

Research Article

Cricket Players Selection for National Team and Franchise League Using Machine Learning Algorithms

Md. Robel¹, Md. Ashikur Rahman Khan^{1*} , Ishtiaq Ahammad² , Md. Mahbubul Alam¹, Kamrul Hasan¹

¹Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh

²Department of Computer Science and Engineering, Northern University Bangladesh, Dhaka, Bangladesh

E-mail: ashik@nstu.edu.bd

Received: 7 October 2023; **Revised:** 27 November 2023; **Accepted:** 27 November 2023

Abstract: Cricket player selection is a crucial task for both national teams and franchise leagues. Traditionally, selectors rely on their experience and knowledge to evaluate a player's physical fitness, batting, and bowling performance. However, with the advancements in machine learning algorithms, it is possible to automate and improve the selection process. In this research, a Machine Learning (ML)-based approach is proposed for cricket player selection. This approach uses a combination of physical fitness data, batting and bowling statistics, and other relevant metrics to create a comprehensive player profile. Then, three ML algorithms—linear regression, support vector regression, and random forests—are employed to identify the most promising players. Data is collected on a large number of cricket players and their performance in national and franchise leagues from the two most prominent cricket websites, ESPNcricinfo and Cricbuzz, to evaluate the proposed approach. Then, the proposed approach trains and tests the ML models on this data and compares their accuracy and performance. Based on the performance scores obtained from these models, two squads are selected for the national team and one squad for each franchise league team. The results demonstrate that the proposed approach can significantly improve the selection process and identify players with high potential. Furthermore, the results found that the support vector regression algorithm outperformed other ML models in terms of prediction and player selection.

Keywords: machine learning, cricket, player selection, national team selection, franchise league

1. Introduction

There are usually billions of dollars invested in the sports sector [1]. In the modern sporting world, a vast array of coaches, specialists, physiotherapists, dietitians, and trainers support athletes and teams internationally [2]. Among the most well-known sports on this planet, cricket is the one. A bat and ball sport, cricket is practiced in 106 International Cricket Council (ICC) member countries and has grown into a multi-billion-dollar industry [3]. The Twenty-Twenty Cricket (T20) format, One Day International (ODI) cricket, and Test cricket are the three main formats played in cricket. Because both the ODI and T20 formats are evolving, cricket has quickly gained popularity as a team sport [4]. Given this rise in popularity, franchise T20 leagues have formed in many countries around the world.

Each successful cricket team now relies heavily on data analysis. The results of cricket data analysis offer a deeper understanding of the players as well as the game, which is particularly beneficial to those associated with the sport, including existing players, technical experts, coaches, and young players [5, 6]. Cricket administrators routinely look for novel solutions to improve the performance of their cricket team while giving themselves an edge over others because of how quickly cricket has evolved. Maximizing player performance whilst reducing the chance of injury is related to player performance administration [7]. The analysis of cricket data is crucial to this procedure. Because of many restrictions, data analysis for sports was unable to fully rely on conventional statistical methods. Assumptions regarding the data are typically the foundation of standard statistical procedures, and conventional cricket data and analysis might fail to satisfy the necessary assumptions. The choice of data analysis methodologies is of utmost importance because cricket-related activities are not autonomous and are often affected by a variety of human variables [8]. Additionally, because there are so many interrelated variables, the majority of conventional metrics are unable to account for modern game assumptions.

Arguably the most crucial element in cricket that can impact how well a squad performs is the team selection. Cricket player selection has historically depended on the skills and expertise of coaches and selectors, who made proper choices based on players' ability, physical makeup, and prior performance. Due to its length and subjectivity, this traditional player selection process, however, is possibly open to biases and errors. Machine Learning (ML) techniques offer a data-driven approach to select players, enabling analysis of past player data to identify trends and project potential outcomes. ML models might be educated on a range of data sources like player information, fitness data, and tournament circumstances, to identify the key factors that typically affect a player's performance.

Sports Analytics benefit greatly from the expansion and acceptance of ML algorithms [9]. The Internet of Things (IoT) has made technological strides that allow sports analysts to collect data more quickly and precisely [10]. The recognition of ML approaches employed for sports data analysis has skyrocketed thanks to the quantity of computing resources as well as data. The influence of ML strategies on data collection [11], data extraction from edge devices [12], as well as the analysis of the collected data to better understand the edge users [13, 14] has increased the effectiveness of ML approaches. The majority of these technological devices that are used for data collection are wearable [15, 16]. The combination of these wearables with ML techniques has made it possible to create AI expert systems within the sports domain [17]. These ML-based solutions swiftly learned sports data analysis to build models and forecast future results using sports data already available to give precise predictions to improve sports decision-making. Therefore, across every sport, data analysts utilize ML approaches to carry out a variety of activities, from squad selection to player performance assessment, and to increasing their team's possibilities of winning [18-20].

The primary goal of ML approaches is to replace time-consuming, and laborious human operations with an increased degree of mechanization in the knowledge development process and therefore, they are used in various sectors like education [21-23], finance [24], medicine and healthcare [25, 26], and clustering [27, 28]. These ML solutions must be developed with domain-specific expertise. The procedure used to improve ML strategies is known as feature engineering. This is done to maximize the efficiency of the ML strategies by extracting features from unprocessed information with the aid of domain-specific knowledge. Although there is no shortage of sports-related data, implementing ML to big data raises interesting questions because it necessitates domain-specific expertise, the implementation of learning techniques, as well as software engineering [29].

In the "Literature Review" chapter, this paper conducted a comprehensive review on predicting player performance in cricket. There are only a few works that might predict a squad for the national team. But their research doesn't include franchise league squad prediction. According to this study, there is no such research that solely focuses on predicting the squad for the national team, as well as the franchise league. In addition, these previous studies have considered a very limited number of parameters for performance prediction. These limitations motivated us to address this research gap by developing an ML-based approach that considers a broader range of performance metrics and predicts squads for national teams as well as franchise leagues using ML algorithms.

This research is motivated by the need to update and improve the traditional methods of selecting cricket players, taking into account the constraints and potential biases present in human decision-making processes. We are driven to make use of ML algorithms' potential as a formidable instrument to improve the player selection process' objectivity and efficiency in this era of rapid technological breakthroughs. The abundance of information accessible, including measurements related to physical fitness, batting and bowling statistics, and other relevant data, presents a chance to

develop a detailed and sophisticated player profile. The dynamic nature of modern cricket, particularly in light of the T20 leagues' recent rise in popularity and the ensuing changes to the rules and methods of the game further motivates this research. Our goal is to provide a new approach that fits the changing demands of cricket by expanding the use of ML applications beyond individual player performance to squad predictions for national teams and franchise leagues, thereby, ultimately improving the accuracy and equity of player selection procedures.

To create a well-balanced cricket team, the captain, wicketkeeper, batsman, and bowler must be selected based on their strengths and capabilities. To accomplish this, this research developed a detailed methodology for selecting the best players for each position, which considers various parameters and equations to assess each player's suitability. By selecting the best player for each position, we aimed to improve the team's overall performance. Data is collected from prominent cricket websites, *espnricinfo*, and *cricbuzz*, to analyze the physical fitness characteristics of cricket players and calculate their fitness scores. Using these scores, a group of players is selected with the highest fitness scores. We then calculated each player's overall performance, which is a combination of batting and bowling statistics. Using statistical methods, we computed separate scores for batting and bowling performance and combined them using a statistical formula to determine the overall performance.

To predict player performance, three different ML algorithms were utilized: Linear Regression (LR), Support Vector Regression (SVR), and Random Forest (RF). These algorithms are used to predict the fitness, batting, and bowling performance scores of each player, and then integrated these datasets to predict the players' overall performance. Finally, these performance values are used to select the best players for both national teams and franchise leagues. In summary, the major contributions of this research are:

- Conducted a comprehensive literature review on predicting player performance in cricket. The review shows limited works predict squads for the national team, and none include franchise league squad prediction.
- Addressed the identified research gap by developing an ML-based approach for predicting squads for both national teams and franchise leagues in cricket.
- Overcame limitations of previous studies by considering a substantial number of parameters for measuring player performance, enhancing the predictive accuracy of the model.
- Developed a detailed methodology for selecting players for specific positions (captain, wicketkeeper, batsman, and bowler) based on individual strengths and capabilities.
- Collected data from reputable cricket websites, *espnricinfo*, and *cricbuzz*, to analyze the physical fitness characteristics of cricket players and calculate their fitness scores.
- Implemented three different ML algorithms (LR, SVR, and RF) to predict fitness, batting, and bowling performance scores for each player. Integrated ML algorithm outputs predict the overall performance of players, providing a comprehensive assessment of their abilities.
- Utilized predicted performance values to select the best players for both national teams and franchise leagues, showcasing the practical application of the developed methodology.

Overall, the research represents a substantial contribution to the field of cricket player selection by introducing a comprehensive methodology that considers multiple performance parameters and leverages ML algorithms to improve squad predictions for diverse cricket contexts.

2. Literature review

Currently, player predictions are an extremely crucial and difficult duty in any kind of sport, particularly cricket. From a squad of 15 to 20 players, the team officials, coach, and captain must select the top eleven players for the specific tournament or series. Normally, they have primary concerns regarding the player's recent performance versus the team, location, or a particular series. Therefore, analyzing every player to select the best becomes a difficult assignment for a team selector. As a result, numerous researchers have attempted to forecast the performance of players using ML techniques. In reality, ML has been applied to every aspect of sports. ML algorithms are employed in a wide range of applications, including self-awareness [30], the application of medicine [31], player's performance assessment [32], match outcome predictions [33], player's injury estimation [34], tracking of player's fitness [35], smart sports analysis [36], online content generation tactics [37], selecting a perfect team [38], as well as fantasy sports [39].

Researchers have mostly employed ML algorithms, particularly in cricket, to forecast match outcomes [40-42] as well as assess players [43]. A few selected works that are associated with cricket player selection using ML methodologies are discussed below.

The research's [44] objectives were to establish profiles of psychological strength in teenage cricket players and compare how these profiles differed in terms of cognitive assets as well as adverse mental states. Once weighed against cricketers with a medium degree of mental strength, the findings indicate that individuals with a high degree of mental strength claimed to have more cognitive assets and less adverse mental states. In [45], Shah describes new approaches to assess players' performance. The new standard for batsmen takes into account the characteristics of each bowler they are facing, and the new standard for bowlers takes into account the characteristics of each batsman they are bowling to. The overall score of a batsman's performance is the sum of each performance against each bowler. The bowler's overall score of performance is determined by the sum of his or her performances against all of the batsmen.

Clustering techniques have been employed to represent data appropriately to forecast the most effective batsman or bowler slated to be appointed next for a particular match at an Indian Premier League (IPL) Cricket Match in the study [46]. Here authors have demonstrated how neural networks and K-Means or Hierarchical clustering might be utilized in conjunction. The researchers in [47] defined a methodology for assessing a player's worth for the IPL auction in their work. Their prescribed model took into account variables including the player's past bid price, expertise, strike rate, etc. The most effective bowler for the IPL was identified by the study [48] using bowling rate. Researchers took into account a bowler's ability to take wickets, fielding prowess, and total career catches. Another study [49], employing artificial neural networks, attempted to forecast bowlers' IPL performances. They made use of metrics for bowlers' effectiveness from past ODI and T20I games. After evaluation, the authors rank the players based on their performance. Ideal players might then be chosen for matches and competitions based on these rankings.

Social network analysis was used by researchers [50] to evaluate the performance of a team's bowlers and batters. Employing player-vs.-player data accessible for test and ODI matches, the authors created a directed as well as weighted group of batsmen and bowlers. An entirely novel set for comparison and then choosing batsmen for limited over cricket team performed by the work [51]. The authors created a new calculation for the strike rate which was largely concentrated on the probability of being out. The strike rate was plotted on a 2D graph with the output on the y-axis. Then authors choose a selection standard for selecting the player. In a different piece of work [52], authors modeled and predicted ODI matches using data mining approaches. For modeling the present situation of the match, authors utilized both past match data (like mean runs scored, the average number of wickets lost, etc.) and immediate match data (like the team that bats are playing at home or away or at a neutral venue, performance characteristics of the two batsmen currently in play). The match's result is then predicted utilizing ML techniques like nearest-neighbours clustering and LR.

The success of Indian bowlers against the international cricket side that regularly plays India was predicted by the researchers [53]. To forecast the number of runs a bowler is going to give as well as how many wickets a bowler is expected to take in a specific ODI match, researchers used backpropagation networks and radial basis function networks. To forecast bowlers' effectiveness, Lemmer [54] created a brand-new metric termed Combined Bowling Rate. Three conventional bowling metrics (which are: bowling average, strike rate, and economy) are integrated to create the Combined Bowling Rate. In the case of batsmen, the work [55], uses a hierarchical linear framework to predict batsmen's success in a test series. By considering the ground, playing pitch, opposition, location, and balls faced, the authors [56] use ML algorithms to assess a batsman's performance. This suggested model thoroughly examines a range of factors and their effects on the runs that batsmen accomplish. In their study, authors [51] employed neural networks to forecast player performance. Researchers divided players into three groups- performer, middling, and failure - as batsmen and bowlers, respectively. Researchers then suggest whether a player deserves to be included in the selection list to compete in the 2007 World Cup depending on how frequently that player has gotten various ratings. In their study [57], the authors used ML classifications to divide all-rounders into four groups: Performer, Batting All-Rounder, Bowling All-Rounder, and Underperformer. The proposed classification model was created using the data of 35 all-rounders who participated in the initial three seasons of the IPL, and it was then utilized for predicting the expected categories of six potential all-rounders. Naive Bayes (NB), K-Nearest Neighbours, and RF were used to make the prediction. The k-fold cross-validation approach is used to test every single one of the analytical findings. The results of the experiment show that RF has substantially greater accuracy in forecasting.

