Sports Performance Analysis System: A Survey and Taxonomy of Comparison

Tulaisiram Naidoo, Alice Loh Yew Huey, Yu Zexu andShamala K. Subramaniam
[1] Department of Communication Technology and Networking,Faculty of Computer Science
[2] Universiti Putra Malaysia, Selangor 43400, Malaysia

ABSTRACT

In recent years, machine learning has been applied to various important fields. Using algorithms and big data analysis can make every job in this field simple and accurate. Today, machine learning can be used in a variety of contexts, machine learning can use big data to learn, analyse and improve. This makes the data accurate. In this regard, machine learning can also be successfully applied in fields related to sports science by analysing the physical data of athletes. This paper has therefore focused on the related research of machine learning in sports science and is deliberated in detail reviewing several research works and highlighting the open issues.

Keywords: **Sport science, Machine learning, Big Data Analysis, Athletes, Performance Analysis**

1. INTRODUCTION

In daily training and competition, it is crucial for athletes to master their own body data. Athletes break through their physical limits, and machine learning is used to analyse thedata of the athletes themselves [1]. At this time, sensors are needed to record calorie consumption, changes in body temperature, heartbeats per minute, and blood oxygen saturation to analyse and remind the athletes whether their training meets the requirements. Sensors are utilized to record data and transmit it to athletes and coaches for visual graphics and intuitive data presentation upon computations. In addition, machine learning has a role in preventing accidental injuries in athletes in the field of sports science. For example, the use of wearable devices for machine learning of the athletes' usual movement posture to detect whether the athletes are at risk of accidental falls in the near future. The sensor can keenly capture any potentially dangerous posture and remind it, which can prevent athletes from receiving accidental injuries during sports. There are many elements of the sports eco-system which are leveraging machine learning. Among the main areas is sports science. Big data analysis based on multiple athletes can make sports science have more accurate data and functions in the field of machine learning. This paper focuses on analyzing substantial machine learning findings in the area of sports science and presents a taxanomy of comparison. The paper is presented in the following order. Section II, discusses the related work.

2. RELATED WORK

The new era of sport is becoming very competitive by the day, especially on a professional level [2]. Countless leagues are created all over the world and the pressure to compete and win in them is heavy. Thus, numerous strategies, training methods and methods of training the athletes are tested and implemented to improve their respective teams and to ensure they are victorious in the matches. New techniques and innovations are always being sought for the athletes to excel on their performance [3]. The most prominent cause is that many sports require a dynamic intermittent team where the players need a varying mastery in anaerobic and aerobic fields [4]. Hence, sports performance analysis has become vital due to the numerous variables involved.

Machine learning is applied widely in sports performance analysis. sports injuries prediction, and team formations [5]. Sports analysis technology has become popular in the market. One of the software is Kinovea is a 2D motion sports analysis software to process kinematic parameters and provide accurate angular and linear measurements through digitalization the data captured, a solid and applicable result can be generated through application of Kinovea [6]. Besides, Catapult is a sport analysis system that applies Global Positioning System (GPS), accelerometers, gyroscopes, and magnetometers to collect data from multiple sports such as American football or rugby football such as speed, acceleration, duration of game played, etc for workload measurement [7]. Moreover, Catapult has improved further by having a new invention of electronically trackable ball; the hardware device is placed inside the ball [8].

One of the methods applied in sport analytics researchis machine learning. In football, coaches apply machine learning to interpret football statistics into useful information for decision making or opponent teams' assessment [9]. Machine learning is categorised into supervised learning model and unsupervised learning model. Supervised learning model applies classification according to labelled data fed, and unsupervised learning model which applies clustering following the structure of unlabelled data fed [10]. Besides, machine learning is applied in sports for prediction and classification [11]. To determine the difference between actual and predicted results, root-mean-square (RMS) is suitable to be used as a metric [12].

