AutoCoach: An Automated Playing Eleven Selection Framework in Cricket

Pulkit Sharma
School of Electronic Engineering and Computer Science
Queen Mary University
London, United Kingdom

Bilal Hassan
School of Computing and Digital Media
London Metropolitan University
London, United Kingdom

Muhammad Farooq Wasiq
Department of Creative Technologies
Air University
Islamabad, Pakistan

***Abstract*— Cricket is one of the famous international sports where both teams select their best eleven against each other from the available pool of players usually a total size of 15 to 16. The team selection process is usually performed by the coaches using their observation and consultations etc. In simple words, by looking at the possible combinations of opposite sides, the best possible combinations are suggested. In our work, we tried to automate this process using machine learning models like Support Vector Regressor, Linear Regression Random Forest regressor etc. To train and test the model, the data was crawled from the Indian Premier League (IPL), which follows 20 over format and data for more than 10 seasons is used. More importantly, multiple related features for batters and bowlers were accumulated to develop all possible combinations for each playing team in a single match. In total, we selected 4 different teams and developed three combinations for each. Then, in a single match between two teams, the combinations were compared based upon accumulative score e.g., matching scores for different combinations are compared with ground truth. In our opinion, this is one of the unique contributions made for automated team selection in cricket specific to the format and encompassing performance indicators not only accumulated from IPL but also from the International Cricket Council (ICC).**

Keywords— Cricket, Team Selection, Regression, SVM, Random Forest

I. INTRODUCTION

Cricket One of the most popular sports in the world now is cricket. The Twenty20 format has had significant fan support in recent years, generating significant cash from sponsorships and international media participation. According to the latest estimates for the income of the International Cricket Council and the Indian Premier League (IPL) as of 2018, the sport is worth £5.983 billion in India alone.

We decided to conduct our research using the IPL dataset because it gives players the essential stage on which to display their ability. The selection committee, which is made up of seasoned cricketers, chooses the squad with great care, taking into consideration several player factors such as past performance, morale, abilities, and traits that are particularly useful for team building. Using workshops and lengthy talks, a fifteen-man roster is chosen from among hundreds of eligible individuals. The process of choosing a cricket team is still subjective when there is no suitable approach in place. Since the sport currently produces a significant amount of data, it is essential to examine this unexplored source and get precise selection results. Researchers have looked into several cricket-related numerical characteristics throughout many eras. Lemmer and colleagues state that other writers have previously developed several performance indicators in this field [1]. Lemmer devised a method for evaluating a bowler's performance that combines the three traditional bowling attributes—bowling strike rate, bowling average, and bowling economy rate—and refers to it as a combined bowling rate. He used the harmonic mean approach to get an aggregate bowling rate [2]. In addition to this research, he made an effort to assess batsmen's batting performance in the limited-overs cricket format [3]. Pranavan, et al [4] looked at the collection of characteristics that significantly affect a game's outcome using machine learning. Conversely, Lewis's [5] measure is predicated on ball-by-ball data from the match Nevertheless, gathering ball-by-ball data from a match or a sequence of matches takes a lot of time. Farhana et al. [6] presented a support vector machine-based method wherein a user selects the playing 11 from a rated roster of players according to their skill level. To accomplish the optimization, Singla [7] employed the Gurib Python module. Shah et al. [8] in their investigations of whether players' batting performances are more important than their bowling performances.

