Prediction of player position for talent identification in association netball: a regression-based approach

Nur Hazwani Jasni [a,1], Aida Mustapha [a,2,*], Siti Solehah Tenah [b,3], Salama A Mostafa [a,4,*], Nazim Razali [a,5]

[a] Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, Malaysia
[b] Research Management Centre, University Tun Hussein Onn Malaysia, 86400 Johor, Malaysia
[1] nurhazwanijasni@gmail.com; [2] aidam@uthm.edu.my*; [3] solehah@uthm.edu.my; [4] salama@uthm.edu.my; [5] nazim.iium@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received August 23, 2021
Revised November 12, 2021
Accepted December 21, 2021
Available online March 31, 2022

Keywords
Talent Identification
Sports Analytics
Association Netball
Data Mining
Regression

ABSTRACT

Among the challenges in industrial revolutions, 4.0 is managing organizations' talents, especially to ensure the right person for the position can be selected. This study is set to introduce a predictive approach for talent identification in the sport of netball using individual player qualities in terms of physical fitness, mental capacity, and technical skills. A data mining approach is proposed using three data mining algorithms, which are Decision Tree (DT), Neural Network (NN), and Linear Regressions (LR). All the models are then compared based on the Relative Absolute Error (RAE), Mean Absolute Error (MAE), Relative Square Error (RSE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Relative Square Error (RSE). The findings are presented and discussed in light of early talent spotting and selection. Generally, LR has the best performance in terms of MAE and RMSE as it has the lowest values among the three models.

1. Introduction

In sports development, talent identification is a key element that aims to evaluate, recognize, and position certain sports players with specific skills and competencies to achieve competitive success during a game or match [1]. Identifying talents is crucial in developing a team with higher chances of winning at a higher level. The process involves predicting performance over a while to ensure consistency by measuring physical, physiological, cognitive, and sociological attributes. These attributes are then tested during games to see actual athletes' performance under pressure [2]. Overall, talent identification is not based on specific skills measured during the evaluation process but on how the skills are executed.

The era of the Fourth Industrial Revolution (IR 4.0) has begun to promote Artificial Intelligence (AI) techniques for automation in diverse work sectors, including sports [3], [4], and is expected to change how people live, work, and communicate as well as how values are assessed in the future [5]. In Malaysia, sports are still evaluated and managed based on traditional approaches. The coach will manually select the first seventh players in the netball team solely based on their own experience and knowledge. However, this method sometimes may cause a bias towards players since sometimes the coach has their favorite players, hence the potential for conflict of interest.

Whether in team sports or individual sports, recognizing outstanding at early stages and nurturing the talents is highly important [6]. Talent searching is a never-ending process in every association when forming a new team. Talent identification is not just assessing the athletes in an athletic event but also measuring specific and consistent physical and physiological requirements. Searching based on talent identification is better because the coach does not have to force someone to be his/her player in a team. Talent means that person already has a skill or ability about something they can do. This way, the coach has to focus on practices for their players to maintain momentum at every subsequent match or game.

Netball is a popular sport played [7] in more than 60 countries globally. The players consist of all girls only and have become the favorite game among girls nowadays [8]. Usually, netball matches are played between two teams, where each team is made of seven players with a maximum of 12 players per team. The seven positions in a netball team are Center (C), Wing Attack (WA), Wing Defense (WD), Goal Attack (GA), Goal Shooter (GS), Goal Defense (GD), and Goal Keeper (GK). All position has their basic techniques to be at those positions. These needed each other to be as a team to achieve the goals. A reference of skills required among netball players is shown in Table 1. These skills and movements in netball are high intensity, short bursts of movements but less intense, such as 2 to 3-meter sprints per time, jumping, pivoting, and catching. Netball requires the players to focus and maintain a high endurance, strength, speed, power, agility, and flexibility.

Table 1. Skills Required among Netball Players

Physical Skills	Mental Skills	Technical Skills
• speed • strength • agility	• imagery • concentration • self-talk	• jumping • throwing • catching • marking • sidestepping • landing • pivoting • shoot work • shooting

In netball, seven different positions consist of the center court players (C, WA, and WD), the shooters (GA and GS), and the defenders (GK and GD) as follows: 1) Centre (C) is the free player on the court. It is because it can play on any court part except in two-goal circles. These positions must be played by aerobically fit and have an excellent running ability. One of these centers' roles in court is taking the first pass after every second restart. The Centre assists the wing attack to move the ball to the goal shooters. The Center must also defend WA and WD, limit their influence, and closely aim to pressure the opposition when attacked with WD; 2) Wing attack (WA) must be played by those who have good attacking skills and be able to play smart because they must work purely to create scoring opportunities for the GA and GS by working as their defender over. The wing attackers are allowed in two attacks of thirds of the court, but they are not allowed to enter the goal circle as same as Centre (C). Usually, they attack the end of the center pass-through and work closely with the Centre to create as much scoring as possible. WA also defends the opposition WD to prevent counterattacks. They act as the cornerstone of the attack, a link between the defensive and the attacking of the ground; 3) Wing Defense (WD) is the last position in the court, and it works to defend two-thirds of the court without entering the defensive goal circle. It is also defending the opposition WA and limiting their influence wherever possible. As mentioned before, WD makes the attacking move to pass the ball quickly across the court to WA, C, and GA.; and 4) Goal Shooter (GS) only focuses on shooting goals within its perimeter.