The following are some case studies in which player selection is done through machine learning. First, decision tree, and naive bayes are used to assist coaches in selecting suitable athletes for a match line-up according to the athletes' physiological variables [13]. Next, the research shows comparisons of Random Forest, Naïve Bayes and AdaBoost classifiers in football players' selection, Random Forest has the highest accuracy among the classifiers. However, the authors mention that there is a high possibility of bias and less accurate results to weaker players since the dataset used is from a top leagues and teams' performance result [14]. Hit/it Assistant, an electronic sport system manufactured by Performa.nz Company, is used to measure the control, technique, reaction time, speed, agility, coordination and surround control skill of a player. Different algorithms such as multilayer perceptron algorithm, sequential minimal optimization algorithm (SMO), logistic model tree, logistic regression algorithm (LR), random forest, and SimpleCART are used to train the data set, which SMO and SimpleCART achieve higher accuracy results. Goalkeeper data is excluded from the research because there is no matched Vector Machines (SVM), Neural Networks and Linear Logistic Regression, SVM achieved the higher accuracy, 90%, Logistic Regression is the lowest accuracy, 80%, because linear regression model can only explain partial characteristic of the athletes' performance, neural network fits linear and nonlinear characteristics of the athletes, and SVM does not require huge samples like neural network, so SVM is the best algorithm in the research [19]. Besides, research conducted by Harsha Vardhan Goud et al. in 2019 [20] outlines the role or use of machine learning for Invasion sports refers to sports where separate teams compete with each other to gain possession of the ball or puck and to score using it. By using this and the path/trajectory that the ball takes when being passed, machine learning can be used to bring out the successful cases and beyond that, the most successful strategies. The use of this method on the sports of football was able to measure an absolute mean error of just 0.17 [21] and was also able to involve more parameters like the number of yellow and red cards, successful and unsuccessful attempts, fouls, penalty kick and more into consideration.

Wearable technology helps to collect insights needed for the performance analysis and are used in real time for sports activities. Wearable sensors are great for the players as it can not only function as a sensor for multiple uses like body-worn sensors, heartbeat sensors, speed and acceleration sensors, but it also can help the athletes during training such as to help them keep themselves hydrated when their internal temperature rises. These sensors can be fitted into their equipment such as helms and bats. However, this research does not go in-depth with the aspect of implementing machine line-up [15].

Machine learning is also extensively used to analyze game results. A research project with 10 teams from the Chinese Football Super League is constructed to predict game results using k- nearest neighbour algorithm (KNN), LR, and back propagation neural network (BPNN) which BPNN has the highest accuracy of 80% [16]. Besides, naïve bayes was used to predict a basketball match result, prediction accuracy is only 67% [17].

Furthermore, machine learning is used to analyse athletes' performance. To evaluate 59 handball athletes' performance, neural network and machine learning models such as linear regression, decision tree and support vector regression were used, but the outcome of machine learning was not optimal where the results were fluctuating, while neural network models achieved stable results [18]. To analyse the athletes' running performance, 500 runners' 100m running readings trained using Support

performance analysis of the players using wearable technologies. The sports are categorised into several categories, each with its distinct way of performance analysis, parameters and attributes as follows [20]:

a) Invasion sports
b) Net and wall games
c) Striking and fielding games

learning and using it in real- time but only discusses its role for doing so. Nevertheless, it does give significant insight into the subject matter and also shows how to implement the use of wearable devices and use it to get the necessary information for machine learning.

3. TAXONOMY OF COMPARISON

There are many varying aspects in the field of sports that use machine learning-related technologies for systematic analysis. Each type of sport requires different approaches when it comes to collecting data, training the data and the machine learning model used. In the field of basketball, parallel programming models can be used for fog and edge computing infrastructures to monitor player data collection in basketball games [22]. Machine learning is used to collect the accurate data from basketball players to prepare the athletes and to focus on improving certain skills in competitions. In the field of cricket, machine learning is used to make predictions for cricket matches as it is able to predict and analyse the outcome of the matches [23].