The authors [58] applied evolutionary multi-objective optimization to choose a suitable cricket team in a situation that involves a squad selection. They looked at their batting average and bowling average to gauge each player's effectiveness. To do multi-objective genetic optimization over the team, authors first characterized team selection to be a bi-objective optimization issue and then applied non-dominated sorting genetic algorithms. In a different study, authors [59] chose a team using genetic algorithms. They determined a team's fitness by taking into account the personal fitness of every player. A player's fitness level is determined by how well he performs while fielding, wicketkeeping, batting, and bowling, as well as by how physically fit he is and how much experience with games he has. The authors also took into account the squad's performance recently, against a specific team, and on a specific pitch. The squad was subsequently expressed as a string with every string bit denoting a player using the genetic process.

Using ML approaches, the authors of [60] predict emerging stars in the cricket region. More accurately, they forecast future stars in both the batting and bowling departments. To this end, different aspects as well as mathematical expressions are given, and the notions of co-players, teams, and opposing teams are integrated. Lastly, comparisons are made between the ICC rankings and profiles of the top ten emerging cricket players according to emerging star scores, weighted average, and game performance evolution. Authors of [61] presented a method in which the sporting ability of cricket players is evaluated to establish team composition and training schedules. The LR, K-means, and RF models are used to analyze cricket players' performances. The top players on the chart are chosen for teams by this study, which raises the probability of winning games. By grouping players into five categories, the study [62] seeks to precisely determine cricket teams in the ODI format. The players are ranked as excellent, very good, good, satisfactory, or poor depending on their past and present performances. This paper presents an improved cricket framework where a team of eleven players is selected using a fair method.

From the above studies, it can be seen that most of them are particularly focused on predicting the performance of either a bowler or a batsman. Most of these previous studies focused on predicting the performance of individual bowlers or batsmen, with very few attempting to predict squads for national teams or franchise leagues. Furthermore, the number of parameters considered in these studies was limited, which motivated us to address this research gap by developing an ML-based approach that considers a broader range of performance metrics.

3. Methodology

Here, a detailed description of the research procedures used in this study is provided. It covers the workflow diagram, data collection, and preprocessing, parameters considered for captain, wicketkeeper, batsmen, and bowler selection, performance calculation techniques for batsmen and bowlers, performance prediction models, player selection procedures for the national team and franchise league, and performance prediction parameters. Figure 1 depicts the workflow diagram of this research, which outlines the specifications and steps taken to conduct this investigation as well as mentions the list of approaches for each step.

3.1 Details of dataset, data collection, and processing

The initial step of this research is to collect project-related data from a variety of sources. Required data is gathered from several sources and for some specific tournaments including the Dhaka Premier League (DPL), Bangladesh Premier League (BPL), and ICC tournaments like the World Cup and Champions League. The websites *espncricinfo* and *cricbuzz* were used to manually collect all of this data. After data collection, there are five separate areas where these data are divided. They are the Captain, Wicket-Keeper, All-Rounder, Batsmen, and Bowler.

3.1.1 Dataset parameters

3.1.1.1 Fitness parameters

Cricket players' fitness is a key factor in player selection. Information was gathered about cricketers' fitness during the data collection process to analyze their performance. A variety of parameters are employed to analyze cricket players' fitness. Table 1 shows these parameters.

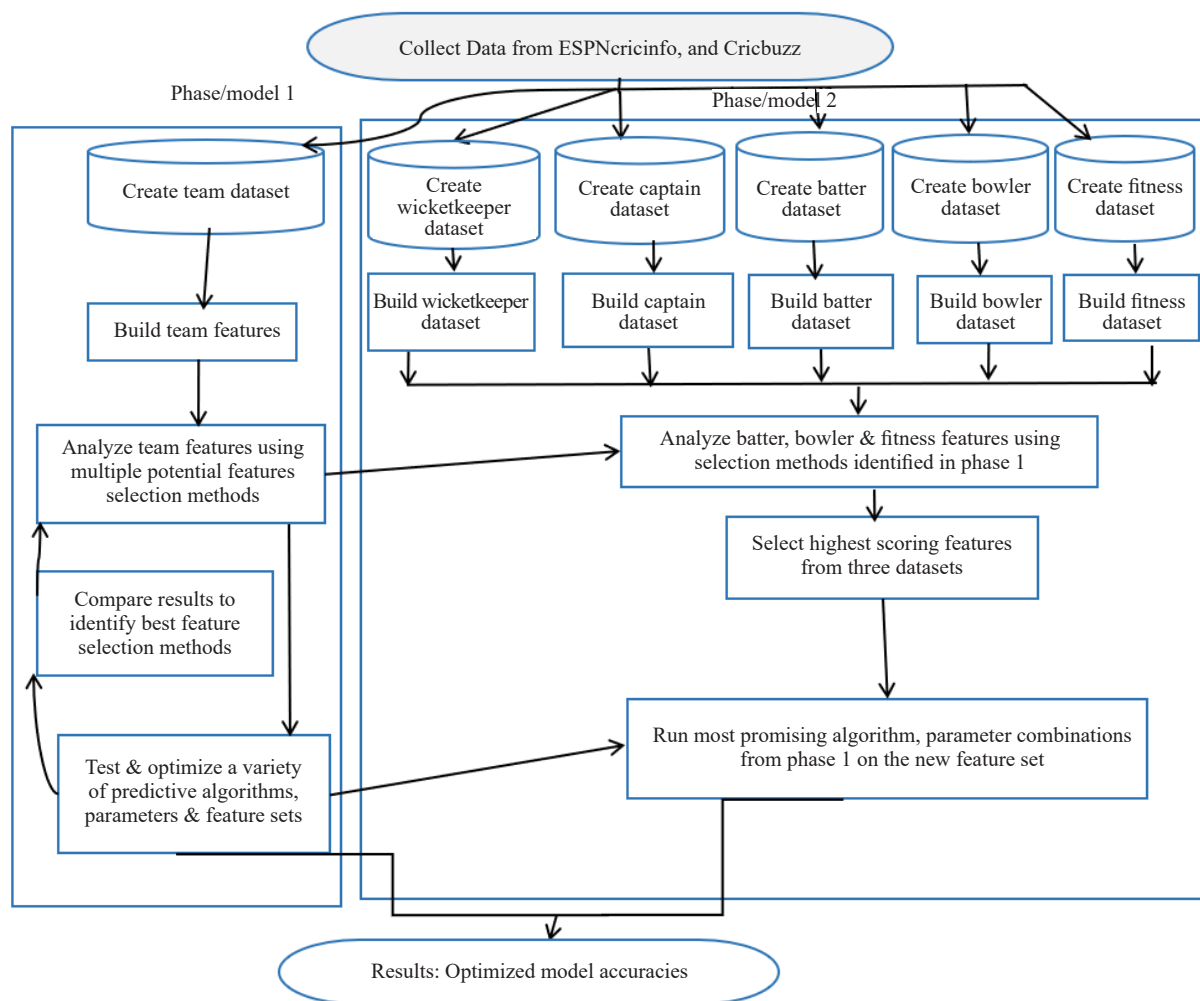


Figure 1. Workflow diagram

BMI: BMI is a measurement that is used to assess both adult men's and women's body fat based on height and weight. A person's weight and height can be evaluated to see if it falls within a healthy range or not using BMI.

The following formula is used to determine BMI:

$$\text{BMI} = \text{weight (kg)} / (\text{height (m)})^2 \quad (1)$$

A player's initial healthiness can be assessed based on the BMI score. These can be:

- BMI less than 18.5: Indicates underweight.
- BMI between 18.5 to 24.9: Indicates normal or healthy weight.
- BMI between 25 and 29.9: Indicates overweight.
- BMI between 30 and 34.9: Indicates Class I obese.
- BMI between 35 to 39.9: Indicates Class II obesity.
- BMI 40 or higher: Indicates Class III obesity or morbid obesity.

Blood Pressure: Blood pressure is the force of blood against the walls of arteries as the heart pumps it around the body. It is measured in millimeters of mercury (mm Hg) and is expressed as two numbers: systolic and diastolic.

Bench Press: For cricket players, upper body power and strength are crucial for a variety of techniques like throwing, striking, and catching. The bench press can be a useful exercise for increasing upper body strength and

power. But to increase total sports performance and lessen the risk of injury, it should be done in conjunction with other exercises that target the back, core, and lower body muscles.

Table 1. Fitness parameters of cricketers

Sl No.	Parameter
1.	Age
2	Height
3	Weight
4	BMI
5	Blood pressure
6	Bench press
7	Squat
8	Speed
9	Body fat percentage
10	Deadlift
11	40-yard dash time
12	Vertical jump
13	Endurance
14	Agility score
15	Flexibility test

Squat: Squats are a compound exercise that works the glutes, hamstrings, and quadriceps among other muscular groups. They can aid in boosting the strength and power of the lower body, which is advantageous for the sprinting, jumping, and explosive movements needed for cricket.

Speed: Speed is a crucial element in cricket since it enables players to move quickly around the field, chase down the ball, and run faster between wickets. Sprints, shuttle runs, and agility drills are a few examples of speed training exercises.

Body Fat Percentage: This is the amount of fat in the body as a percentage of total body weight. Even while a certain amount of body fat is important for good health and performance, too much body fat can hurt a person's speed, agility, and endurance. Because of this, cricket players must maintain a healthy body fat percentage through a combination of good nutrition and activity.

Deadlift: Another complex exercise that may be used to build general strength and power is the deadlift. In particular, they concentrate on the lower back, glutes, and hamstrings, which are crucial muscular areas for running, jumping, and explosive motions in cricket.

Time in the 40-yard Dash: The 40-yard dash is a sprint that gauges an athlete's quick acceleration and speed. Even while it might not be a precise indicator of cricket performance, it can help evaluate general speed and acceleration, which are important in the sport.

Vertical Jump: In cricket, the vertical jump can be useful for jumping, hitting, and fielding. Because it is a measure of explosive power. Exercises like jump squats, box jumps, and plyometrics can be used to train it.

Agility Score: Being agile in cricket refers to having the capacity to change directions swiftly and effectively. Exercises that increase agility include shuttle runs, ladder exercises, and cone drills.

Endurance Score: Cricket matches can run for several hours and need constant effort. Thus, endurance is crucial. Exercises like running, cycling, and rowing can be used to increase cardiovascular endurance.

Flexibility Test: Joint health and injury prevention both depend on flexibility. Stretching activities can help to increase flexibility over time, while flexibility exams like the sit and reach test can be used to evaluate general flexibility.

3.1.1.2 Parameters for captain selection

This captain selection characteristic indicates if a player is capable of being the team leader. For captain selection, we have used some performance parameters which are presented and described in Table 2. Then based on these parameters, the 'leadership Score' is calculated. The equation to calculate the leadership score is:

$$\begin{aligned} \text{leadership_score} = & \text{batting_weight} \times \text{batting_norm} + \text{bowling_weight} \times \text{bowling_norm} \\ & + \text{fielding_weight} \times \text{fielding_norm} + \text{experience_weight} \times \text{experience_norm} \end{aligned} \quad (2)$$

The higher the leadership score is, the more likely to be the player will be selected as a captain.

Table 2. Captain selection parameters

Parameter	Description
Batting average	Batting Average = Runs Scored / Times Out
Bowling average	Bowling Average = Number of Run Conceded / Wicket Taken
Fielding accuracy	The fielding accuracy of a player is a measure of performance of catch, running in fields, stumpings, and hit of stumps
Win percentage	Win percentage = (Number of matches won by the captain / Total number of matches led by the captain) × 100
Loss percentage	Loss percentage = (Number of matches lost by the team/Total number of matches played by the team) × 100
Experience	Total playing years

Table 3. Wicket-keeper selection parameters

Parameter	Description
Total matches	Number of matches the player played during his carrier
Total stumpings	Total stumpings in his carrier
Total catches	Total catches out
Total run-outs	Total run out by wicketkeeper
Dismissal	Total dismissal out by wicketkeeper
Max dismissal per inn	Per match dismissal rate
Dismissal/innings	Dismissal per match
Total runs	Total runs in his career
Run rate	Total runs / total_inngs
Batting strike rate	Runes Score × 100 / Balls Faced

3.1.1.3 Parameters for wicket-keeper selection

This wicketkeeper selection characteristic indicates if a player is capable of being a wicketkeeper. The majority of wicketkeepers are batters. They are anticipated to score more runs because of their hitting specialization and are less worn out than other players as a result of less physical activity when fielding. For wicketkeeper selection, we have used some performance parameters which are presented and described in Table 3.

3.1.1.4 Parameters for batsmen selection

The selection of the perfect batsmen is very important. The performance of the squad can be significantly impacted by choosing the best batsmen for the given match. Before choosing the batsmen, the selectors take into account several variables, including the player's current form, the pitch conditions, the opposition, and the team's batting order. The players' batting statistics, including their average, strike rate, and overall performance under various conditions, are also taken into consideration by the selectors. They also take into account the player's hitting style, such as whether it is aggressive or defensive, and how well it fits with the team's batting plan. For batsmen selection, we have used some performance parameters which are presented and described in Table 4.