Hossain et al. [9] employed a genetic algorithm, which is employed in the process of choosing the top 30 cricket players for Bangladesh. Sharp et al. [10] put out an approach that uses integer programming to solve the maximizing problem of T20 team selection, selecting players based on a certain skill set. Kamble et al. [11] put out a hierarchical analytical method for selecting a specific group of cricket players. Another study [12] used integer programming to choose the players they wanted for a cricket fantasy league squad. Lemmer [13] integrates the batsman's performance and the probability of dismissal for a coherent and integrated assessment of all the performance parameters of a wicketkeeper. Dey et al. [14] have created a strict method based on the Analytical Hierarchy Process (AHP) for determining player price in the IPL. To gauge the players' effectiveness, Chaudhary [15] employed a Data Envelopment Analysis (DEA). To determine the participants' ranking, C. D. Prakash [16] introduced a Recursive Feature estimate approach based on Random Forests and employed a bespoke set of metrics that took into account both T20 International and IPL matches. In the study [17], authors developed a new Deep Performance Index using a machine learning technique, which is used to rate bowlers and batters in Twenty20 cricket. Prior studies have concentrated on the usual player traits that are available, but none have examined the competing side pool. The innovative technique that this study suggests aims to apply machine learning models to various input variables. This innovative method will be utilized to determine various indices and project player ratings to assess each player's performance for the upcoming season. An automated method will then be employed to determine the best eleven-man team to play against a variety of opponent combinations.

II. METHODOLOGY

The proposed framework in our research considers various phases which are represented in the architecture (Fig. 1). The methodology conducted starts with designing a crawler for data extraction, followed by data cleaning and preprocessing, extracting the essential features, normalizing the data, calculating players rating with new techniques, data splitting into training and validation, applying three modules for future rating calculation of the players for performance evaluation and finally using the algorithm for selecting the optimum team against the various combinations of the opponent.

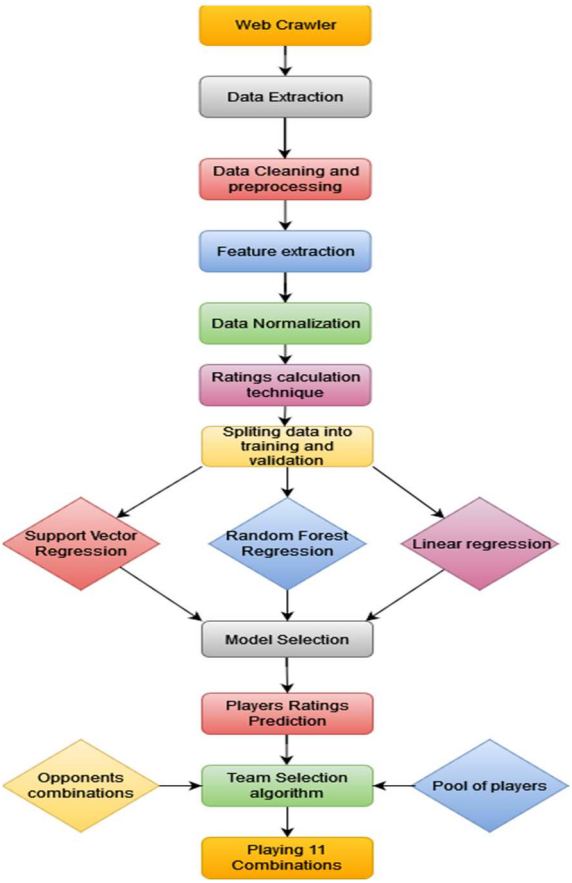

Fig. 1. Flowchart of methodology

A. Data Extraction and Preprocessing

A web crawler was made to produce the dataset. We created a crawler to collect player data from websites like espncricinfo.com and crickinfo.com on every player that participated in the Indian Premier League. The players' extracted dataset includes information from the last twelve IPL seasons, which runs from 2008 to 2019. For bowlers, all-rounders, and batters, we used twelve qualities , see Fig. 2 and Fig. 3.