As for the models used, neural network seem to perform better for many sports such as cricket [13], handball [18], volleyball [24] and football [16]. The major benefit of using neural network technology is its ability to solve classification tasks handily in order to make a realistic model for sports performance predictions from previously selected sports performance related issues criteria [25]. Moreover, machine learning and neural networks are used to monitor 14 elite volleyball athletes's performance to reduce injury risk by answering daily questionnaires, ending up with a total of 1112 questionnaires [26].

The related work has been analyzed using the proposedtaxonomy of comparison shown in Table1. The metrics of focus are the objective of the system developed, the Machine Learning methodology developed and implemented and the outcome which has enabled in depth analysis of the respective sport.

Reference	Sports	Goal	Methodology	Dataset	Results
Zhu and Sun, 2019	Running	Player's performance analysis	Support Vector Machines (SVM), Neural Networks and Linear Logistic Regression	Results of 100m running selected from 500 sports school student	SVM has achieved higher accuracy results, with around 90% accuracy
Miljković D. et al. 2019	Basketball	Game result prediction	Naive bayes	778 games from 2009/2010 NBA season	Prediction accuracy is 67%.
Oytun et al., 2020	Handball	Player's performance analysis	Neural network and machine learning models such as linear regression, decision tree and support vector regression	59 players from the North Cyprus Women's Handball Super League.	Neural network has achieved a stable result.
Kapadia et al., 2020	Cricket	Sport analytics for cricket game results using machine learning	Using machine learning to predict and analyse the outcome of cricket matches	IPL Cricket matches for 10 years (2008 to 2017)	KNN:64.2% FP:48.35% Decision Tree algorithm:48% Precision and Recall:49%
Muazu Musa et al., 2021	Volleyball	Performance analysis in volleyball	Application of machine learning on volleyball match performance analysis	50 archers in Malaysia who target to state and national archery competitions	ANN provides a reasonable prediction (CA of up to 92.5%) Machine learning models yield reasonably good predictions.
Ćwiklinski et al., 2021	Football	Player selection	Random Forest, Naïve Bayes and AdaBoost	Top 8 most popular leagues based on the Union of European Football Associations	Random Forest has the highest accuracy

Didem Abidin, 2021	Football	Player selection and team formation	Multilayer perceptron algorithm, sequential minimal optimization algorithm (SMO), logistic model tree, logistic regression algorithm, random forest, and SimpleCART	100 sequences of trainings with Hit/it	SMO and SimpleCART have achieved higher accuracy results MLP 92.6407% SMO 89.1775% LMT 90.4762 % Logistic 85.7143% Naïve Bayes 79.6537% RF 93.9394% SimpleCART 79.2208%
Arie-Willem de Leeuw et al., 2022	Volleyball	Player's performance analysis	Machine learning model and neural network	Collect data from 14 athletes body analysis and answer questionnaires from 1112 athletes.	Machine learning has achieved the stable result.
Hong Y. and Li Y. 2022	Football	Game result prediction	KNN, LR, BPNN	10 teams from the Chinese Football Super League 2008-2018	KNN: 40% LR: 60% BPNN: 80%
Cortez A., Trigo A., Loureiro N. 2022	Football	Player selection	Decision Tree, Naive Bayes	Physiological variables of football players in the game of a Portuguese team from the Second Regional Division of AF Santarem in 2018/2019 season.	Descision Tree & Naive Bayes according to their position: Centre Midfielder [83%,65%] Central Defender [83%,73%] Forward [70%,57%] Full Back [71%,77%] Winger [73%,68%]

Table 1: Taxonomy of comparison

Most of the above research is conducted in developed countries such as countries from Europe which are well managed and organised. In Malaysia, the athletes from Paralympic mentioned lack of facilities such as sharing training space and uncomfortable equipment have affected the training experience [27]. Besides, national athletes also feedback that lack of opportunities and facilities offered to non-national athletes which leads to difficulty in developing young and talented athletes [28]. Moreover, sports in Malaysia faces issues with poor management, politics, and financial support. Financial support is important especially for the team without a good performance result will face difficulty to get funding [29]. Besides, implementation of sports technology with artificial intelligence such as machine learning and cloud computing will consume high cost [30]. Table 2 shows the summary of problems that occurred in Malaysia's sports field.