Table 4. Batsmen selection parameters

Parameter	Description
Career span	Cricketers career start and end session
Career start	Year of career start
Total matches	Total matches in the squad
Total innings	Total innings played as batsmen
Ducks	Number of matches with no runs in his career
Total runs	Total runs in his career
Highest score	Highest score in his career
Batting average	The batting average is commonly referred to as the average number of runs scored per inning. This attribute indicates the run-scoring capability of the player
Balls faced	The total number of he faced
Strike rate	Strike rate is the average number of runs scored per 100 balls faced
No. of 100 s	Number of hundreds in his career
No. of 50 s	Total 50 s in his career
Total sixes	The number of sixes he hits
Total fours	The number of fours he hits
40+ batting average	The total number matches he doing 40+ runs per innings
90+ batting strike rate	Total number of matches he doing 90+ strike rate per innings

3.1.1.5 Parameters for bowlers selection

In cricket, choosing the correct bowlers is important since they can assist a side in winning a match. The bowling attack of a team must be planned and well-balanced, including velocity, swing, spin, and accuracy. The choice of

bowlers is frequently made based on the pitch's characteristics, the strengths and weaknesses of the opponent side, and the bowlers' physical condition and form. For bowler selection, we have used some performance parameters which are presented and described in Table 5.

Table 5. Bowlers selection parameters

Parameter	Description
Career span	Career start and end year
Career start	Year of start career
Total matches	Total matches he played
Total innings	Total innings he bowled
Total overs	Total overs in his career
Maidens	No runs giving in his over
Runs conceded	Total runs conceded in his total overs
Wickets taken	Total wickets taken in his career
Best bowling in an innings	Best bowling in an innings is a statistic used in cricket that describes a bowler's top individual bowling effort within a single match innings
Bowling average	It's calculated by dividing the total runs conceded by the number of wickets taken. A bowler who has a lower average is more effective
Economy rate	Economy Rate = Total runs conceded / Total overs bowled
Bowling strike rate	Bowling strike rate = Total number of balls bowled / Total number of wickets taken
Four wickets in an innings	Four wickets were taken in one innings
Five wickets in an innings	Five wickets taken in one innings
200+ wickets taken	Total 200+ wickets taken in his career
< 35.00 bowling avg	Total innings bowled with less than 35 bowling average
< 4.00 economy rate	Total innings bowled with less than 4 economy rates
< 40.00 bowling strike rate	Total innings bowled with less than 40 bowling strike rates
Catches	Number of catches taken
Stumpings	Number of stumping outs performed

3.1.2 Data preprocessing

Preparing and cleaning raw data before actual implementation is crucial as the accuracy of the results will be directly impacted by the quality of the data. We might get a lot of data from many sources. Therefore, part of it might have junk values or be null. The dataset may occasionally contain values that are ambiguous, duplicated, or missing. Moreover, some data may have numerous dimensions, which can be troublesome. Because they greatly extend training time. Hence, data preprocessing is essential to clean up those ambiguous raw data. The handling of duplicates, missing values, dimension reduction, etc. is included in data preprocessing methods. Figure 2 indicates the steps involved in data preprocessing.

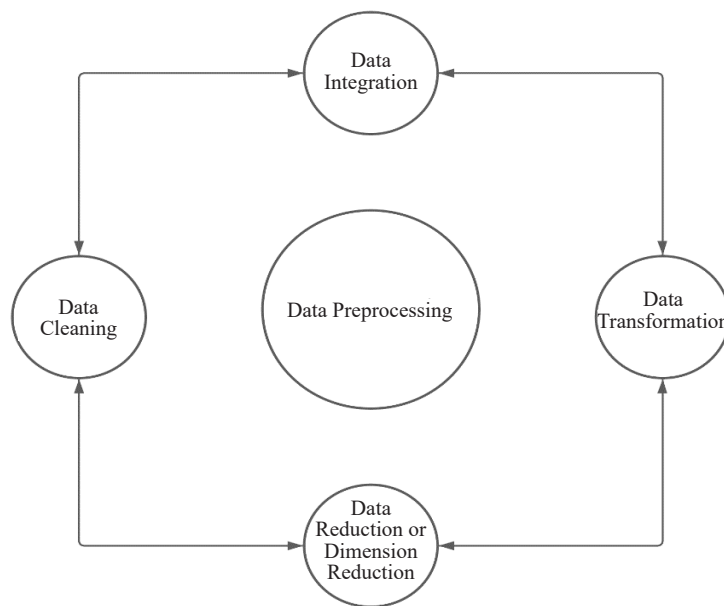


Figure 2. Tasks involved in data preprocessing

3.1.2.1 Data cleaning

This research has faced difficulty when compiling the collected data for the wicketkeeper, captain, all-rounders, and batsmen. Because there were several null, duplicate, and missing values. We also get a lot of unnecessary data during data collection. Thus, data cleaning was necessary. Both the Pandas and Scikit-Learn libraries make it simple to implement the preprocessing of the dataset. Methods like *rename()*, *replace()*, *dropna()*, *loc()*, *iloc()*, and *drop()* are available in the Pandas library. In the case of our dataset, duplicate or missing values were dealt with using the *drop()* method.

3.1.2.2 Data transformation

Data transformation for cricket player selection based on batting, bowling, and fitness datasets involves manipulating and reshaping the data to make it suitable for analysis and decision-making. Initially, relevant data is gathered for the batting, bowling, and fitness performance of cricket players. This data may include statistics such as batting average, strike rate, bowling average, economy rate, and various fitness metrics like speed, agility, and endurance. Generally, data transformation involves several steps. From the collected dataset, first, any irrelevant or duplicate data points are removed and missing values are handled in the “Data Cleaning” step. Then consistency in data formats is ensured, such as converting numerical values to a consistent unit of measurement. After that, new features are created that can enhance the analysis and decision-making process. For this, normalization is used. We can normalize the data to bring all the variables to a similar scale. This step is important when dealing with variables that have different units or ranges, as it ensures that no single variable dominates the analysis due to its magnitude. Normalizing the dataset’s input variables is referred to as feature scaling. Scaling the features of the data is one of the most crucial changes that must be applied. ML algorithms typically struggle to perform effectively when the scales of the numerical attributes are extremely varied. Hence, scaling the data becomes essential. It is possible to address this issue by standardization. With the help of this technique, some data with a scale of 1 to 1,000 can be converted into a value between 0 and 1. Thereby, the ML algorithms’ capacity to anticipate outcomes is enhanced. It’s very easy to standardize something. To ensure that standardized values always have a zero mean, it first subtracts the attribute’s mean value from each of the attribute’s values. Next, it is divided by the variance to create a distribution with a unit variance. This is the best method since standardization is considerably more resistant to the effects of outliers. The equation for the normalization is presented below:

$$X_{\text{normalized}} = (x - \min(x)) / (\max(x) - \min(x)) \quad (3)$$

Python's Scikit-Learn library has classes named *StandardScaler*, *MinMaxScaler*, and *RobustScaler* which perform standardization. Scaling variables like runs, wickets, batting average, and bowling economy rate are used to pick cricket players. By normalizing these variables, the model's accuracy has increased and certain variables can be kept from taking over the model.

Finally, data is aggregated at a suitable level, such as player level or team level, depending on the scope of the analysis. This step involves calculating summary statistics for each player, such as average batting and bowling performances. We might assign different weights to different performance indicators based on their relative importance.

3.1.2.3 Data integration

In the context of cricket player selection, data integration refers to the process of combining data from different sources to create a unified dataset that can be used for squad selection using ML. In this research, we have separate datasets for batting, bowling, and fitness information. The goal is to integrate these datasets into a cohesive dataset that captures all relevant information for player selection. Here's a step-by-step approach for data integration in cricket player selection which we applied in this research.

Data Collection: Gather the individual datasets containing batting, bowling, and fitness information of the players. These datasets may come from various sources such as match records, player profiles, or fitness assessments.

Data Cleaning: Perform data cleaning steps on each dataset separately to handle missing values, remove duplicates, correct inconsistent entries, and address outliers. This ensures that the individual datasets are clean and reliable.

Data Alignment: Before integration, it is crucial to align the data in each dataset so that the information for each player is consistent across all datasets. Ensure that the player identifiers or unique player IDs are consistent among the datasets, allowing proper merging and alignment.

Merging Datasets: Merge the batting, bowling, and fitness datasets using a common identifier, such as the player ID. This merges the datasets into a single dataset that contains information from all three domains. The merged dataset should have a row for each player and columns for batting statistics, bowling statistics, and fitness attributes.

Handling Missing Values: Address any missing values that may arise during the merging process. Depending on the extent of missing data, we can choose to impute missing values or remove instances with substantial missing information. Imputation techniques may involve mean imputation, regression imputation, or sophisticated methods such as multiple imputation.

3.1.2.4 Data reduction

Data reduction for cricket player selection based on batting, bowling, and fitness datasets involves reducing the dimensionality of the data by selecting the most relevant features or variables that contribute significantly to the selection process. Here's a description of the steps involved in data reduction:

Feature Selection: It is uncommon to use all of the variables in a dataset while creating an ML model in the actual world. The model's capacity for generalization is impacted by the addition of pointless or redundant information, which may also lower a model's overall accuracy. The least number of assumptions is necessary for the best solution to a problem. As a result, feature selection becomes a crucial step in creating ML models. To find useful features for enhancing the effectiveness of a model, labeled data can be subjected to ML feature selection approaches. For instance, in our fitness dataset, we have selected three features for batsmen, bowlers, and all-rounders. These are the batsman fitness score, bowler fitness score, and all-rounder fitness score. All other features are not considered here. Then based on these three selected features, we calculate the overall fitness score.

Correlation Analysis: A correlation analysis is conducted to identify the relationships between different variables. Variables that are highly correlated with each other may provide similar information, and it may be possible to eliminate one of them without losing much information. This helps to reduce redundancy and simplify the dataset.

Dimensionality Reduction Techniques: Apply dimensionality reduction techniques to reduce the number of variables while retaining the most important information. Two common techniques are Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). PCA identifies linear combinations of variables that

capture the maximum amount of variation in the data, while t-SNE is useful for visualizing high-dimensional data by mapping them into a lower-dimensional space.

Feature Ranking: Rank the features based on their importance or relevance to the player selection process. This can be done using techniques like information gain, chi-square test, or feature importance from ML models. Features with higher rankings are considered more influential and should be retained, while features with lower rankings can be eliminated.

Subset Selection: Select a subset of features based on their rankings or importance scores. This involves choosing a threshold or a fixed number of features to retain in the dataset. The subset selection process can be performed manually or using automated feature selection algorithms.

3.2 Player performance calculation

Calculating the performance of a player requires taking into account several key parameters. For this reason, we are taking some specific selected parameters for calculating a player's batting and bowling performance. But before that, we have calculated the fitness score of a player to ensure that he is capable of playing.

3.2.1 Fitness score calculation

The fitness score of each player is calculated by taking into account the scores for the seven different physical fitness parameters. These parameters are bench press, squat, speed, deadlift, vertical jump, agility score, and endurance test score. First, the sum of these parameters is calculated and then the sum is divided by 7. Finally, the resulting fitness score is added as a new column in the data frame. Then sorts the DataFrame in descending order based on the fitness score column and selects the top 30 players with the highest fitness scores.

3.2.2 Batsmen performance score calculation

We have calculated the batting performance score by considering several batting statistics for a set of cricket players. At first, we calculate the batting average, strike rate runs per inning, and boundary percentage for each player. These are the considered batting statistics from which we compute a batting performance score by taking the average of them.

Among the considered batting statistics, first, we calculate the batting average for each player. Then, the strike rate is calculated, which is the number of runs scored per 100 balls faced. Next, we calculate the runs per inning by dividing the total runs scored by the number of matches played. Lastly, the boundary percentage is calculated, which is the percentage of total runs scored that come from boundaries (4 s and 6 s), using the formula: $((4 \text{ s} \times 4) + (6 \text{ s} \times 6)) / \text{Runs}$. Ultimately, we calculate the batting performance score as the average of the batting average, strike rate, runs per inning, and boundary percentage.

This method provides a way to quantify a player's batting performance by giving more weight to a certain batting statistic based on their relative importance to the team's goals. The weights assigned to each metric are subjective and may vary depending on the context and the specific requirements of the team.

3.2.3 Bowler performance score calculation

The calculation of bowling performance score involves several statistics of the bowler. These are Maidens (Mdns), Wickets (Wkts), Strike Rate (SR), Economy Rate (Econ), 4-wicket hauls (4), 5-wicket hauls (5), Catches (Ct), and Stumpings (St). To calculate the bowling performance score, first, the sum of Maidens and Wickets is calculated and it's multiplied by the Strike Rate. The result of that calculation is then divided by the economic rate. The result of that division is then multiplied by the sum of 4-wicket hauls and 5-wicket hauls. Finally, the result of that multiplication is divided by the sum of Catches and Stumpings.

Based on the above calculation description, the equation for the bowling performance score can be presented as follows:

$$\text{Bowling Performance Score} = (\text{Mdns} + \text{Wkts}) \times \text{SR} / \text{Econ} \times (4 + 5) / (\text{Ct} + \text{St}) \quad (4)$$

3.3 Splitting the dataset for training and testing

It is not feasible to use every dataset at once. It is common to divide the dataset into two sets. A dataset is often divided into a training set and a testing set, with the majority of the data used for training and the remainder saved for subsequent testing. The consequences of data discrepancies can be reduced by using similar data for training and testing, and the properties of the model can be better understood. In this research's dataset, 60% of the data are used for training and the remaining 40% is kept in stock for testing. To divide the dataset, we used the Scikit-Learn's *train_test_split* class. This research's proposed model is tested by making predictions against the test set after being processed using the training set.

3.4 Players selection

3.4.1 Players selection for national team

National team player selection in cricket is typically based on various performance metrics, including batting, bowling, and fitness performance score. The process of selecting players involves a careful evaluation of these metrics, as well as other factors such as opponent, playing condition, form, and team dynamics.