Fig. 1 and Fig. 2 show a netball court's layout and schematic diagram. A netball court is divided into three equal sections, and only specific players can move in a specific section. Netball is dominant with a female sports game played by seven players. A score is counted when the ball goes through the goal ring. It is at the top 3-meter-high pole by only a GS or a GA. All players must not hold the ball for more than three seconds, and they are not allowed to run with the ball.

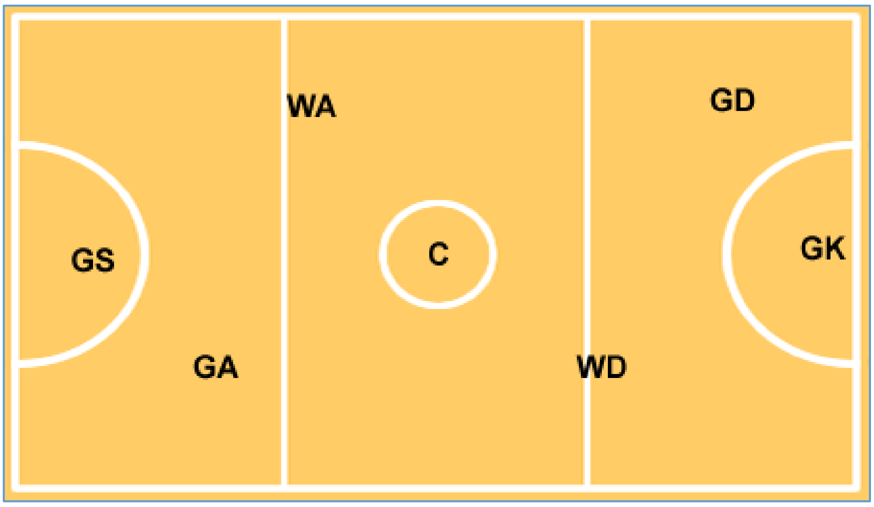

Fig. 1. The layout of a netball court.

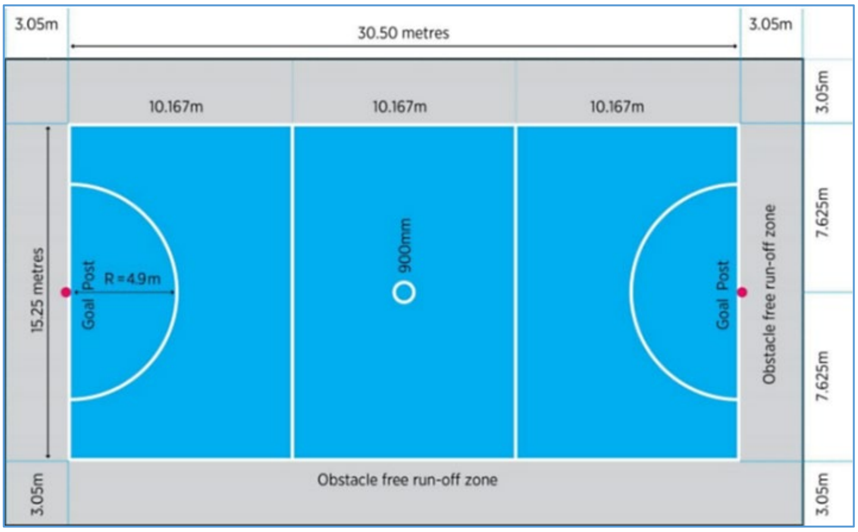

Fig. 2. Schematic diagram of the netball court.

This project attempts to predict all players' talent identification to determine different positions in netball [9], [10]. In achieving the research aim, the following objectives are specified. The first is to predict player position for talent identification in association netball. The second is to apply three algorithms for netball players' talent identification, which are Decision Tree (DT), Neural Network (NN), and Linear Regressions (LR). The third is to evaluate and compare the best model performance by calculating the correctness of the prediction models based on the error between the predicted value and actual value. This research is conducted by focusing on the player position for talent identification in association netball using Decision Tree (DT), Neural Network (NN), and Linear Regressions (LR).

The remaining of this paper proceeds as follows. Section 2 presents the details about the method or technique that has been used in the literature. Section 3 presents the research methodology, Section 4 presents and discusses the results, and finally, Section 5 concludes the study.

2. Method

2.1. Related Work

This section presents an overview of the existing performance prediction model and the respective measurements used to evaluate the models. Research in [11] used data mining techniques such as the Cultural Algorithm and Decision Tree to characterize each society's individuals based on their language, power, individualism, masculinity, and uncertainty avoidance. Jones et al. [12] employed a Neural

Vol. 8, No. 1, March 2022, pp. 84-96

Network model has to classify the giftedness among learners and extract the learning path in the distance learning environment using the K-Means Clustering algorithm. The results were evaluated by comparing the average of students and the gifted students distinguished by the neural network model. The ANOVA analysis showed that all *p*-values are less than 0.001 proving that gifted students are identified with better memory, evaluation, and higher cognitive and logic skills.

Meanwhile, Jantan et al. [13] studied the holistic development profile among the England cricket players using a machine learning approach via a classification experiment with Support Vector Machine (SVM), Multilayer Perceptron (MLP), Naïve Bayes, and Nearest Neighbor. The study reported that SVM produced the highest accuracy of 92.9% and excelled in predicting competitive performance categories using developmental features. In another study [14], the researchers predicted and classified the employees' performance patterns using the Decision Tree classification algorithm based on the 33 selected attributes. Their experimental results showed that C4.5 achieved the highest classification accuracy, at 95.08%.