Research	Participant	Problems
Fitri et al. (2022)	Paralympic athletes	There is no specific training facility or space provided. The athletes have to share facility with the public most of the time. Equipment is not comfortable and size does not fit.
Daniel Azizan (2022)	National national athletes from various sports	Lack of support from the government in developing fresh potential national athletes. There is less equipment, facilities and opportunities provided for non-national athletes.
Fiah et al. (2014)	Interviews with sports administrators and senior sport journalists	Poor management skills, lack of financial support, corruption and occurrence of politics in sports.
Himeur Y. et al. (2022)	Research from Scopus, Elsevier, Wiley, and IEEE.	Deployment of smart technology such as deep learning architectures and cloud servers are expensive.

Table 2: Feedback about sports facility in Malaysia

IV. OPEN ISSUES

Data management in sports is important since the data of every athlete is valuable, but it is also a complex process for data protection. Furthermore, the development of sport science and technology is rapid in Malaysia, but it has not been optimally implemented among the sports community. There is less investment in technology implementation, leading to an insufficient number of wearable sensors. Some players might not own the required wearable sensors, so the data set size for machine learning will be limited.

In the research, it is clear that there are various problems with hardware shortages as its limited by cost and quantity. For many athletes, this professional hardware is a necessity but due to its high price, not everyone can use it. The high cost of these equipment is way beyond the reach for most people and especially those from grassroot. It is not cost-efficient for them to buy and use it as they would be ableto gain better benefit by instead purchasing sports equipment, better facilities, coaches, etc. Hence, this topic has seen limited study. Sports considers generality, not just a certain sport. The same is true in sports science, and many indicators need to be counted. For example, rehabilitation, nutrition, or sports psychology, etc. It takes more analysis based on demographics to get results.

Other than that, there is a lack of off-field research and study conducted over the years regarding performance analysis as it's harder to quantify. More thought has to be putinto how to better correlate the off-field inputs of those involved in the sports field such as the athletes to their performance as various factors can play a role in affecting their performance. Performance analysis is also mostly donefor professional sports in the elite level.

Furthermore, accessibility also plays an important role in adapting these new technologies for commercial use by major teams and by the demographics. If the coaches and athletes are unable to use the system easily and effectively, they would opt out of using the system entirely. The introduction of the new technologies and the need to get this hardware also might turn them off from using it as they may not want to experience the drastic changes in their daily routine. It's more likely they follow the same training routine and feel that the technology given is redundant and unnecessary.

V. CONCLUSION

In the present work, a retrospective analysis of the field of sports science related to machine learning is carried out.

This paper critically analyses the improved strengths and areas of machine learning and neural networks or sports science neural network engines in sports science, which is expected to benefit from this report in the future based on sports science research.

The result of the match is usually based on the prediction data obtained by the athlete's own training. Machine learning models are increasingly used in the field of sports science because athletes can better understand theirown problems in their own training and competitions for predictive analysis. Coaches can also monitor each athlete, so as to know the athlete's physical data, as well as injury recovery. Sports scientists and trainers would also profit from their algebraic representation to identify optimal training programming without taking any simulations.

However, today's machine learning that only relies on big data analysis still has a little data difference, but with the addition of a neural network engine, the blessing of smart wearable devices can make the data more accurate.