The first step in selecting a national team player based on batting performance is to analyze the player's batting statistics. For this, several parameters are considered and these are discussed previously. These metrics are used to assess a player's consistency, ability to score runs quickly, and ability to perform under pressure. Similarly, when we are evaluating a player's bowling performance, we consider various bowling-related statistics which are also previously discussed. These metrics are used to assess a player's consistency, ability to take wickets, and ability to restrict the opposition's scoring.

In addition to assessing a player's batting and bowling performance, we also consider their fitness. Fitness has become an increasingly important factor in player selection, particularly in limited-overs cricket. We assess the player's physical fitness, agility, and endurance to ensure they can perform at their best for the duration of the match or series. Fitness assessments will include tests such as the Yo-Yo test, which measures a player's aerobic fitness.

To select players for the national team, we first calculate the player's fitness score. Then based on this fitness score, a group of players are selected having the highest fitness score. Then we carry out the actual squad selection of players by calculating the player's overall performance, which is a combination of batting and bowling statistics. One common method is to use a weighted average, where each statistic is assigned a weight based on its importance. Here, batting average might be assigned a weight of 50%, bowling average might be assigned a weight of 30%, and economy rate might be assigned a weight of 20%. These considered statistics and associated weights can be adjusted based on the specific needs of the team. Using this method we select a player's overall performance score, and then a list of players is selected for the national team.

3.4.2 Player selection for franchise league

To select the players for the franchise league, we consider the franchises of BPL in this research. To select players for the franchises, BPL uses the snake draft system. The snake draft system is a player selection process that is commonly used in fantasy sports leagues and franchise leagues. In this system, each team takes turns selecting players, and the order of selection is reversed after each round. This means that the team that selects last in the first round will select first in the second round, and so on. This ensures that each team has an equal opportunity to select top-performing players.

When using the snake draft system to select players for franchises in the BPL, we create a dataset of potential players that we want to select based on their performance metrics and world rankings. We take turns selecting players from this list, with each team trying to select the best available player based on their needs and the overall strategy of the team. In addition to performance metrics and world rankings, other factors that may influence player selection in the snake draft system include the team's overall strategy, the strengths and weaknesses of other teams in the league, and the player's availability and willingness to play for the team.

3.5 Cricketer's performance evaluation algorithms

To perform the performance prediction, initially, we analyze cricket players' physical condition based on their physical fitness characteristics and calculate a fitness score. Then, using the statistical method, we compute batting and bowling performance scores. Next, combining those results, a statistical formula is used to determine the overall performance. Ultimately, to perform the prediction, the ML algorithms are used. In this research, we have used LR, SVR, and RF. Utilizing these ML algorithms, we have predicted the fitness, batting, and bowling performance of players. In the end, we integrated the three datasets and used those ML algorithms to predict the overall performance of the players. Based on those overall prediction performance values, we choose players for the national team and franchise league. The considered algorithms are Random Forest, Linear Regression, and Support Vector Regression.

3.5.1 Random forest

Random Forest Regression is an ML algorithm used for regression tasks. It is an ensemble method that combines multiple decision trees and produces a single output by averaging the outputs of individual decision trees. Random Forest can handle a large number of input variables and is less prone to overfitting than decision trees.

Each tree in the RF Regression algorithm is built using a different subset of the training data. This process is known as bootstrap aggregating or bagging. During the training process, the algorithm creates a random subset of the training data by selecting observations with replacements. Each tree is then built on this random subset of the data. When a new observation is presented to the model, the algorithm runs the observation through each decision tree in the forest and produces an output. The final output of the model is the average of the outputs from all the decision trees in the forest.

The mean of the predictions made by each tree tends to be the RF's final prediction for regression-related issues. It is the modal (most probable class) in issues with classification.

$$\text{Prediction}(x) = 1 / N \sum_{i=1}^n \text{tree}_i(x) \quad (5)$$

An RF is built in two stages: the first involves combining the N decision trees, and the second involves making predictions for each tree that was built in the first stage. The following steps can be used to demonstrate the working process:

Step 1: Select K data points at random from the training set.

Step 2: Build decision trees for the selected data points (Subsets).

Step 3: Decide on N for the decision trees you want to build.

Step 4: Repeat steps 1 and 2.

Step 5: Locate each decision tree's forecasts for new data points, and then assign the new data points to the category that has received the most votes.

To build this classification algorithm, we have used Python's Scikit-Learn, which provides the *Random Forest Classifier* class.

3.5.2 Linear regression

A statistical technique for examining the relationship between two continuous variables. The best-fit line that can describe the relationship between a dependent variable (also known as the response variable) and one or more independent variables (also known as predictor variables or explanatory variables) is specifically what is sought after.

The correlation coefficient indicates the strength of the association among two variables. This coefficient is ranging from -1 to +1. The degree of the correlation between the data that was collected for the two variables is shown by this coefficient. The formula for a linear regression line equation is:

$$N = i + jM; \quad (6)$$

where,

M is the independent variable and N is the dependent variable plotted along the x-axis and y-axis respectively, and

the slope of the line is j , and i is the intercept.

In basic linear regression, there is only one independent variable, and a straight line is used to model the relationship between the independent and dependent variables. Finding the equation of this line with the shortest possible difference between the expected and actual values of the dependent variable is the objective. A hyperplane is used to model the relationship between the independent and dependent variables in multiple linear regression, which involves two or more independent variables. Finding the hyperplane equation that minimizes the difference between the expected and actual values of the dependent variable is the objective.

3.5.3 Support vector regression

Support Vector Regression is an ML algorithm that is used for regression analysis. It is a type of Support Vector Machine (SVM) that can be used to predict continuous values, such as stock prices, housing prices, or temperature readings. The goal of SVR is to find a hyperplane in a high-dimensional space that has the maximum distance from the actual data points, while still fitting the data within a certain margin of error. The main difference between SVR and traditional regression models is that SVR uses a kernel function to transform the original data into a higher-dimensional space, where the hyperplane can be found. The kernel function helps to identify nonlinear relationships between the independent and dependent variables, making SVR a powerful tool for modeling complex datasets.

In SVR, the objective function seeks to decrease both the degree of model complexity as well as the error among the predicted and actual values. A loss function is utilized to calculate the error. The following is the fundamental equation:

$$1/2 \|x\|^2 + Y \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (7)$$

Here,

The vector for weight is x .

A regularization factor called Y regulates the trade-off between obtaining a minimal model complexity as well as a reduced training error.

Slack variables ξ_i and ξ_i^* permit a few training units to exhibit a positive error or to be outside of the margin.

SVR works by first defining a set of support vectors that represent the training data. These support vectors are used to find the hyperplane that best fits the data. The hyperplane is then used to predict the values of the dependent variable for new data points. One of the key advantages of SVR is that it can handle outliers and nonlinear relationships between variables, which makes it more robust than traditional regression models.

3.6 Performance evaluation metrics

Performance evaluation metrics are used to evaluate the performance of the utilized ML algorithms. These metrics provide insights into how well the algorithm is performing on the given dataset and can help guide improvements and optimizations. For this research, three performance metrics have been considered, which are: accuracy, Mean Squared Error (MSE), and Mean Absolute Error (MAE).

Accuracy: The model's accuracy is calculated as the proportion of accurate guesses to all other predictions. A common presentation method is to multiply the result by 100 to get the percentage. The percentage of findings that are truly positive (both truly positive) is represented by accuracy's numerical value and genuine negativity in the chosen population. The equation is displayed below.

$$\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / (\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}) \quad (8)$$

Mean Squared Error: This metric measures the average of the squared differences between the predicted and actual values. It is useful for regression problems.

$$MSE = 1 / n \times \sum (y_i - \hat{y}_i)^2 \quad (9)$$

Where n is the number of samples, y_i is the actual value, and \hat{y}_i is the predicted value.

We have implemented this metric in this research using Python environments sklearn's *mean_squared_error* method.

Mean Absolute Error: The average size of the errors in all of the predictions is what this metric measures. We are aware that an error is essentially the disparity in absolute terms between the true or real values and the expected values. Because of the absolute difference, results with a negative sign are ignored. Hence,

$$MAE = \text{True values} - \text{Predicted values} \quad (10)$$

The outcome of MAE is the mean of this error over all samples in a dataset. We have implemented this metric in this research using Python environments sklearn's *mean_absolute_error* method.

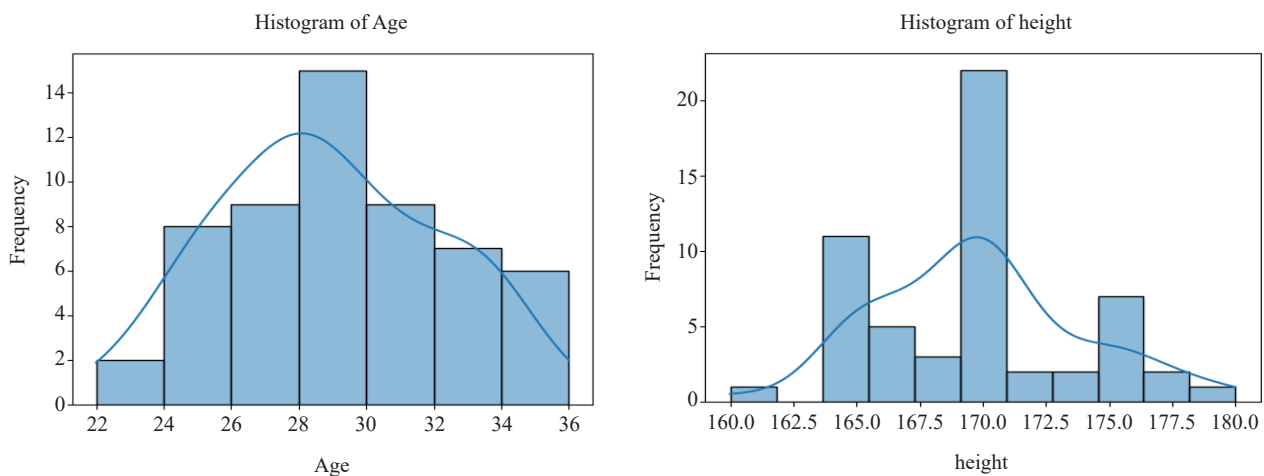
4. Results and discussion

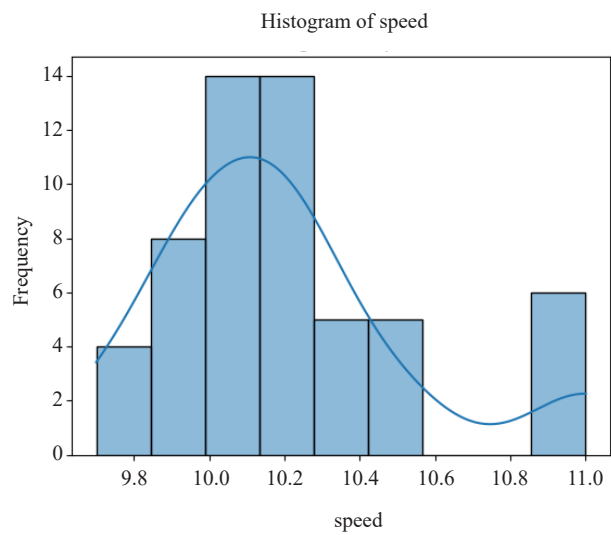
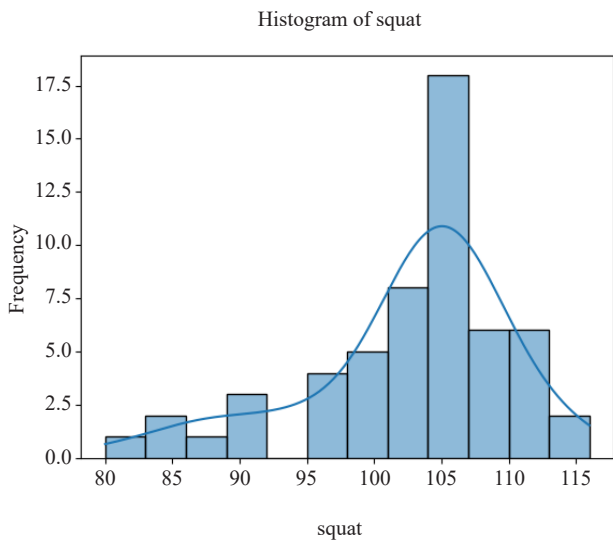
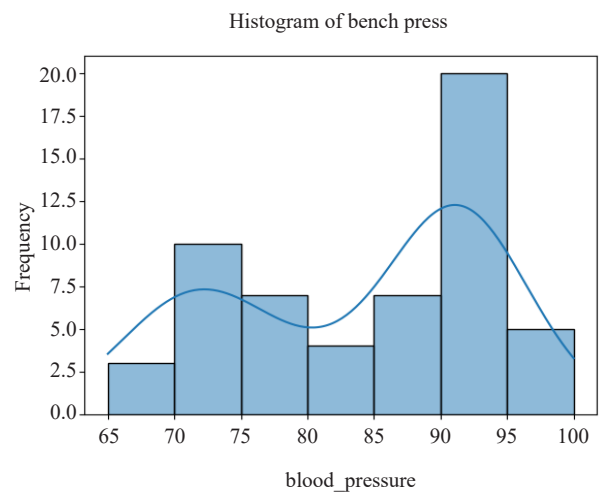
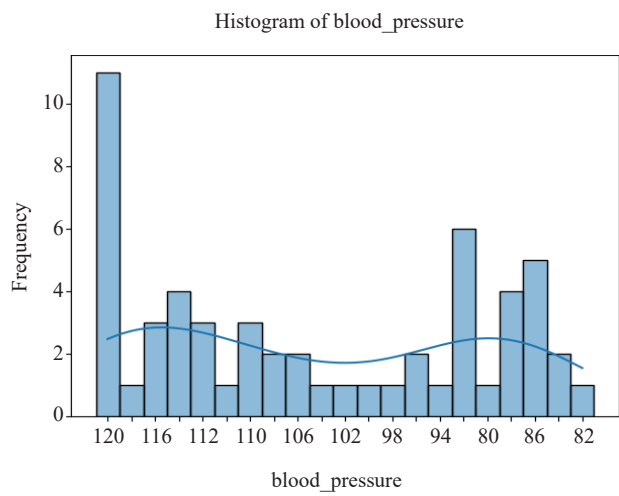
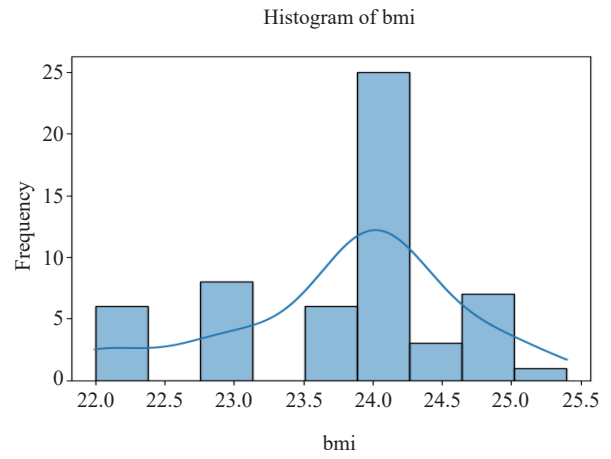
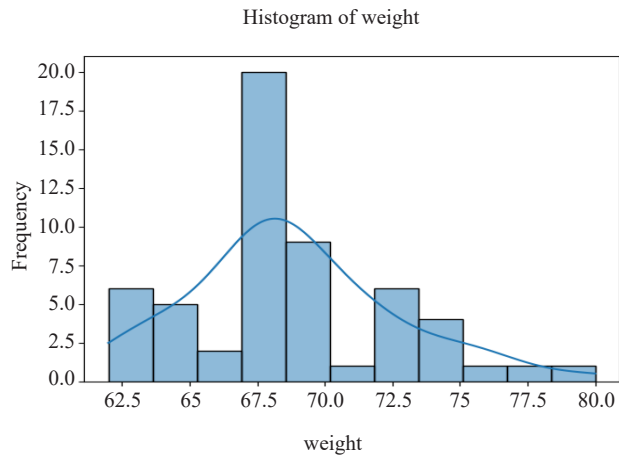
4.1 Cricketers fitness data analysis