Finally, [15] predicted employee performance using three classification techniques, which were C4.5, Naïve Bayes, and Support Vector Machine. The result showed that SVM was the most suitable classifier and had the most excellent prediction accuracy, 86.9% using WEKA tools in executing the experiment. Table 2 summarizes related works in terms of data mining approaches and techniques.

Table 2. Summary of Related Work

Author(s)	Approaches	Method / Algorithms
Ochoa et al. [6]	Classification	Cultural Algorithms, Decision Tree
Bael and Parka [11]	Clustering, Classification	k-Means Clustering Algorithm
Jones et al. [12]	Classification	Support Vector Machine (SVM), Multilayer Perceptron (MLP), Naïve Bayes, Nearest Neighbor
Jantan et al. [13]	Classification, Prediction	Decision Tree
Nasr et al. [14]	Classification	C4.5, Naïve Bayes, SVM

2.2. Knowledge Discovery in Databases (KDD) Methodology

This section presents the methodology used in completing this project [16]. The research methodology chosen for the project is the Knowledge Discovery in Databases (KDD) methodology. Knowledge Discovery in Databases (KDD) methodology is a type of data mining methodology that describes extracting knowledge from the historical dataset using specific data mining methods [17]. Fig. 3 shows the KDD methodology extracted from [17].

The KDD methodology is used in many research areas such as Data Mining, Pattern Recognition, Machine Learning, Expert Systems, and Data Visualization. Note that the underlying principle in KDD is the step-by-step process of transforming data into knowledge in the context of large databases. Based on Fig. 3, the KDD methodology begins with the selection phase, preprocessing, transformation, data mining phases, interpretation or evaluation phases, and knowledge.

2.2.1. Data Selection Phase

The netball dataset is selected in the data selection phase. Data selection aims to generate a target dataset from the large, initial database [18]. As previously mentioned, the data is organized based on the best algorithm analyzed using Microsoft Azure tools. The dataset has been selected for the process from a large database. It is usually referred to as target data. It consists average of 500 data of each place that was described by age, height, body mass, shooting, marking, catching, throwing, jumping, and landing data will be analyzed. All the features in the data are related to netball skills.

2.2.2. Data Cleaning Phase

In this phase, the dataset is cleaned and checked for missing values. All missing values are replaced with the mean value of the feature as provided by the Azure tool. This phase also involved finding hidden data correlations among the variables and identifying the most useful variables from the dataset that can give the best prediction models [19].

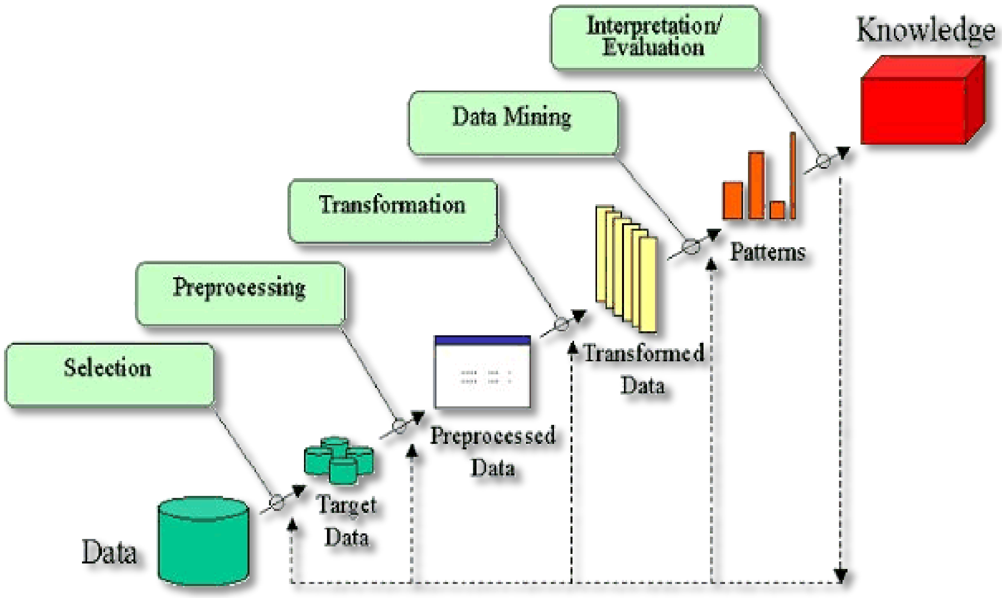

Fig. 3. Knowledge Discovery in Databases (KDD) process [17].

2.2.3. Choosing the Appropriate Data Mining Task

In this research, the type of data mining task used is prediction using the regression model. The prediction task maps the netball datasets to predefine the targets [20]. The prediction aims to predict some cases with attributes to classify the objects [21].

2.2.4. Choosing the Data Mining Algorithm

Once the data mining task has been chosen, this step is to select the appropriate algorithms to be used for building the prediction or classification models [22]. In this project, the Decision Tree, Neural Network, and Linear Regression algorithm have been chosen based on the literature reviews and are ready for the next steps.

2.2.5. Implementation of the Data Mining Algorithm [23]

Finally, the algorithms are employed to build the experiments' regression models in this stage.

2.2.6. Evaluation

In this stage, the performance of the classification or prediction models is evaluated based on the error measures by comparing the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE) and Coefficient of Determination (R^2) between the three models.