Fig. 2. Batsmen Attributes

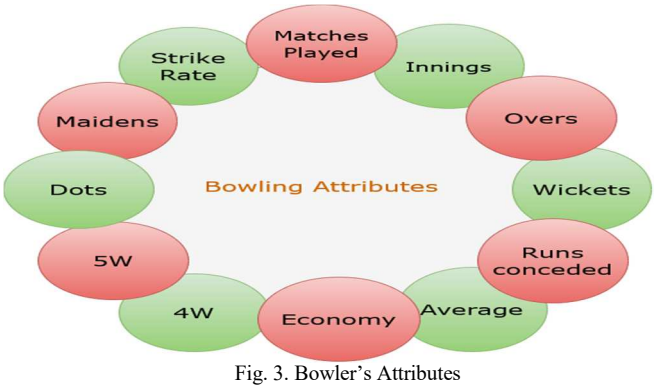

Fig. 3. Bowler's Attributes

B. Data Normalization

Data normalization is used to address these inconsistencies in data as the attributes are linked to various units of measurement. A positive or negative influence may be related to many factors that are taken into consideration with the newly established indices when assessing the overall rating of the individual players. The batting average, runs scored, and total number of boundaries scored are among the metrics that positively correlate with a player's talent. Conversely, the quantity of runs given up and the economy rate are negatively correlated with the player's ability to bowl. Therefore, due consideration must be used while using the normalization procedure for such parameters. The features normalized as follows if it has a positive correlation with the player's ability:

$$\frac{X - \min(X)}{\max(X) - \min(X)} \qquad (1)$$

The characteristic is normalized when it has a negative correlation with the player's skill in the following way:

$$\frac{\max(X) - X}{\max(X) - \min(X)} \qquad (2)$$

Following the player's association with the attributes, we determined the normalization value using the aforementioned formulae, where X denotes the attribute's value, min(X) denotes its minimum value, and max(X) denotes its highest value. To use the normalized data for the rating calculation approach, we computed them using the aforementioned formulae.

C. Ratings Calulation technique

While simple averages give each component the same weight when calculating ratings, a weighted composite measure enables the relative significance of the factors to be assessed. When evaluating players in order to determine ratings, we take into account the relevant player traits. Many characteristics fall under the various player classifications of bowlers, all-rounders, and batters. The addition of weight to a player's qualities aids in assigning equal significance to other attributes. We have put out a novel approach and assigned relative weights to various qualities in order to determine the player ratings for each of the twelve Indian Premier League seasons. Six key characteristics were taken into consideration to assess the batting ratings. The formula that shows six attributes of a batsman is given below:

$$RS = Number\ of\ runs\ scored$$

$$SR = \frac{Total\ numbers\ of\ runs\ scored}{Total\ number\ of\ balls\ faced} \quad (3)$$

$$Milestone\ rate = \frac{100's + 50's}{Number\ of\ innings} \quad (4)$$

$$4B = Number\ of\ four\ boundaries$$

$$6B = Number\ of\ six\ boundaries$$

$$NO = Number\ of\ innings\ played - Outs \quad (5)$$

Based on the batsmen's qualities listed above. The runs that the batters have scored in each and every game is referred to as the RS. The term "strike rate" (SR) refers to the ratio of the total number of balls faced to the total number of runs achieved by the batters. The ratio of the total number of games played to the sum of the batsmen's hundreds and half-centuries is known as the milestone rate. The total amount of fours that the batters have hit is 4B. The total amount of sixes that the batters have hit is 6B. The word NO represents the number of not-outs, which is the result of subtracting the batter's innings played and the number of occasions they were out. The player ratings were multiplied by the corresponding weights and added for the purpose of calculating the ratings. Using the criteria listed above, the rating for batting is determined as follows:

$$\boldsymbol{BTAR} = 1 * RS + 1 * SR + 1 * SR + 1 * MR + 4 * 4B + 6 * 6B + 10 * NO \quad (6)$$

Similar to this, we took into account a total of six essential features while determining the bowler's rating. The following defines the six characteristics of the bowlers:

$$SR = \frac{Number\ of\ bowls\ bowled}{Number\ of\ wickets\ taken} \quad (7)$$

$$WT = Number\ of\ wickets\ taken$$

$$MO = Number\ of\ maiden\ overs$$

$$4W = Number\ of\ four\ wickets\ taken$$

$$5W = Number\ of\ five\ wickets\ taken$$

$$DB = Number\ of\ dot\ balls$$

According to the bowler's previously mentioned characteristics, SR stands for strike rate, which is calculated as the ratio of bowls bowled to wickets claimed. WT is a characteristic that indicates how many wickets the bowler has taken. The quantity of overs in which no runs are given up defines MO. 4 wickets obtained in a match is known as 4W. The match's five wicket total is represented by the letter 5W. Each related weightage seeks to give each characteristic equal weight, preventing any one attribute from outweighing the others and devaluing the ratings for each one individually. The player ratings were multiplied by the corresponding weights and added linearly for the rating computation. The following formula is used to determine the Bowling (BWAR) rating using the previously listed attributes:

$$\boldsymbol{BWAR} = 1 * SR + 10 * WT + 5 * MO + 40 * 4W + 50 * 5W + 1 * DB \quad (8)$$