REFERENCES

1. Alzamer H, Abuhmed T, Hamad K.: A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications. Electronics. 10(16), 1956 (2021). https://doi.org/10.3390/electronics10161956

2. Haas, A., Wäsche, H., Wittelsberger, R., Nieken, P., & Woll, A: Social Skills and sports: Pupils of an Elite School of sports are more competitive andcooperative. German Journal of Exercise and Sport Research (2022). https://doi.org/10.1007/s12662- 022-00827-w

3. Richard Butler: Sports psychology in action. CRC Press, London (2020). https://doi.org/10.4324/9781003058939

4. Cunniffe, E., Grainger, A., McConnell, W., Persson, U. M. C., Delahunt, E., Boreham, C., and Blake, C. A comparison of peak intensity periods across male field hockey competitive standards. Sports, 9(5), 58 (2021). https://doi.org/10.3390/sports9050058

5. Rory B. and Teo S.: The Application of Machine Learning Techniques forPredicting Match Results in Team Sport. Journal of Artificial Intelligence Research 73, 1285-1322 (2022). https://10.13140/RG.2.2.22427.62245

6. Puig-Diví, A., Escalona-Marfil, C., Padullés-Riu, J. M., Busquets, A., Padullés-Chando, X., and Marcos-Ruiz, D.: Validity and reliability of the Kinovea program in obtaining angles and distances using coordinates in 4 perspectives. PLOS ONE, 14(6), e0216448 (2019), https://doi.org/10.1371/journal.pone.0216448

7. Fury, M. S., Oh, L. S., & Berkson, E. M.: New Opportunities in Assessing Return to Performance in the Elite Athlete: Unifying Sports Medicine, Data Analytics, and Sports Science. Arthroscopy, Sports Medicine, and Rehabilitation, 4(5), 1897– 1902 (2022). https://doi.org/10.1016/j.asmr.2022.08.001.

8. Shaun Holthouse: Tracking balls in sports (9586099). Catapult Group International Pty Ltd.(2017),

9. https://patents.justia.com/patent/9586099.

10. Thabtah, F., Cowling, P., and Peng, Y.: A study of Predictive Accuracy for Four Associative Classifiers. Journal of Digital Information Management, 3(3), 202–205. Digital Information Research Foundation (2005).

11. Alexandra, M., and Florea, A. M.: Emotional Belief-Desire-Intention Agent Model: Previous Work and Proposed Architecture. International Journal of Advanced Research in Artificial Intelligence, 2(2) (2013), https://doi.org/10.14569/ijarai.2013.020201

12. Muazu Musa, R., P. P. Abdul Majeed, A., Taha, Z., Chang, S. W., Ab. Nasir, A. F., and Abdullah,
M. R.: A machine learning approach of predicting high potential archers by means of physical fitness indicators. PLOS ONE, 14(1), e0209638 (2019). https://doi.org/10.1371/journal.pone.0209638

13. Hoog Antink, C., Braczynski, A. K., and Ganse, B.: Learning from machine learning: prediction of age-related athletic performance decline trajectories. GeroScience, 43(5), 2547–2559 (2021). https://doi.org/10.1007/s11357-021-00411-4

14. Cortez A., Trigo A., Loureiro N.: Football Match Line-Up Prediction Based on Physiological Variables: A Machine Learning Approach. Computers 11(3): 40 (2022). https://doi.org/10.3390/computers11030040

15. ĆWiklinski, B., Giełczyk, A., and Choraś, M.: Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers. Entropy, 23(1), 90 (2021), https://doi.org/10.3390/e23010090

16. ABİDİN, D. (2021). A case study on player selection and Team Formation in football with machine learning. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTERSCIENCES, 1672–1691. https://doi.org/10.3906/elk-2005-27

17. Hong Y. and Li Y: Prediction of Football Match Results Based on Edge Computing and Machine Learning Technology. International Journal of Mobile Computing and Multimedia Communications 13(2), pp. 1-10 (2022). https://10.4018/IJMCMC.293749