2.2. Datasets

The dataset used in this study is sourced from the international (Thomas) association netball. A total of 500 pieces of data have been selected, consisting of 9 attributes: height, body mass, shooting, marking, catching, throwing, pivoting, jumping, and landing. The main goal is to find the best algorithm to predict the netball player's position with the smallest error measure. The excerpt of the netball players' dataset is shown in Fig. 4.

1	Age	Height	Body Mass	Shooting	Marking	Catching	Throwing	Jumping	Landing
2	15	164	59	1.70	1.65	5.10	10.1	6.27	5.72
3	13	176	64	1.31	1.48	4.19	10.1	5.99	4.93
4	15	177	63	1.57	1.64	4.84	9.9	5.59	5.86
5	15	178	61	1.55	1.48	4.88	9.3	6.04	4.90
6	15	168	65	1.49	1.65	5.04	9.7	5.10	5.45
7	15	178	65	1.74	1.59	4.87	9.4	4.19	5.55
8	14	167	60	1.87	1.95	6.25	9.5	4.84	5.81
9	14	175	69	1.64	1.56	5.08	9.6	4.88	6.15
10	14	169	59	1.96	2.01	6.38	8.9	5.04	5.44
11	16	170	64	1.63	1.54	4.90	7.9	4.87	5.68
12	13	177	68	1.62	1.78	5.32	9.0	6.25	5.74
13	15	166	63	1.53	1.46	5.01	7.4	5.08	5.68
14	14	177	60	1.65	1.82	5.27	7.1	6.38	6.04
15	14	165	63	1.56	1.54	5.00	7.0	4.90	5.12
16	15	164	67	1.26	1.70	4.30	6.7	5.32	5.95
17	16	167	60	1.93	1.94	6.05	6.6	5.01	5.83
18	14	169	59	1.72	1.69	5.41	5.9	5.27	5.91

Fig. 4. The excerpt of netball players.

2.2. Algorithms

This section presents the details of the prediction algorithms used in this research project. In this research project, Decision Tree (DT), Neural Network (NN), and Linear Regression (LR) are selected for the regression approach.

Decision Tree (DT). A DT is a collection of nodes that form a tree with a root. A root is a node with no incoming edges [24]. The algorithm is based on a sequence of information nodes that will branch out to different outcomes, as illustrated in Fig. 5. In such settings, decision trees [25] work by predicting each incoming x-data case into one of the class labels for the outcome by traversing the tree [26]. It applies to data where the y-value is a class label, and the dependent variables use x-variables available in the data.

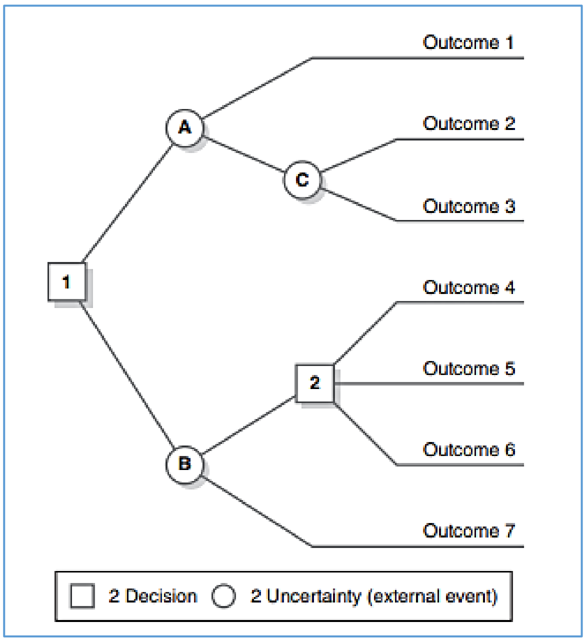

Fig. 5. Example of Decision Tree [19].

Neural Networks (NN). NN works by mimicking the human brain. Conceptually, it is inspired by the neural networks in human brains, but the implementation in Machine Learning is only an expectation and far from reality. Machine learning works by taking multiple inputs and then producing a single output, as shown in Fig. 6.

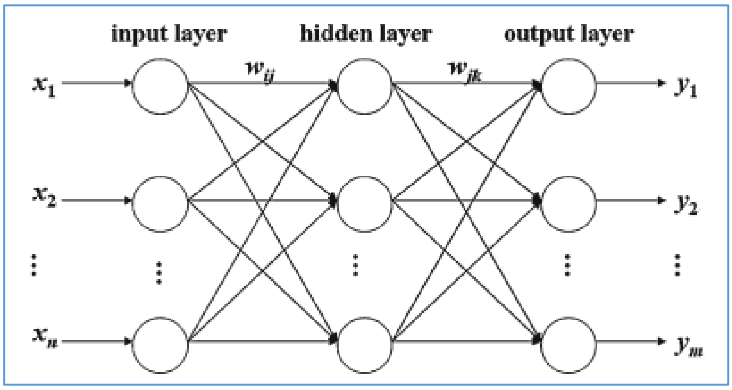

Fig. 6. The topology of Neural Network [27].

Most Neural Networks (NN) used the layers (three or more) to present the data to produce the output in response to a given input. Meanwhile, intermediate laters are used to model other features [27]. In machine learning, Neural networks can emerge data as one of the most important for prediction. Neural networks in prediction methods in real-life data are challenging due to the massive size of the data, high dimensionality, and the presence of seasonal variations [28].