Lastly, all-rounders are described as players who score over the thresholds for both the batting-associated rating and the bowling-associated rating. Future rating projections are also made using the computed batting and bowling linked ratings.

D. Model Selection and Prediction

The ratings for the upcoming season are calculated by dividing the dataset into two categories: 75% of the dataset is designated as the training dataset and the rest of 25% as the validation dataset. This process is done after the ratings of each player in each of the twelve IPL seasons are calculated in separate categories. To forecast the ratings for the upcoming season, we have put three distinct machine learning models into practice utilizing all of the crucial characteristics of each individual player. A measure of accuracy used to compare various models was the mean squared error. The supervised learning strategy known as "Support Vector Regression" is used to predict discrete values. SVM and SVR have the same base. Finding the best-fitting line is support vector regression's main goal. To obtain a minimal mean square error, we experimented with various parameter settings when putting our model into practice. We used various kernels to the model, including sigmoid, poly, rbf, and linear. When we used gamma as the scale, 1.0 as the regularization parameter, and 'rbf' as the kernel to estimate each player's individual rating for the upcoming season, a smallest mean squared was produced. The mean squared error was 3.47 for the bowler's ratings and 5.65 for the batsmen's.

We employed the multiple linear regression model in order to lower the error in predicting individual player ratings for the next Indian Premier League season. Within the fields of machine learning and statistics, multiple linear regression is among the most widely used and comprehended methods. If the objective is prediction, forecasting, or error reduction, multiple linear regression can be used to build a predictive model to an observed dataset made up of response and explanatory variable values. The connection in multiple linear regression is demonstrated under the assumption that both independent and response variables have a linear relationship. Using the key characteristics of the multiple linear regression model, we were able to forecast the individual player ratings for the upcoming season. The mean squared error for the bowler's ratings was 4.01 and for the batsmen's ratings, it was 7.82.

We employed the random forest regression model to further reduce the mean square for the forecast of the individual ratings in the next Indian Premier League season. When dealing with issues involving the training of several decision trees, such as regression and classification, random forests are employed as part of the ensemble learning technique. Throughout the training phase, this model builds a large number of decision trees, and the projection of each tree produces an output that is the mean of the classes. Using the key characteristics of the random forest regression model, we projected each player's ratings for the upcoming season. We used n_estimators of 1000 and a random state of 42, and the mean squared error for batsmen's ratings was found to be 4.52, while it was found to be 3.25 for bowlers. The random forest regressor produced the smallest error

within the range of models that were employed. As a result, we chose the model to forecast player ratings for the upcoming IPL season, which will be utilized in the selection process.

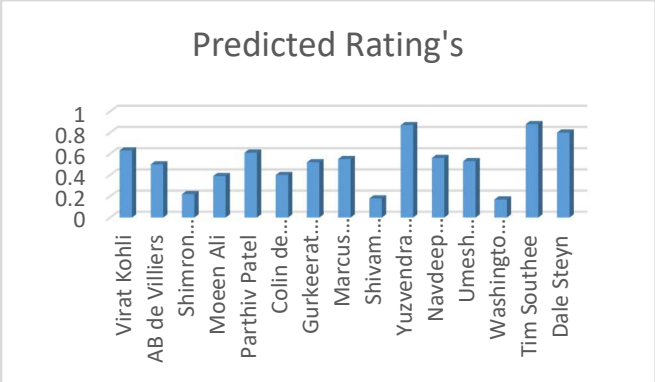

Fig. 4. Bowler's Attributes

E. Team Selection Algorithm

The data frame containing the projected ratings for the upcoming season was then applied to four teams. The sides engaged in a total of six matches. The strongest playing eleven is chosen from a roster of fifteen players for each squad. The three combinations that made up each side were then separated into three groups: bowlers, batsmen, and all-rounders. As a result, there are currently thirty-six (36) distinct categories of batsmen, bowlers, and all-rounders in addition to twelve (12) possible team combinations, see Fig. 5.