18. Miljkovic, D., Gajic, L., Kovacevic, A., and Konjovic, Z.: The use of data mining for basketball matches outcomes prediction. IEEE 8th International Symposium on Intelligent Systems and Informatics (2010),

https://doi.org/10.1109/sisy.2010.5647440

19. Oytun, M., Tinazci, C., Sekeroglu, B., Acikada, C., and Yavuz, H. U.: Performance Prediction and Evaluation in Female Handball Players Using Machine Learning Models. IEEE Access, 8, 116321–116335 (2020). https://doi.org/10.1109/access.2020.3004182

20. Zhu, P., Sun, F. Sports Athletes' Performance Prediction Model Based on Machine Learning Algorithm. In: Abawajy, J., Choo, KK., Islam, R.,Xu, Z., Atiquzzaman, M. (eds) International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019. ATCI 2019. Advances in Intelligent Systems and Computing, vol 1017. Springer, Cham. https://doi.org/10.1007/978-3-030-25128-4_62 (2019)

21. Harsha Vardhan Goud, P. S., Mohana Roopa, Y.,& Padmaja, B.: Player performance analysis in sports: With fusion of Machine Learning and Wearable Technology. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (2019). https://doi.org/10.1109/iccmc.2019.8819815

22. Kumar, P. S., & Pranavi, S.: Performance Analysis of Machine Learning Algorithms on diabetes dataset using Big Data Analytics. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS) (2017). https://doi.org/10.1109/ictus.2017.8286062

23. Jiang, H., Qiu, T., & Deepa Thilak, K.: Application of deep learning method in automatic collection and processing of video surveillance data for basketball sports prediction. Arabian Journal for Science and Engineering (2021). https://doi.org/10.1007/s13369-021-05884-1

24. Kapadia, K., Abdel-Jaber, H., Thabtah, F., & Hadi,W.: Sport analytics for cricket game results using machine learning: An experimental study. Applied Computing and Informatics, 18(3/4), 256–266 (2020). https://doi.org/10.1016/j.aci.2019.11.006

25. Muazu Musa, R., Abdul Majeed, A. P., Suhaimi,M. Z., Mohd Razman, M. A., Abdullah, M. R., & Abu Osman, N. A.: Nature of volleyball sport, performance analysis in volleyball, and the recent advances of machine learning application in sports. Machine Learning in Elite Volleyball, 1–11 (2021). https://doi.org/10.1007/978-981-16-3192- 4_1

26. Namatēvs, I., Aleksejeva, L., & Poļaka, I.: Neural network modelling for Sports Performance Classification as a complex socio-technical system. Information Technology and Management Science, 19(1) (2016). https://doi.org/10.1515/itms-2016-0010

27. Arie-Willem de Leeuw, Stephan van der Zwaard, Rick van Baar & Arno Knobbe Personalized machine learning approach to injury monitoring in elite volleyball players, European Journal of Sport Science, 22:4, 511-520, DOI: 10.1080/17461391.2021.1887369 (2022)

28. Fitri M., Zainal A. N. E., Novan N. A., Kumalasari I., Haris F., Mulyana B., Khoo S., and Yaacob N.: Accessibility of Inclusive Sports Facilities for Training and Competition in Indonesia and Malaysia. Sustainability, 14(21), 14083 (2022). https://doi.org/10.3390/su142114083

29. Daniel Azizan: National Athletes' Perceived Image of the Government in the Development of Sports in Malaysia. Theses and Dissertations, 1514(2022).

30. Mohamed, A., Mazlan, D., & Osman, N. : Issues in Malaysian sports management. Recent Trends in Social and Behaviour Sciences, 393–396 (2014). https://doi.org/10.1201/b16658-69

31. Himeur Y., Elnour M., Fadli F., Meskin N., Petri I., Rezgui Y. Bensaali F., and Amira, A.: AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artificial Intelligence Review (2022). https://doi.org/10.1007/s10462- 022-10286-2