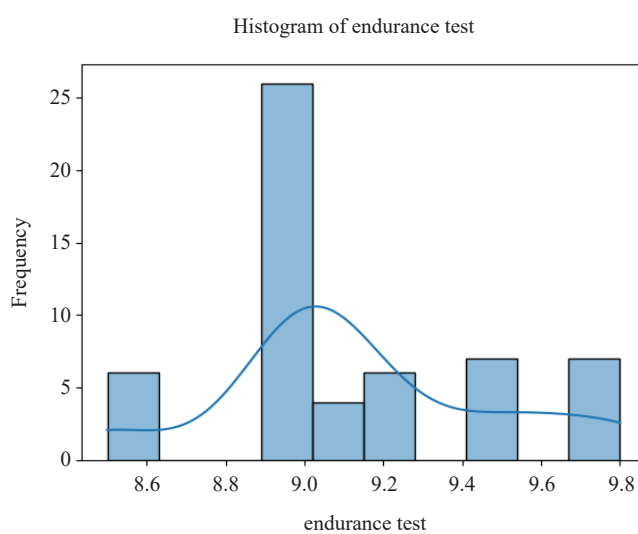
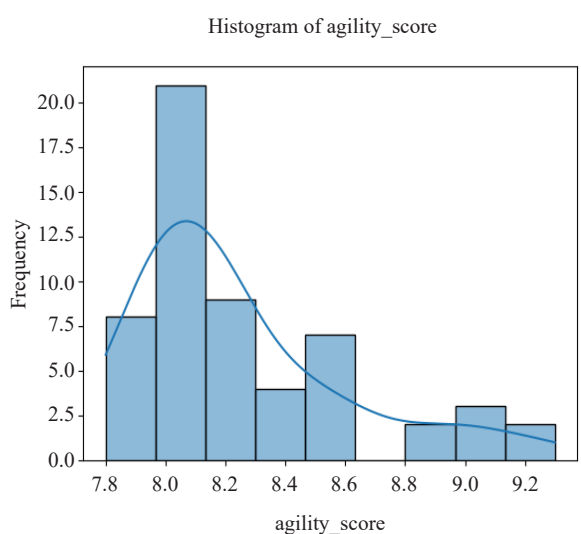
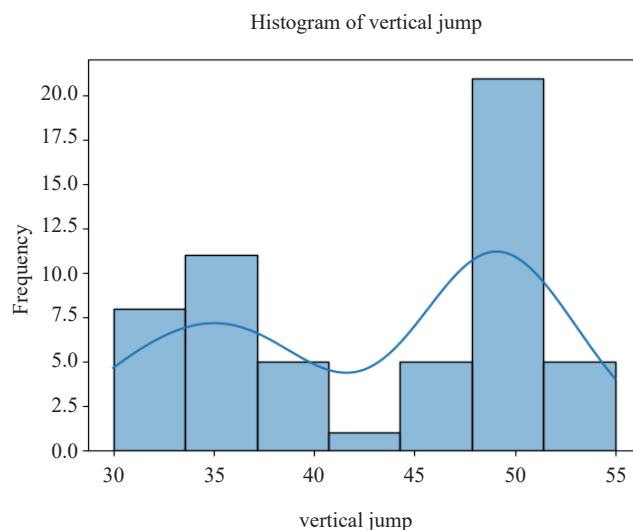
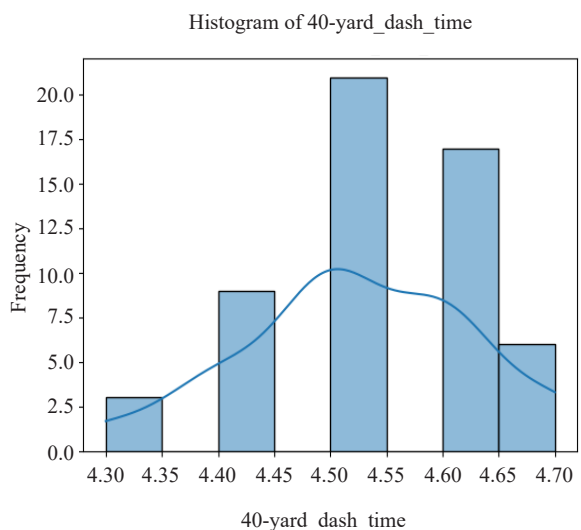
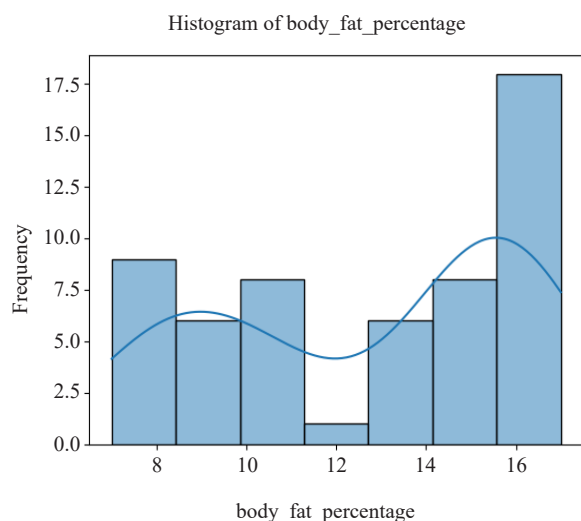
We perform various data exploration and visualization tasks using Python's pandas and matplotlib libraries on the fitness dataset. The results of these analyses are used to make informed decisions about the fitness and health of the players. Figure 3 shows the histogram of all columns which are considered for this research. The basic statistics of these data reveal information about the distribution of values for each variable. The mean and standard deviation of BMI and agility scores can provide insights into the fitness levels of the players. Similarly, the correlation matrix shows the strength and direction of linear relationships between pairs of variables. The histograms, boxplots, and scatter matrix plots provide visual representations of the data distribution and relationships between variables. These plots can help identify outliers, skewness, and non-linear relationships.

The scatter plot of agility score vs. endurance test, colored by BMI category, can help identify potential relationships between these variables and the influence of BMI on agility and endurance. Figure 4 visually represents this.

The creation of a new column 'total_strength' as the sum of bench press, squat, and deadlift, and the subsequent calculation of mean of total strength by age group using the group by method, can help identify age-related changes in strength. The age group is used to bin the ages into intervals of 5 years between 20 and 35 years. This calculates the average total strength for players in each age group. In this research, we divided three age groups which are: 20-25, 25-30, and 30-35. The mean value of 'total_strength' for each age group is presented in below Table 6.







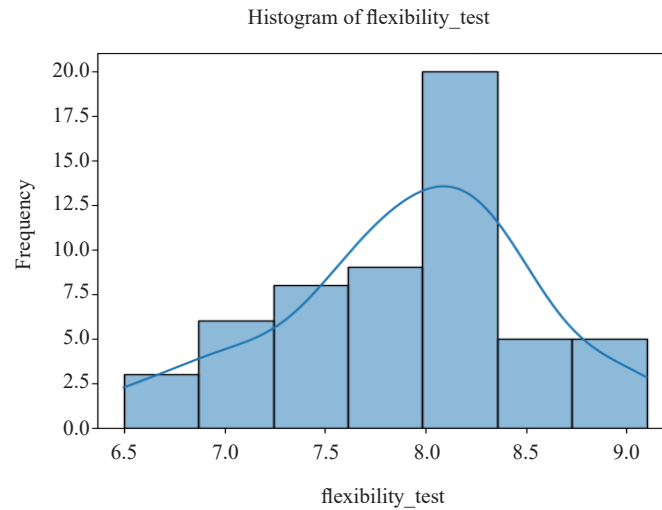


Figure 3. Histogram of the considered fitness parameters

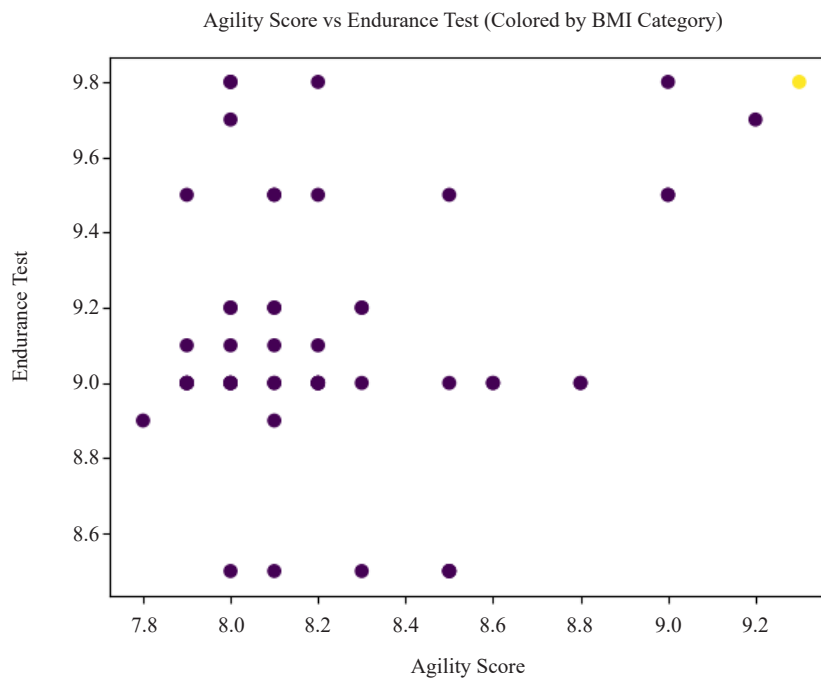


Figure 4. Agility score vs endurance test graph

Table 6. The mean value of 'total_strength' for each age group

Age	Mean value of 'total_strength'
(20, 25]	273.954545
(25, 30]	281.392857
(30, 55]	266.00000

For the 'bench press' column, we create a box plot showing the distribution of bench press values by age group. This is presented in Figure 5. The box plot shows the median, quartiles, and outliers for each age group. From the dataset, we find that the maximum bench press value is 550 lbs and the mean bench press value is 240.7 lbs, which is significantly lower than the maximum value. This suggests that there are likely some outliers in the data pulling the maximum value higher. The box plot of bench press by age group shows that the median bench press value increases as the age group increases, with the 30-35 age group having the highest median bench press value. However, there is also a lot of variability in the data for each age group, with some outliers having much higher bench press values.