Two teams were considered in the process: our target team, designated as Team B, and the opposing team, designated as Team A. Additionally, three iterations of the opposing team underwent testing. Three groups of the opposing team's batsmen, bowlers, and all-rounders were assigned to each combination. We have fifteen players on our team B roster, divided into three groups: bowlers, all-rounders, and batters. All three categories were compared using the team selection method. The ratings of team B and the opposing team, team A, are kept in an array, for example, when team selection takes into account the category of batsmen. To record the ratings of Team B that are higher than those of Team A, a null array called "c" has been constructed. Following this, we start the for-loop, which compares each batsman's rating on team A with each batsman's rating on team B. Following this comparison, all of the ratings that were either higher or equal to those of the opposing team A were chosen. They were also double-checked to make sure they were in the c array; if they weren't, the chosen value would be added. Subsequently, the algorithm determines if the batters in our team B, which are generated by inserting certain values after comparison, are greater than the batsmen in the opponent team A. The programmed looped over and over again as the difference, calculating the remaining number of batters (Flag) between our side B and the other team A The cycle was continued until the flag reached zero, at which point the number of batters left were chosen from the rosters of the fifteen teams. The c array now contains their ratings. Following this selection procedure, the ratings were compared to the original team data frame to obtain the players' real names, which allowed Team B to choose its batters versus Team A. The eleven-player pick from the group of surviving bowlers and all-rounders was made using the same algorithm once again. As a result, the best possible playing eleven was chosen to oppose the opponent team's lineup. This algorithm, which has 36 subcategories and twelve combinations, was performed for each of the four teams. The following lists the results of the three combinations of the opposition team A and the best team chosen for team B, see Fig. 6 – Fig. 8.