Linear Regression (LR). An LR can remember a line or define it as a linear approach [29] to modeling the relationship between dependent and independent variables, as shown in Fig. 7. It is based on the least square estimation and should be linear.

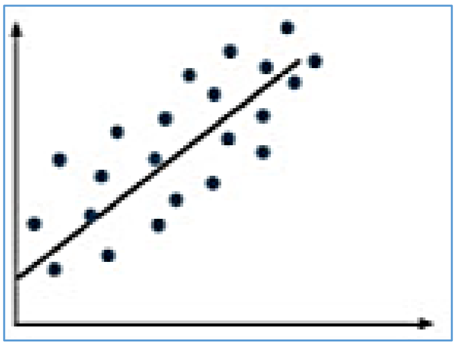

Fig. 7. The topology of Neural Network [27].

LR then creates a trendline based on the data that have been plotted. The trendline can show factors apart from the correlation between the two variables [30]. The best fit line is when the errors between the predicted and observed values are minimal.

2.3. Evaluation Metrics

The evaluation methodology can be categorized based on threshold techniques, probability techniques, or ranking techniques [31]. In this research project, evaluation is critical in order to compare the performance of all three prediction models, which are Decision Tree (DT), Neural Network (NN), and Linear Regression (LR) [32]. The performance of the prediction model is evaluated based on error measures in predicting the position of netball players in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE), and Coefficient of Determination (R^2).

To illustrate the formula for all metrics, consider a as the actual target and p as the predicted target. Equation (1) to Equation (8) shows the formula for each evaluation metric used in this study. Note that the Coefficient of Determination (R^2) is composed of four equations from (4) to (8), which include the Sum of Squares Total, Sum of Squares Regression, Sum of Squares Error, and finally, the Coefficient of Determination (R^2).

- Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{i=1}^{n} | p_i - a_i |}{n} \tag{1}$$

- Root Mean Squared Error (RMSE) [33]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}} \tag{2}$$

- Relative Absolute Error (RAE)

$$RAE = \frac{\sum_{i=1}^{n} | p_i - a_i |}{\sum_{i=1}^{n} | \bar{a} - a_i |} \tag{3}$$

- Relative Squared Error (RSE)

$$RSE = \frac{\sum_{i=1}^{n} (p_i - a_i)^2}{\sum_{i=1}^{n} (\bar{a} - a_i)^2} \tag{4}$$

- Coefficient of Determination ($\boldsymbol{R^2}$)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{5}$$

Where, Sum of Squares Total (SST), Sum of Squares Regression (SSR), and Sum of Squares Error (SSE) are given in (6) to (8).

$$SST = \sum (y - \bar{y})^2 \tag{6}$$

$$SSR = \sum (\grave{y} - \bar{y})^2 \tag{7}$$

$$SSE = \sum (y - \grave{y})^2 \tag{8}$$

3. Results and Discussion

This paper focused on the Decision Tree (DT), Neural Network (NN), and Linear Regression (LR) algorithm. The performance analysis is done to find the error measures of the models. Experiments were carried out using the Azure Machine Learning tool on 500 data with 9 different attributes, and the results were compared. The error measures used are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE), and Coefficient of Determination (R^2).

The best outcomes for Decision Tree were obtained when the instruction was split by 60% instruction and 60% research by achieving the lowest RMSE, and it means MAE since performance is most accurate when the errors are low. The decision tree is considered the best model for this method, as shown in Table 3 and the Error Histogram in Fig. 8.

The best performances for Neural Network were obtained when the instruction was split by 60% instruction and 60% research. It has the lowest possible to get a significant error. It is more accurate in predicting the position of netball players, as shown in Table 4 and the Error Histogram in Fig. 9 as the best result for this methodology.

Table 3. Result for Decision Tree Algorithm

Split	MAE	RMSE	RAE	RSE	R^2
(50, 70)	12.449931	15.83025	0.971146	1.079859	-0.079859
(60, 60)	**12.126708**	**15.747484**	**0.992057**	**1.154336**	**-0.154336**
(70, 50)	12.747655	16.3529	1.12261	1.429377	-0.429377

Table 4. Result for Neural Network Algorithm

Split	MAE	RMSE	RAE	RSE	R^2
(50, 70)	12.693134	15.162538	0.990117	0.990684	0.009316
(60, 60)	**12.008018**	**14.289889**	**0.982347**	**0.950534**	**0.049466**
(70, 50)	12.835259	15.64378	1.130377	1.308099	-0.308099

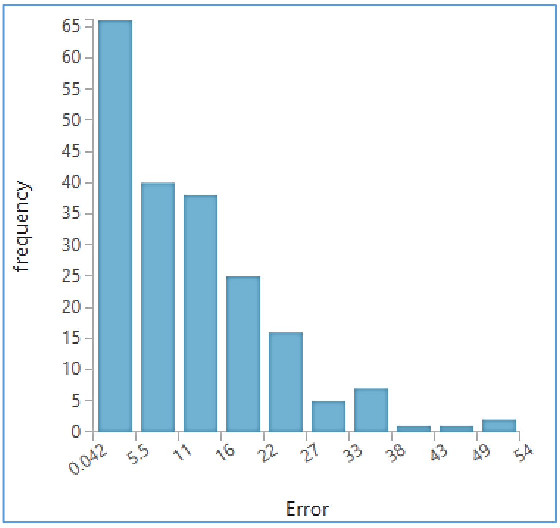

Fig. 8. Error Histogram for Decision Tree in split data (60, 60).