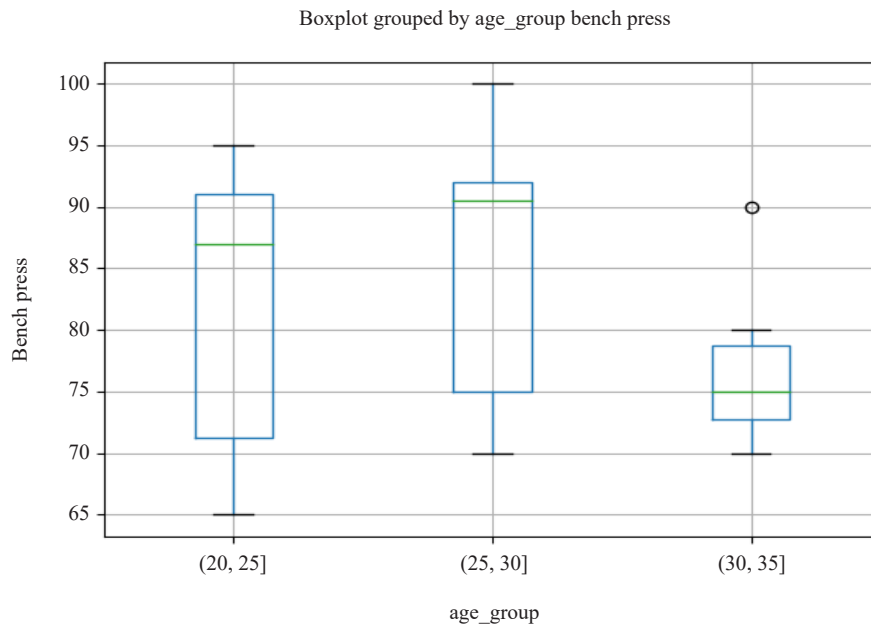


Figure 5. Bench press for different age

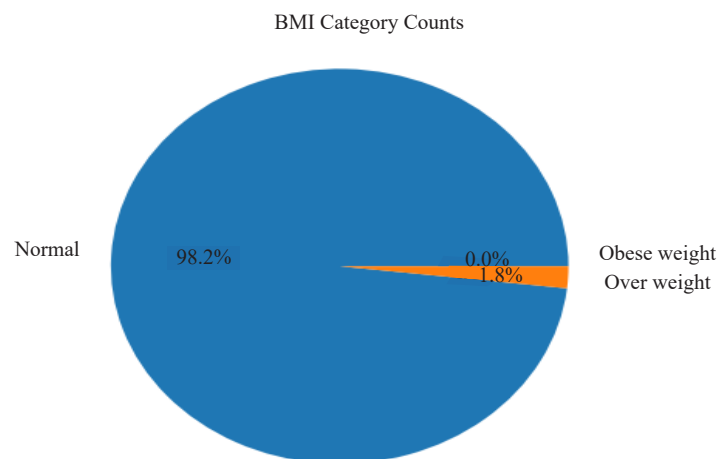


Figure 6. BMI category

The creation of another new column called 'bmi_category' based on BMI value, and the subsequent counting of players in each BMI category using the value_counts method, can provide insights into the distribution of BMI values

and the prevalence of underweight, normal, overweight, and obese players in the dataset. We divided the BMI of crickets in 'bmi_category' based on the 'bmi' values of each player. The labels for the categories are 'Underweight', 'Normal', 'Overweight', and 'Obese', and the cutoff values for the categories are [0, 18.5, 25, 30, 100]. The pie chart of Figure 6 shows the BMI category counts which provides a clear visualization of this distribution. We have considered a total of 56 players out of which 55 have a 'Normal' BMI score, and 1 has an 'Overweight' score. No players have having 'Underweight', or 'Obese' score.

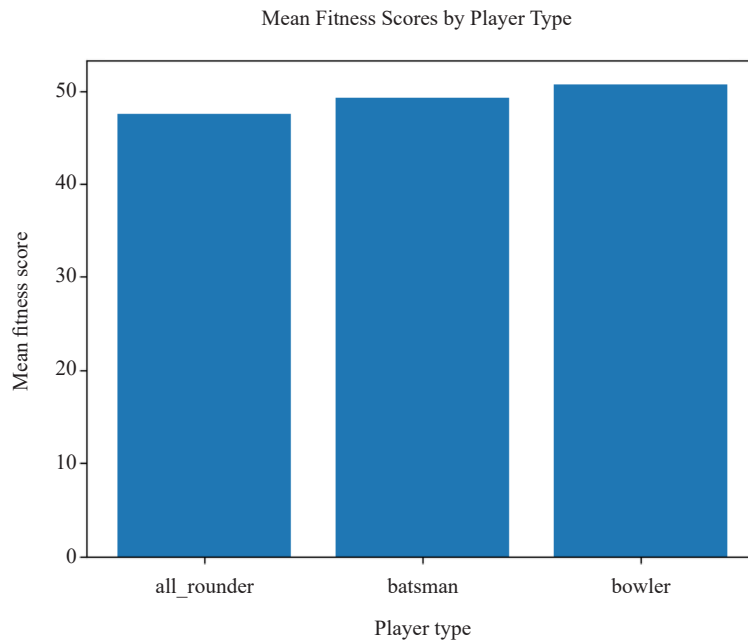


Figure 7. Players mean of fitness score

4.2 Player selection based on fitness data

To select the players, at first, we read the data into the panda's data frame and then filtered the data frame based on the player type (i.e., batsmen, bowlers, or all-rounders). After filtering, we calculate the overall fitness score for each player type by taking the mean of seven different fitness measurements, which are: bench press, squat, speed, Deadlift, vertical jump, agility score, and endurance test. We calculate their fitness score using the mathematical formula which we already discussed in Section 3.4.1. The data frame is then sorted in descending order based on the fitness score, and the top 15 players for batsmen and bowlers and the top 10 players for all-rounders are selected. That means, we find out top fitness players based on their fitness scores and we show them in Table 7. These selected top players are then used in our ML model for further analysis. The players with low fitness scores are optioned out from the developed model for further investigation. Figure 7 shows the mean fitness score of each player type.

4.3 Captain selection

In our dataset, we have 15 players who are competing for the captain position. To select one as a Captain, we have incorporated 'Leadership Score' whose mathematical formula is presented in Equation (2). This "Leadership Score" is a combination of several other parameters that are mandatory for captain selection. The leadership score first defines a dictionary containing data on each player's specific parameters, which are: batting average, bowling average, fielding accuracy, experience, win percentage, and loss percentage. Figure 8 shows the scores for all these considered parameters of the 'Leadership Score'.

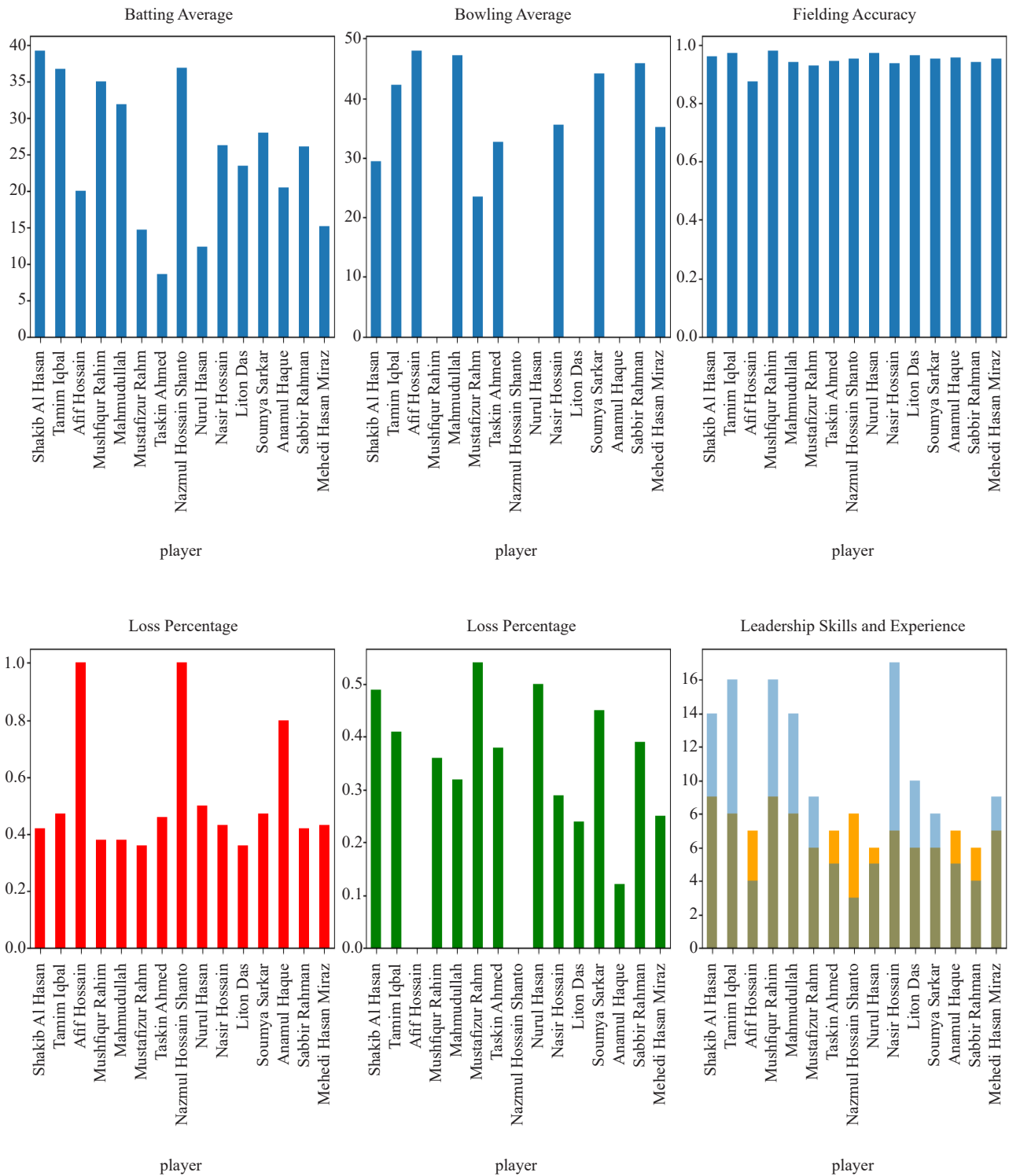


Figure 8. Performance score of all parameters of leadership score

Next, we calculate a leadership score for each player using a weighted formula that incorporates the normalized values for batting, bowling, fielding, and experience. The resulting leadership score is added to the dictionary as a new column. For the ‘Leadership Score’, we didn’t measure only the performance percentage of a player but also measured his performance as a batsman or bowler or both. The reason is that, only because of captaincy, we can’t take a player

in the team. He also should have created an impact on either the batsmen area or bowler area, or both. Figure 9 shows the overall captain selection graph. From the graph, we can see that the top three players for the captaincy position are: Nassir Hossain, Shakib Al Hasan, and Tamim Iqbal. As Nassir Hossain has the highest score among all other players, he is going to be selected as a captain for the squad.

Table 7. Selected players based on fitness data

Top bowler	Top batsmen	Top all-rounder
Fazle	Anamul Haque	Shakib Al Hasan
Arafat Sunny	Mahmudul Hasan Joy	Saifuddin
Haider Ali	Yasir Ali	Sabbir Rahman
Rony Talkdur	Emon	Mahmudullah
Abu Jayed	Shahin Alam	Soumya Sarkar
Mahmud	Rakibul Hasan	Afif Hossain
Avishek Das	Akbar Ali	Nasir Hossain
Al Amin	Tanzid Hasan	Mehidy Hasan Miraz
Shahadat Hossain	Shamsur Rahman	Mosaddek Hossain
Rubel Hossain	Shamim Hossain	Abul Hasan
Kamrul Islam	Mohammad Naim	-
Tanzim Hasan Sakib	Saif Hossain	-
Shahidul Islam	Tamim Iqbal	-
Jubair Hossain	Muktar Ali	-
Shoriful Islam	Liton Kumar Das	-

4.4 Wicketkeeper selection

In our dataset, we have 12 players who are competing for the wicketkeeper position. To select one, we have measured each player's performance by considering some specific parameters, which are: total matches, total stumpings, total catches, total run outs, dismissals, maximum dismissal in one innings, dismissal/innings, total run, run rate, and batting strike rate. Figure 10 shows the overall performance scores of all players by considering the mentioned parameters.

The player who ought to have affected all considered parameters will be selected as a wicketkeeper. Because we can't include a player as a wicketkeeper in the team by simply looking at one statistic. As seen from Figure 10, Mushfiqu is the best candidate to play the wicket-keeping position.

4.5 Performance metrics

To predict the cricket players' performances, we employed LR, SVR, and RF regression in the batting, bowling, and the combination of batting, bowling, and fitness datasets. The metrics which are considered for performance measure are Accuracy, MSE, and MAE. The details about them are presented in the 'Methodology' section. Table 8 presents the prediction performance for the batting dataset, Table 9 presents the prediction performance for the bowling

dataset, and Table 10 presents the prediction performance for the combined dataset.

