimination Algorithms. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 9(2), pp.196-207.
23. Ustebay, S., Turgut, Z. and Aydin, M.A., 2018, December. Intrusion detection system with recursive feature elimination by using random forest and deep learning classifier. In 2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT) (pp. 71-76). IEEE.
24. Richhariya, B., Tanveer, M., Rashid, A.H. and Alzheimer's Disease Neuroimaging Initiative, 2020. Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomedical Signal Processing and Control, 59, p.101903.
25. Mustaqim, A.Z., Adi, S., Pristyanto, Y. and Astuti, Y., 2021, June. The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection. In 2021 International conference on artificial intelligence and computer science technology (ICAICST) (pp. 270-275). IEEE.
26. Hossin, M. and Sulaiman, M.N., 2015. A review on evaluation metrics for data classification evaluations. International journal of data mining knowledge management process, 5(2), p.1.