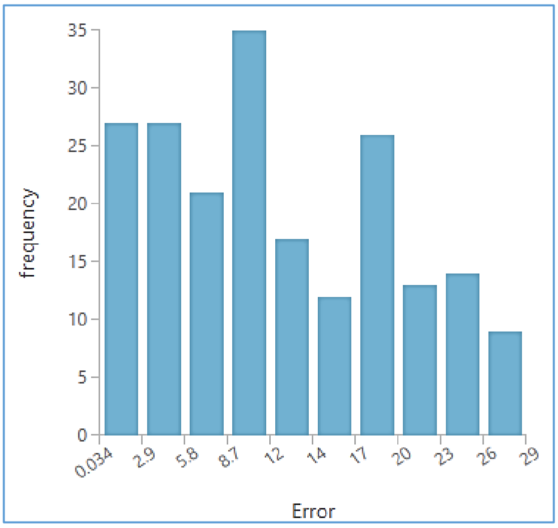

Fig. 9. Error Histogram for Neural Network in split data (60, 60).

The linear Regression algorithm is the most accurate algorithm when predicting using a split at 70% of instructions and 50% of research. It has achieved the best result for this methodology at the lowest MAE and RMSE. Table 5 and Error Histogram in Fig. 10 show the result for the Linear Regression algorithm.

The Linear Regression algorithm performs better than the Decision Tree and Neural Network. Linear Regression is best because it is the least complex to compare with other algorithms that only try to find the relationship between the independent and dependent variables. Using Linear Regression proves that it is more accurate for predicting the model than other algorithms, as shown in Table 5. Linear Regression has the lowest possibility for error because the MAE, RMSE, RAE, RSE, and Coefficient of Determination (R^2) are the lowest value, as shown in Fig. 10.

Table 5. Result for Linear Regression Algorithm

Split	MAE	RMSE	RAE	RSE	R^2
(50, 70)	11.971853	14.423889	0.933854	0.896512	0.103488
(60, 60)	11.755705	14.216148	0.961706	0.940749	0.059251
(70, 50)	**11.529811**	**14.212382**	**1.015408**	**1.079670**	**-0.079670**

Neural Network results are higher than Linear Regression but lower than the Decision Tree algorithm. It is proven that the possibility of error when predicting the model using a Neural Network is high, as shown in Table 4. It may cause the accuracy in predicting the model is decreasing and may get a bad result. The result is shown in Fig. 9 when used in split data (60, 60) using Neural Network Regression. Decision Tree comes with the result of the highest performance when testing a Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE), and Coefficient of Determination (R^2). It is mean; the result shows the lower accuracy when predicting using a Decision Tree because the possibility of having an error while used is big because of the highest Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE) and Coefficient of Determination (R^2) rather than using Neural Network Regression and Linear Regression. It is proven in Fig. 8, Error Histogram when the frequency for having an error when using Decision Tree Regression is 65.

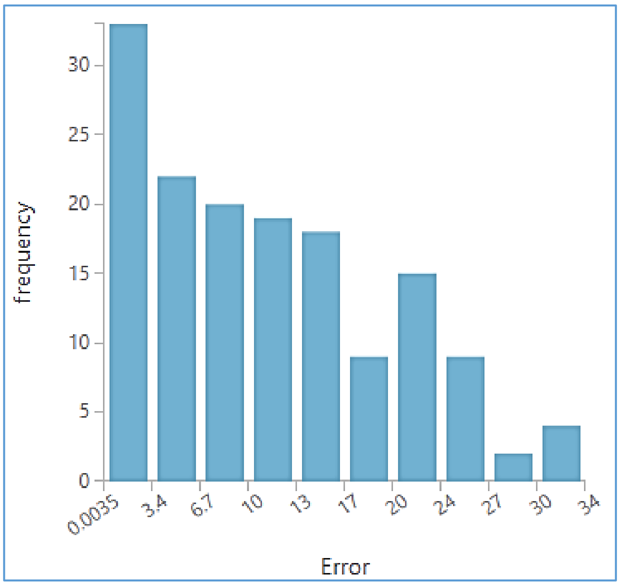

Fig. 10. Error Histogram for Linear Regression in split data (70, 50).

Note that the primary purpose of this experiment is to obtain a comparison of performance between Decision Tree Regression, Neural Network Regression, and Linear Regression the result are presented in Table 6 shows the difference in Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Relative Square Error (RSE) and Coefficient of Determination (R^2) to train the model using Throwing, one of the dataset's training modules. Overall, Table 6 shows the final performance comparison between DT, NN, and LR.

Table 6. Comparison of Performance between DT, NN, and LR

	MAE	RMSE	RAE	RSE	R^2
DT	12.126708	15.747484	0.992057	1.154336	-0.154336
NN	12.008018	14.289889	0.982347	0.950534	0.049466
LR	**11.529811**	**14.212382**	**1.015408**	**1.079670**	**-0.079670**

An additional technique for summarizing the performance of a prediction algorithm is given in Fig. 11 and Fig. 12, which gives a better understanding of different types of errors across all three algorithms used. The comparison between the three algorithms proves that Linear Regression is more accurate than other algorithms used in this experiment because it has the lowest MAE and RMSE values.