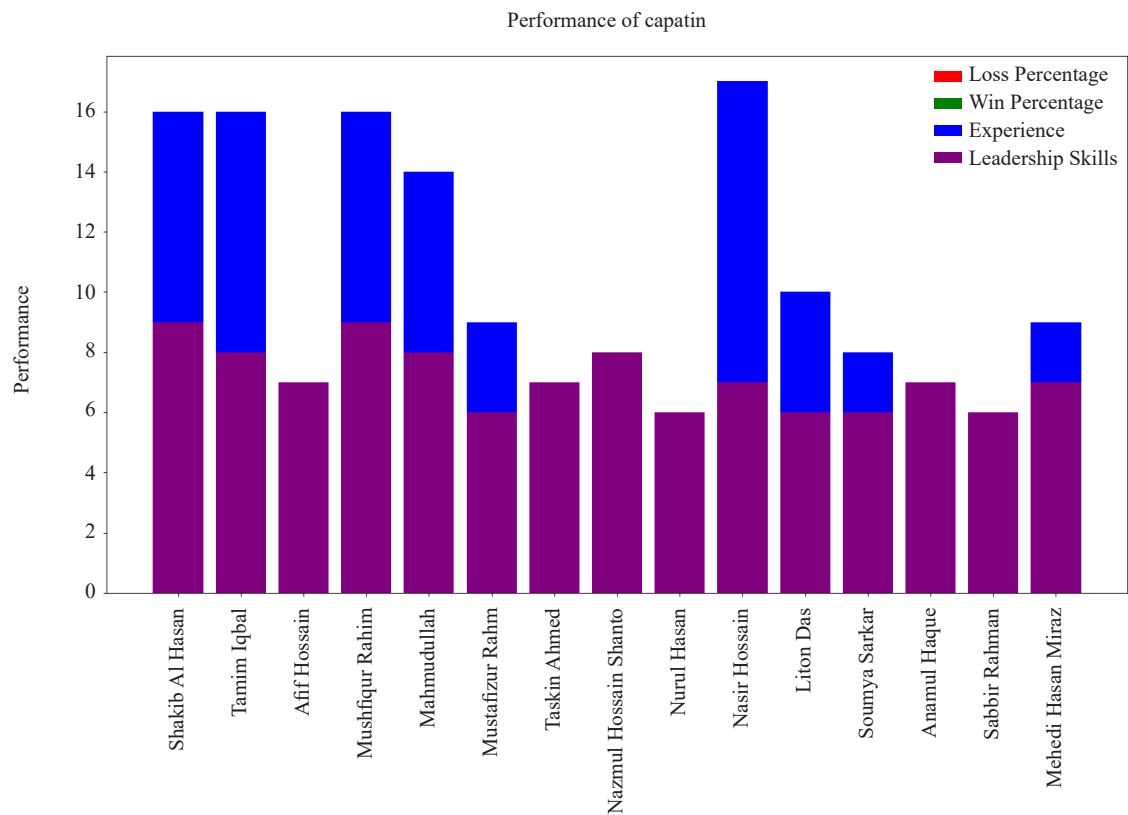


Figure 9. Captain selection graph

Table 8. Prediction performance for batting dataset

Performance prediction metrics	Linear regression	SVR	Random forest
MSE	23.18	9.18	33
MAE	8	8.1	3
Accuracy	84%	99%	86%

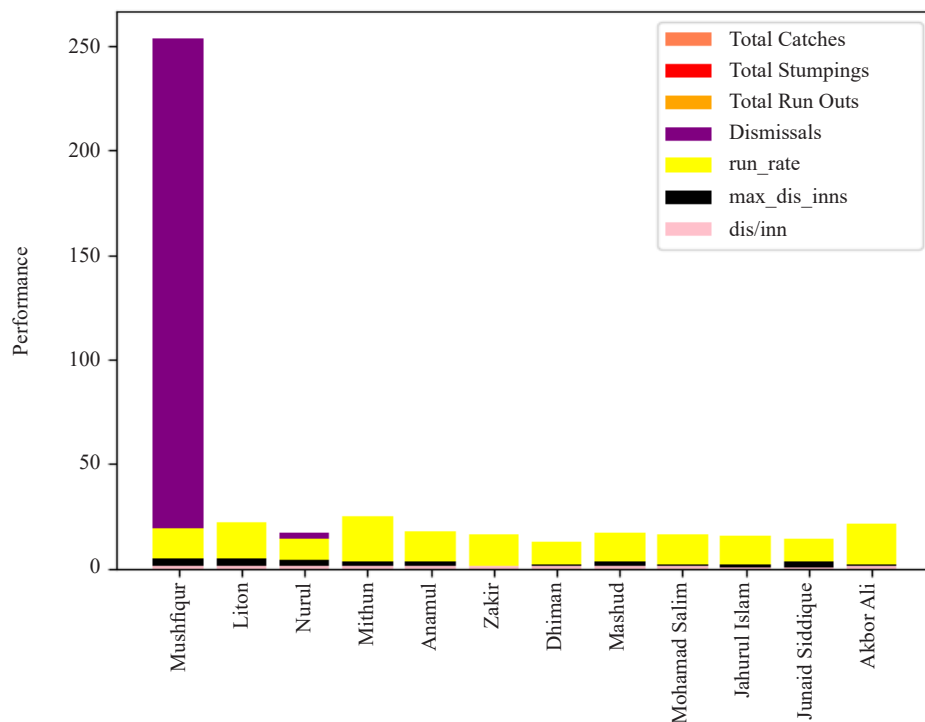
Table 9. Prediction performance for bowling dataset

Performance prediction metrics	Linear regression	SVR	Random forest
MSE	339.33	30.7206	148
MAE	33.25	2.3819	49.50
Accuracy	93%	99%	72%

Table 10. Prediction performance for combined dataset

Performance prediction metrics	Linear regression	SVR	Random forest
MSE	6.778	2.43	8.13
MAE	2.43	5.10	1.92
Accuracy	98%	99%	97%

From the Tables, we can see that the SVR gives us the best prediction accuracy for both batting, bowling, and combined dataset.

Figure 10. Wicket keeper performance graph**Figure 10.** Wicket keeper performance graph

4.5.1 National team player selection

For selecting the National team players, first, utilizing all relevant performance-related characteristics, we compute bowling performance from a bowling dataset and batting performance from a batting dataset. When all is said and done, we merge the three datasets (i.e., batting, bowling, and fitness), and then calculate the overall performance of this combined dataset. Finally, we select the top players based on the results of the combined dataset. We have selected two squads (each containing 16 players) from the chosen top players which are presented in Figure 11.

	squad 1	squad
0	Imrul Kayes	Jahurul Islam
1	Dolar Mahmud	Nasum Ahmed
2	Anamul Haque	Mustafizur Rahman
3	Nurul Hasan	Soumya Sarkar
4	Litton Das	Fazle Mahmud
5	Mominul Haque	Ziaur Rahman
6	Sabbir Rahman	Nasir Hossain
7	Hasan Mahmud	Mosaddek Hossain
8	Mohammad Saifuddin	Mahbubul Alam
9	Afif Hossain	Raqibul Hasan
10	Naeem Islam	Raqibul Hasan
11	Yasir Ali	Rubel Hossain
12	Shamsur Rahman	Mehidy Hasan Miraz
13	Mohammad Mithun	Shuvagata Hom
14	Dhiman Ghosh	Nazimuddin
15	Taskin Ahmed	Arafat Sunny

Figure 11. Players selected for national team

	Dhaka	Chattogram	Cumilla	Rajshahi	Khulna	Rangpur	Sylhet	Barishal
0	TA Boult	Mashrafe Mortaza	Rashid Khan	Imran Tahir	DW Steyn	JJ Bumrah	SL Malinga	CH Gayle
1	B Kumar	LRPL Taylor	RA Jadeja	V Kohli	MS Dhoni	M Morkel	AB de Villiers	HM Amla
2	RG Sharma	F du Plessis	JE Root	Shakib Al Hasan	AJ Finch	KS Williamson	Q de Kock	MJ Guptill
3	Fakhar Zaman	Tamim Iqbal	Mushfiqur Rahim	JM Bairstow	B Kumar	S Dhawan	Babar Azam	EJG Morgan
4	Imam-ul-Haq	TK Curran	SD Hope	AT Rayudu	JC Buttler	Aamer Yamin	SE Marsh	JJ Roy
5	SO Hetmyer	MP Stoinis	SS Iyer	Haris Sohail	KM Jadhav	DA Miller	AD Hales	PR Stirling
6	BA Stokes	TM Head	Imad Wasim	MK Pandey	HM Nicholls	EC Joyce	E Lewis	EC Joyce
7	C Jonker	TD Astle	KA Maharaj	Shafiqullah	RR Pant	TL Seifert	N Dickwella	Hazratullah Zazai
8	DSK Madushanka	CJ Dala	Rumman Raees	OP Stone	Asif Ali	B Mavuta	D Pretorius	SN Thakur
9	MD Shanaka	MR Marsh	Imrul Kayes	SW Billings	Soumya Sarkar	C Munro	MM Ali	C Zhuwao
10	MDKJ Perera	MJ Henry	Rahmat Shah	Hasan Ali	AR Nurse	MA Wood	C de Grandhomme	AU Rashid
11	F Behardien	AK Markram	MJ Santner	Sabbir Rahman	Rashid Khan	PWH de Silva	TG Southee	JDS Neesham
12	GJ Maxwell	AJ Tye	Mujeeb Ur Rahman	CH Morris	LE Plunkett	AD Russell	Mohammad Nawaz (3)	Yasir Shah

Figure 12. Players selected for franchise league

4.5.2 Franchise league player selection

In the player selection process of the Franchise League, our initial step involves gathering batting and bowling performance data in the T20 format from players representing various countries. To evaluate overall performance, we compute batting performance from the batting datasets and bowling performance from the bowling datasets. Additionally, we consider the competitiveness of the selected athletes. Subsequently, we merge these three datasets and calculate the overall performance of the combined dataset. Ultimately, the top players are chosen based on the results obtained from this combined dataset. The selection process employs a snake draft technique, and Figure 12 showcases the chosen top players, with one squad selected for each Franchise.

5. Conclusion, limitations, and future scopes

The objective of this research is to assess the playing techniques and fitness of individual cricket players based on multiple parameters, and subsequently select squads for the national team and franchise league. To accomplish this goal, data is gathered from two prominent cricket websites, namely *espncricinfo*, and *cricbuzz*, and conducted preprocessing on the data. The parameters utilized for selecting captains, wicketkeepers, batsmen, and bowlers are discussed, and player performance scores are calculated based on unique performance calculation methods. We then present our performance prediction ML models, implemented using LR, SVR, and RF algorithms. Ultimately, based on these performance prediction scores, two squads for the national team and one squad for each franchise league team are selected.

However, this research on the application of ML for cricket player selection is not without limitations. Some of them are:

- The study makes use of information gathered from *cricbuzz* and *espncricinfo*, which might not be an exhaustive dataset. This could limit our ML model's capacity to make generalizations throughout broad cricket environments as it could overlook players from less-covered tournaments or regions.
- The study discusses employing physical fitness data, batting, and bowling statistics, along with other relevant statistics, which are static and don't account for a player's growing skills or form. As a result, the model can have trouble capturing the dynamic aspect of player performance.
- While the study concentrates on quantitative measures like fitness data and statistics, it can miss non-quantifiable elements like versatility, leadership, and teamwork which are critical in team sports like cricket.
- A player's form, fitness, or playing style may vary over time, and this could be missed if previous game data is the only thing used. It's possible that the models won't be able to accurately predict a player's performance in the future, particularly if their circumstances significantly alter.

These limitations provide several opportunities for additional future investigation and advancement of this research's findings. Among them could be:

- Future studies could investigate the inclusion of real-time data, including current injuries, recent playing circumstances, and present form, within the player selection procedure to improve the model's responsiveness to the dynamic characteristics of cricket.
- Incorporating non-numerical information, such as professional judgments, coaching comments, and retired player interviews, can yield a more comprehensive picture of a player's ability. This would take care of the drawbacks brought on by some performance measurements' subjective character.
- Examine possible biases in the ML models as well as data to make sure that players are chosen fairly. Talk about the moral implications of data privacy and openness in the selection procedure.
- Provide a user-friendly interface that makes it simple for selectors to communicate with the ML models. Selectors can use this to help in a cooperative decision-making procedure by giving the models information as well as input.

The suggested ML-based method for selecting cricket players can be improved, enlarged, and made more flexible to the changing dynamics of the game by pursuing these potential research avenues.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Sanderson J, Brown K. COVID-19 and youth sports: Psychological, developmental, and economic impacts. *International Journal of Sport Communication*. 2020; 13(3): 313-323. Available from: doi: 10.1123/IJSC.2020-0236.
- [2] Deep Prakash C, Verma S. A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket. *Decision Analytics Journal*. 2022; 2: 100025. Available from: doi: 10.1016/J.DAJOUR.2022.100025.
- [3] Kampakis S, Thomas W. *Using Machine Learning to Predict the Outcome of English County Twenty over Cricket Matches*. 2015. Available from: doi: 10.48550/arXiv.1511.05837.
- [4] Bandulasiri A, Brown T, Wickramasinghe I. Factors affecting the result of matches in the one day format of cricket. *Operations Research and Decisions*. 2016; 26(4): 21-32.
- [5] Morgulev E, Azar OH, Lidor R. Sports analytics and the big-data era. *International Journal of Data Science and Analytics*. 2018; 5(4): 213-222. Available from: doi: 10.1007/S41060-017-0093-7.
- [6] Sarlis V, Tjortjis C. Sports analytics — Evaluation of basketball players and team performance. *Information Systems*. 2020; 93: 101562. Available from: doi: 10.1016/J.IS.2020.101562.
- [7] Xu T, Tang L. Adoption of machine learning algorithm-based intelligent basketball training robot in athlete injury prevention. *Front Neurorobot*. 2021; 14: 117. Available from: doi: 10.3389/FNBOT.2020.620378/BIBTEX.
- [8] Horvat T, Job J. Importance of the training dataset length in basketball game outcome prediction by using naive classification machine learning methods. *Elektrotehniški vestnik*. 2019; 86(4): 197-202. Available from: <https://ev.fe.uni-lj.si/4-2019/Horvat.pdf> [Accessed 3rd May 2023].
- [9] Yang Y, Qin T, Lei YH. Real-time eSports match result prediction. In: *Conference on Neural Information Processing Systems*. Spain, Barcelona; 2016. Available from: <http://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/> [Accessed 3rd May 2023].
- [10] Gowda M, Dhekne A, Shen S, Choudhury RR, Yang X, Yang L, et al. IoT platform for sports analytics. *GetMobile: Mobile Computing and Communications*. 2018; 21(4): 8-14. Available from: doi: 10.1145/3191789.3191793.
- [11] Jamil F, Iqbal N, Ahmad S, Kim DH. Toward accurate position estimation using learning to prediction algorithm in indoor navigation. *Sensors*. 2020; 20(16): 4410. Available from: doi: 10.3390/S20164410.
- [12] Weir G, Alderson J, Smailes N, Elliott B, Donnelly C. A reliable video-based ACL injury screening tool for female team sport athletes. *International Journal of Sports Medicine*. 2019; 40(3): 191-199. Available from: doi: 10.1055/A-0756-9659/ID/R6900-0025.
- [13] Rommers N, Rössler R, Verhagen E, Vandecasteele F, Verstockett S, Vaeyens R, et al. machine learning approach to assess injury risk in elite youth football players. *Medicine and Science in Sports and Exercise*. 2020; 52(8): 1745-1751. Available from: doi: 10.1249/MSS.0000000000002305.
- [14] Ahmadi A, Mitchell E, Richter C, Destelle F, Gowing M, O'Connor NE, et al. Toward automatic activity classification and movement assessment during a sports training session. *IEEE Internet of Things Journal*. 2015; 2(1): 23-32. Available from: doi: 10.1109/JIOT.2014.2377238.
- [15] Ahammad I, Rahman Khan MA, Salehin ZU. Advancement of IoT system QoS by integrating cloud, fog, roof, and dew computing assisted by SDN. *International Journal of Ambient Computing and Intelligence (IJACI)*. 2021; 12(4): 132-153. Available from: doi: 10.4018/IJACI.2021100108.
- [16] Ahammad I, Khan MAR, Salehin ZU. QoS performance enhancement policy through combining fog and SDN. *Simulation Modelling Practice and Theory*. 2021; 109: 102292. Available from: doi: 10.1016/J.SIMPAT.2021.102292.
- [17] Yang Z. Research on basketball players' training strategy based on artificial intelligence technology. *Journal of Physics: Conference Series*. 2020; 1648(4): 042057. Available from: doi: 10.1088/1742-6596/1648/4/042057.
- [18] Deep C, Patvardhan C, Vasantha C. Data analytics based deep mayo predictor for IPL-9. *International Journal of Computer Applications*. 2016; 152(6): 6-11. Available from: doi: 10.5120/IJCA2016911875.
- [19] Hvattum LM, Arntzen H. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*. 2010; 26(3): 460-470. Available from: doi: 10.1016/J.IJFORECAST.2009.10.002.
- [20] Singh T, Singla V, Bhatia P. Score and winning prediction in cricket through data mining. *International Conference*