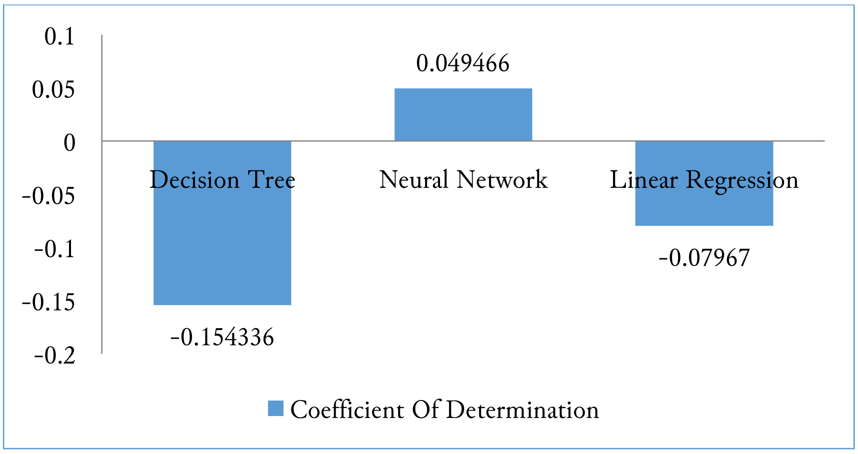

Fig. 11. Comparison of Coefficient of Determination (R^2)

Fig. 12. Comparison of MAE, RMSE RAE, and RSE

4. Conclusion

This work presented three regression models designed to forecast netball players' positions during matches. To be more specific, this netball dataset has been used to predict whether talented players' are suited to their position in netball. The challenge while doing this experiment is the dataset that must clean before testing the dataset. Then another challenging part is choosing the prediction algorithm for this research suitable for the netball dataset. The comparison using three different algorithms, Decision Tree, Neural Network, and Linear Regression, could be used to define the best algorithms that work with datasets. Thus, the result confirms that the Linear Regression algorithm may apply in predicting the position of netball players in association netball in terms of error measures. This research may help the coach choose the player for the position in netball. The presented framework also has the potential to be extended into other team-based sports such as badminton or basketball.

Acknowledgment

This paper is supported by Research Fund E15501, Research Management Centre, Universiti Tun Hussein Onn Malaysia (UTHM).

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] N. Razali, A. Mustapha, F. A. Yatim, and R. Ab Aziz, "Predicting Player Position for Talent Identification in Association Football," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 226, p. 012087, Aug. 2017, doi: 10.1088/1757-899X/226/1/012087.

[2] R. M. Malina, "Early Sport Specialization: Roots, Effectiveness, Risks," *Curr. Sports Med. Rep.*, vol. 9, no. 6, pp. 364–371, Nov. 2010, doi: 10.1249/JSR.0b013e3181fe3166.

[3] Y. Galily, "Artificial intelligence and sports journalism: Is it a sweeping change?," *Technol. Soc.*, vol. 54, pp. 47–51, Aug. 2018, doi: 10.1016/j.techsoc.2018.03.001.

[4] R. R. Nadikattu, "Implementation of New Ways of Artificial Intelligence in Sports," *J. Xidian Univ.*, vol. 14, no. 5, pp. 5983–5997, 2020, doi: 10.2139/ssrn.3620017.

[5] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Commun. ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996, doi: 10.1145/240455.240464.

[6] A. Ochoa, A. Hernández, J. Sánchez, A. Muñoz-Zavala, and J. Ponce, "Determining the Ranking of a New Participant in Eurovision Using Cultural Algorithms and Data Mining," *18th Int. Conf. Electron. Commun. Comput. (conielecomp 2008)*, pp. 47–52, Mar. 2008, doi: 10.1109/CONIELECOMP.2008.27.

[7] C. Downs, S. J. Snodgrass, I. Weerasekara, S. R. Valkenborghs, and R. Callister, "Injuries in Netball-A Systematic Review," *Sport. Med. - Open*, vol. 7, no. 1, p. 3, Dec. 2021, doi: 10.1186/s40798-020-00290-7.

[8] S. Whitehead *et al.*, "The Applied Sports Science and Medicine of Netball: A Systematic Scoping Review," *Sport. Med.*, vol. 51, no. 8, pp. 1715–1731, Aug. 2021, doi: 10.1007/s40279-021-01461-6.

[9] A. H. Mohamad, R. Ramli, and A. F. Ramli, "A Software Engineering Approach in Netball Performance Analysis: Training and Activities Features for Automatic Players Position Selection," *2020 8th Int. Conf. Inf. Technol. Multimed.*, pp. 371–377, Aug. 2020, doi: 10.1109/ICIMU49871.2020.9243476.

[10] E. R. Brooks, A. C. Benson, A. S. Fox, and L. M. Bruce, "Physical movement demands of elite-level netball match-play as measured by an indoor positioning system," *J. Sports Sci.*, vol. 38, no. 13, pp. 1488–1495, Jul. 2020, doi: 10.1080/02640414.2020.1745504.

[11] S. Bae, S. H. Ha, and S. C. Park, "Identifying gifted students and their learning paths using data mining techniques," *Data Min. E-Learning*, pp. 191–206, Jun. 2006, doi: 10.2495/1-84564-152-3/11.

[12] B. D. Jones *et al.*, "The Identification of 'Game Changers' in England Cricket's Developmental Pathway for Elite Spin Bowling: A Machine Learning Approach," *J. Expert.*, vol. 2, no. 2, pp. 92–120, 2019. Available: Google Scholar.