on Soft Computing Techniques and Implementations, ICSCIT, 2015. India, Faridabad: Institute of Electrical and Electronics Engineers Inc.; 2016. p. 60-66. Available from: doi: 10.1109/ICSCIT.2015.7489605.

- [21] Rahman K, Khan Md AR, Ahammad I. Online support for education, medication, agriculture and relief work at COVID-19 pandemic time. *Asian Journal of Research in Computer Science*. 2022; 13(4): 65-85. Available from: doi: 10.9734/ajrcos/2022/v13i430322.
- [22] Prosad R, Khan Md AR, Ahammad I. Design of class routine and exam hall invigilation system based on genetic algorithm and greedy approach. *Asian Journal of Research in Computer Science*. 2022; 13(3): 28-44. Available from: doi: 10.9734/ajrcos/2022/v13i330316.
- [23] Hossain Md A, Ahammad I, Ahmed Md K, Ahmed MI. Prediction of the computer science department's educational performance through machine learning model by analyzing students' academic statements. *Artificial Intelligence Evolution*. 2023; 4(1): 70-87. Available from: doi: 10.37256/aie.4120232569.
- [24] Khan AR, Uzzaman F, Ahammad I, Prosad R, Khan TZ, Ejaz Md S, et al. Stock market prediction in bangladesh perspective using artificial neural network. *International Journal of Advanced Technology and Engineering Exploration*. 2022; 9 (95): 1397-1427. Available from: doi: 10.19101/IJATEE.2021.875852.
- [25] Khan MAR, Akter J, Ahammad I, Ejaz S, Jaman Khan T. Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer perceptron NN and decision tree. *Health Information Science and Systems*. 2022; 10(1): 1-18. Available from: doi: 10.1007/S13755-022-00202-X/METRICS.
- [26] Khan Md AR, Afrin F, Prity FS, Ahammad I, Fatema S, Prosad R, et al. An effective approach for early liver disease prediction and sensitivity analysis. *Iran Journal of Computer Science*. 2023; 6: 1-19. Available from: doi: 10.1007/S42044-023-00138-9.
- [27] Mavrikos E, Melissaris N, Tsekouras G. A hybrid clustering-based approach to color modification of art paintings for protanopes and deuteranopes. *Cloud Computing and Data Science*. 2022; 3(2): 13-38. Available from: doi: 10.37256/ccds.3220221235.
- [28] Khamkar R, Das P, Namasudra S. SCEOMOO: A novel subspace clustering approach using evolutionary algorithm, off-spring generation and multi-objective optimization. *Applied Soft Computing*. 2023; 139: 110185. Available from: doi: 10.1016/j.asoc.2023.110185.
- [29] Koseler K, Stephan M. *Towards the Realization of a DSML for Machine Learning: A Baseball Analytics Use Case*. 2017. Available from: <http://hdl.handle.net/2374.MIA/6224> [Accessed 3rd May 2023].
- [30] Zimmermann-Niefield A, Shapiro RB, Kane S. Sports and machine learning: How young people can use data from their own bodies to learn about machine learning. *XRDS: Crossroads, The ACM Magazine for Students*. 2019; 25(4): 44-49. Available from: doi: 10.1145/3331071.
- [31] Kyle Martin R, Pareek A, Krych AJ, Kremers HM, Engebretsen L. Machine learning in sports medicine: Need for improvement. *Journal of ISAKOS*. 2021; 6(1): 1-2. Available from: doi: 10.1136/jisakos-2020-000572.
- [32] Pappalardo L, Cintia P, Ferragina P, Massucco E, Pedreschi D, Giannotti F. PlayeRank. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2019; 10(5): 59. Available from: doi: 10.1145/3343172.
- [33] Bunker RP, Thabtah F. A machine learning framework for sport result prediction. *Applied Computing and Informatics*. 2019; 15(1): 27-33. Available from: doi: 10.1016/J.ACI.2017.09.005.
- [34] Luu BC, Wright AL, Haeberle HS, Karnuta JM, Schickendantz MS, Makhni EC, et al. Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: An analysis of 2322 players from 2007 to 2017. *Orthopaedic Journal of Sports Medicine*. 2020; 8(9): 2325967120953404. Available from: doi: 10.1177/2325967120953404.
- [35] Reichherzer T, Timm M, Earley N, Reyes N, Kumar V. Using machine learning techniques to track individuals & their fitness activities. In: *Proceedings of 32nd International Conference on Computers and Their Applications, March, 20-22, 2017, Honolulu, Hawaii, USA*. Winona, MN, USA: ISCA; 2017. p. 119-124.
- [36] Lee G, Bulitko V, Ludvig E. Sports Commentary Recommendation System (SCoReS): Machine learning for automated narrative. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. AAAI; 2012. p. 32-37. Available from: doi: 10.1609/AIIDE.V8I1.12505.
- [37] Deep Prakash C, Majumdar A. Analyzing the role of national culture on content creation and user engagement on Twitter: The case of Indian Premier League cricket franchises. *International Journal of Information Management*. 2021; 57: 102268. Available from: doi: 10.1016/J.IJINFOMGT.2020.102268.
- [38] Al-Shboul R, Syed T, Memon J, Khan F. Automated player selection for a sports team using competitive neural networks. *International Journal of Advanced Computer Science and Applications*. 2017; 8(8): 457-460. Available from: doi: 10.14569/IJACSA.2017.080859.
- [39] Beal R, Norman TJ, Ramchurn SD. Optimising daily fantasy sports teams with artificial intelligence. *International*

Journal of Computer Science in Sport. 2020; 19(2): 21-35. Available from: doi: 10.2478/IJCSS-2020-0008.

- [40] Kampakis S, Thomas W. *Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches*. 2015. Available from: <https://arxiv.org/abs/1511.05837v1> [Accessed 3rd May 2023].
- [41] Passi K, Pandey N. Increased prediction accuracy in the game of cricket using machine learning. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. 2018; 8(2): 19-36. Available from: doi: 10.5121/ijdkp.2018.8203.
- [42] Passi K, Pandey N. Predicting players' performance in one day international cricket matches using machine learning. *Computer Science & Information Technology (CS & IT)*. 2018; 111-126. Available from: doi: 10.5121/csit.2018.80310.
- [43] Prakash CD, Patvardhan C, Singh S. A new machine learning based deep performance index for ranking IPL T20 cricketers. *International Journal of Computer Applications*. 2016; 137(10): 42-49. Available from: doi: 10.5120/ijca2016908903.
- [44] Gucciardi DF, Jones MI. Beyond optimal performance: Mental toughness profiles and developmental success in adolescent cricketers. *Journal of Sport and Exercise Psychology*. 2012; 34(1): 16-36. Available from: doi: 10.1123/JSEP.34.1.16.
- [45] Shah P. New performance measure in Cricket. *IOSR Journal of Sports and Physical Education IOSR-JSPE*. 2017; 4(3): 28-30. Available from: doi: 10.9790/6737-04032830.
- [46] Jhansi Rani P, Vidyadhar Kamath A, Menon A, Dhatwalia P, Rishabh D, Kulkarni A. Selection of players and team for an Indian premier league cricket match using ensembles of classifiers. In: *Proceedings of CONECCT 2020 - 6th IEEE International Conference on Electronics, Computing and Communication Technologies*. India, Bangalore: Institute of Electrical and Electronics Engineers Inc.; 2020. Available from: doi: 10.1109/CONECCT50063.2020.9198371.
- [47] Parker D, Burns P, Natarajan H. Player valuations in the Indian premier league. *Frontier Economics*. 2008; 1-17. Available from: <https://www.scribd.com/document/375714267/player-valuations-in-the-indian-premier-league-frontier-paper-pdf#> [Accessed 3rd May 2023].
- [48] Bhattacharjee D, Pahinkar DG. Analysis of performance of bowlers using combined bowling rate. *International Journal of Sports Science and Engineering*. 2012; 6(3): 184-192.
- [49] Jhanwar MG, Pudi V. Predicting the outcome of ODI cricket matches: A team composition based approach. In: *Machine Learning and Data Mining for Sports Analytics, ECML-PKDD'16*. 2016.
- [50] Mukherjee S. Quantifying individual performance in Cricket — A network analysis of batsmen and bowlers. *Physica A: Statistical Mechanics and its Applications*. 2014; 393: 624-637. Available from: doi: 10.1016/J.PHYSA.2013.09.027.
- [51] Barr GDI, Kantor BS. A criterion for comparing and selecting batsmen in limited overs cricket. *The Journal of Operational Research Society*. 2004; 55(12): 1266-1274. Available from: doi: 10.2307/2983068.
- [52] Sankaranarayanan VV, Sattar J, Lakshmanan LVS. Auto-play: A data mining approach to ODI cricket simulation and prediction. In: *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*. Society for Industrial and Applied Mathematics Publications; 2014. p. 1064-1072. Available from: doi: 10.1137/1.9781611973440.121.
- [53] Muthuswamy S, Lam SS. Bowler performance prediction for one-day international cricket using neural networks. In: *Proceedings of the 2008 Industrial Engineering Research Conference*. NY, Binghamton; 2008. p. 1391-1395.
- [54] Lemmer HH. The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation*. 2002; 24(2): 37-44. Available from: doi: 10.4314/sajrs.v24i2.25839.
- [55] Wickramasinghe IP. Predicting the performance of batsmen in test cricket. *Journal of Human Sport and Exercise*. 2014; 9(4): 744-751. Available from: doi: 10.14198/JHSE.2014.94.01.
- [56] Anik AI, Yeaser S, Imam Hossain AGM, Chakrabarty A. Player's performance prediction in ODI cricket using machine learning algorithms. In: *4th International Conference on Electrical Engineering and Information and Communication Technology, iCEEICT 2018*. Institute of Electrical and Electronics Engineers Inc.; 2019. p. 500-505. Available from: doi: 10.1109/CEEICT.2018.8628118.
- [57] Iyer SR, Sharda R. Prediction of athletes performance using neural networks: An application in cricket team selection. *Expert Systems with Applications*. 2009; 36(3): 5510-5522. Available from: doi: 10.1016/J.ESWA.2008.06.088.
- [58] Ovens M, Bukiet B. A mathematical modelling approach to one-day cricket batting orders. *Journal of Sports Science and Medicine*. 2006; 5(4): 495.

- [59] Hossain MJ, Kashem MA, Islam MS, Jannat ME. Bangladesh cricket squad prediction using statistical data and genetic algorithm. In: *4th International Conference on Electrical Engineering and Information and Communication Technology, iCEEICT 2018*. Bangladesh, Dhaka: Institute of Electrical and Electronics Engineers Inc.; 2019. p. 178-181. Available from: doi: 10.1109/CEEICT.2018.8628076.
- [60] Ahmad H, Daud A, Wang L, Hong H, Dawood H, Yang Y. Prediction of rising stars in the game of cricket. *IEEE Access*. 2017; 5: 4104-4124. Available from: doi: 10.1109/ACCESS.2017.2682162.
- [61] Sumathi M, Prabu S, Rajkamal S. Cricket players performance prediction and evaluation using machine learning algorithms. In: *2023 International Conference on Networking and Communications (ICNWC)*. Chennai, India: IEEE; 2023. p. 1-6. Available from: doi: 10.1109/ICNWC57852.2023.10127503.
- [62] Ishi M, Patil J, Patil V. An efficient team prediction for one day international matches using a hybrid approach of CS-PSO and machine learning algorithms. *Array*. 2022; 14: 100144. Available from: doi: 10.1016/j.array.2022.100144.