[13] H. Jantan, A. R. Hamdan, and Z. A. Othman, "Human Talent Prediction in HRM using C4.5 Classification Algorithm," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 8, pp. 2526–2534, 2010. Available: Google Scholar.

[14] M. Nasr, E. Shaaban, and A. Samir, "A proposed Model for Predicting Employees' Performance Using Data Mining Techniques: Egyptian Case Study," *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 1, pp. 31–40, 2019. Available: Google Scholar.

[15] C. Combes, N. Meskens, C. Rivat, and J.-P. Vandamme, "Using a KDD process to forecast the duration of surgery," *Int. J. Prod. Econ.*, vol. 112, no. 1, pp. 279–293, Mar. 2008, doi: 10.1016/j.ijpe.2006.12.068.

[16] L. Subramainan, M. Z. M. Yusoff, and M. A. Mahmoud, "A classification of emotions study in software agent and robotics applications research," *2015 Int. Symp. Agents, Multi-Agent Syst. Robot.*, pp. 41–46, Aug. 2015, doi: 10.1109/ISAMSR.2015.7379128.

[17] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)," *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014. Available: Google Scholar.

[18] Y. Ben-Haim and E. Tom-Tov, "A Streaming Parallel Decision Tree Algorithm," *J. Mach. Learn. Res.*, vol. 11, no. 28, pp. 849–872, 2010, [Online]. Available: http://jmlr.org/papers/v11/ben-haim10a.html.

[19] A. Azevedo and M. F. Santos, "KDD, semma and CRISP-DM: A parallel overview," *IADIS Eur. Conf. Data Min.*, 2008. Available: Google Scholar.

[20] S. Bagga and D. G. N. Singh, "Conceptual Three Phase Iterative Model of KDD," *Int. J. Comput. Technol.*, vol. 2, no. 1, pp. 6–8, Feb. 2012, doi: 10.24297/ijct.v2i1.2605.

[21] S. J. Delany, P. Cunningham, D. Doyle, and A. Zamolotskikh, "Generating Estimates of Classification Confidence for a Case-Based Spam Filter," *Muñoz-Ávila, H., Ricci, F. Case-Based Reason. Res. Dev. ICCBR 2005. Lect. Notes Comput. Sci.*, vol. 3620, pp. 177–190, 2005, doi: 10.1007/11536406_16.

[22] J. A. Lara, D. Lizcano, M. A. Martínez, and J. Pazos, "Data preparation for KDD through automatic reasoning based on description logic," *Inf. Syst.*, vol. 44, pp. 54–72, Aug. 2014, doi: 10.1016/j.is.2014.03.002.

[23] S. Lotfi and M. Rebbouj, "Machine Learning for sport results prediction using algorithms," *Int. J. Inf. Technol. Appl. Sci.*, vol. 3, no. 3, pp. 148–155, Aug. 2021, doi: 10.52502/ijitas.v3i3.114.

[24] M. Mandorino, A. J. Figueiredo, G. Cima, and A. Tessitore, "A Data Mining Approach to Predict Non-Contact Injuries in Young Soccer Players," *Int. J. Comput. Sci. Sport*, vol. 20, no. 2, pp. 147–163, Dec. 2021, doi: 10.2478/ijcss-2021-0009.

[25] L. Rokach and O. Maimon, "Decision Trees," *Maimon, O., Rokach, L. Data Min. Knowl. Discov. Handbook. Springer, Boston, MA*, pp. 165–192, 2005, doi: 10.1007/0-387-25465-X_9.

[26] K. P. Sudheer, A. K. Gosain, D. Mohana Rangan, and S. M. Saheb, "Modelling evaporation using an artificial neural network algorithm," *Hydrol. Process.*, vol. 16, no. 16, pp. 3189–3202, Nov. 2002, doi: 10.1002/hyp.1096.

[27] S. A. Mostafa *et al.*, "Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease," *Cogn. Syst. Res.*, vol. 54, pp. 90–99, May 2019, doi: 10.1016/j.cogsys.2018.12.004.

[28] R. R. Hocking, "A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, no. 1, p. 1, Mar. 1976, doi: 10.2307/2529336.

[29] T. Horvat and J. Job, "The use of machine learning in sport outcome prediction: A review," *WIREs Data Min. Knowl. Discov.*, vol. 10, no. 5, Sep. 2020, doi: 10.1002/widm.1380.

[30] P. Chainok *et al.*, "Modeling and predicting the backstroke to breaststroke turns performance in age-group swimmers," *sport. Biomech.*, pp. 1–22, Dec. 2021, doi: 10.1080/14763141.2021.2005127.

[31] J. Stübinger, B. Mangold, and J. Knoll, "Machine Learning in Football Betting: Prediction of Match Results Based on Player Characteristics," *Appl. Sci.*, vol. 10, no. 1, p. 46, Dec. 2019, doi: 10.3390/app10010046.

[32] W. Chen, D. Sharifrazi, G. Liang, S. S. Band, K. W. Chau, and A. Mosavi, "Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit," *Eng. Appl. Comput. Fluid Mech.*, vol. 16, no. 1, pp. 965–976, Dec. 2022, doi: 10.1080/19942060.2022.2053786.

[33] N. . Saravana Kumar, K. Hariprasath, N. Kaviyavarshini, and A. Kavinya, "A study on forecasting bigmart sales using optimized machine learning techniques," *Sci. Inf. Technol. Lett.*, vol. 1, no. 2, pp. 52–59, Nov. 2020, doi: 10.31763/sitech.v1i2.167.