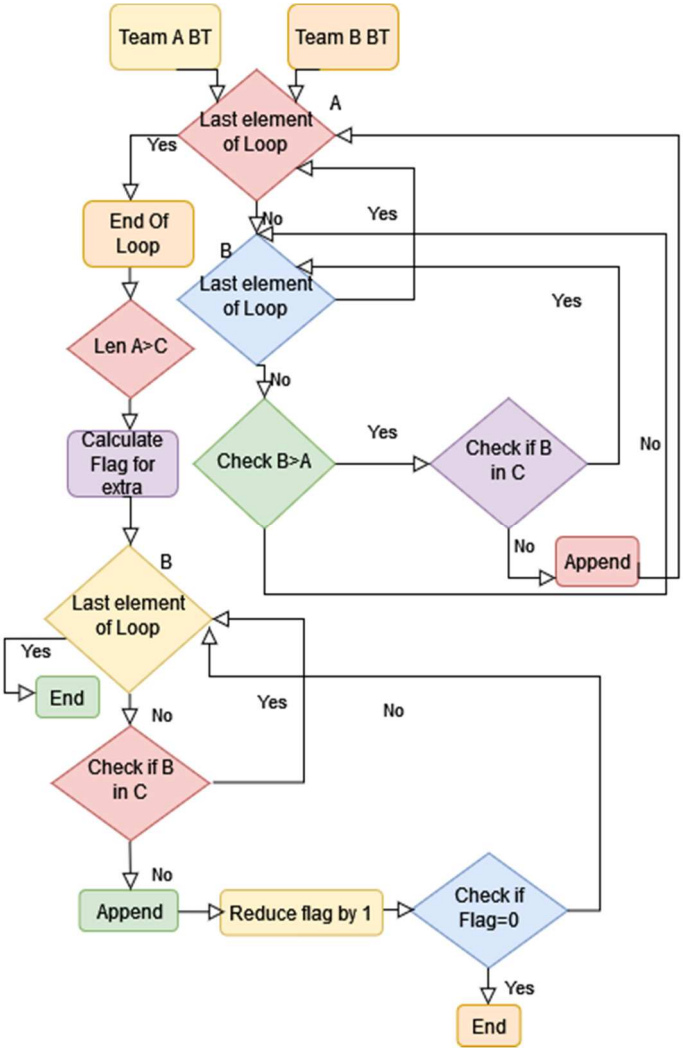

Fig. 5. Team selection algorithm

	Player	Ratings	Role		Player	Ratings	Role
11	Shradul Thakur	0.91	BW	0	Virat Kohli	0.63	BT
7	Shane Watson	0.52	BT	1	AB de Villiers	0.50	BT
8	Murali Vijay	0.70	BT	2	Shimron Hetmyer	0.22	BT
3	Faf du Plessis	0.98	BT	3	Parthiv Patel	0.61	BT
1	Suresh Raina	0.47	BT	4	Colin de Grandhomme	0.40	BT
9	Imran Tahir	0.98	BW	5	Yuzvendra Chahal	0.87	BW
2	Ambati Rayudu	0.53	BT	6	Navdeep Saini	0.56	BW
10	Harbhajan Singh	0.47	BW	7	Umesh Yadav	0.53	BW
12	Mitchell Santner	0.20	BW	8	Washington Sundar	0.17	BW
13	Deepak Chahar	0.90	BW	9	Tim Southee	0.88	BW
4	Dwayne Bravo	1.00	ALL	10	Moeen Ali	0.39	ALL

a) Team A1 *b) Team B1*

Fig. 6. Team A1 and B1 Best Team Assesment

	Player	Ratings	Role		Player	Ratings	Role
6	Ravindra Jadeja	0.94	ALL	0	Virat Kohli	0.63	BT
11	Shradul Thakur	0.91	BW	1	AB de Villiers	0.50	BT
7	Shane Watson	0.52	BT	2	Shimron Hetmyer	0.22	BT
10	Harbhajan Singh	0.47	BW	3	Parthiv Patel	0.61	BT
12	Mitchell Santner	0.20	BW	4	Colin de Grandhomme	0.40	BT
3	Faf du Plessis	0.98	BT	5	Gurkeerat Mann Singh	0.52	BT
0	MS Dhoni	0.47	BT	6	Shivam Dube	0.18	BT
9	Imran Tahir	0.98	BW	7	Yuzvendra Chahal	0.87	BW
4	Dwayne Bravo	1.00	ALL	8	Umesh Yadav	0.53	BW
2	Ambati Rayudu	0.53	BT	9	Navdeep Saini	0.56	BW
1	Suresh Raina	0.47	BT	10	Moeen Ali	0.39	ALL

a) Team A2 *b) Team B2*

Fig. 7. Team A2 and B2 Best Team Assesment

	Player	Ratings	Role		Player	Ratings	Role
5	Kedar Jadhav	0.96	BT	0	Virat Kohli	0.63	BT
8	Murali Vijay	0.70	BT	1	AB de Villiers	0.50	BT
7	Shane Watson	0.52	BT	2	Shimron Hetmyer	0.22	BT
1	Suresh Raina	0.47	BT	3	Parthiv Patel	0.61	BT
2	Ambati Rayudu	0.53	BT	4	Colin de Grandhomme	0.40	BT
11	Shradul Thakur	0.91	BW	5	Gurkeerat Mann Singh	0.52	BT
10	Harbhajan Singh	0.47	BW	6	Yuzvendra Chahal	0.87	BW
14	Karn Sharma	0.94	BW	7	Navdeep Saini	0.56	BW
6	Ravindra Jadeja	0.94	ALL	8	Umesh Yadav	0.53	BW
12	Mitchell Santner	0.20	BW	9	Washington Sundar	0.17	BW
3	Faf du Plessis	0.98	BT	10	Moeen Ali	0.39	ALL

a) Team A3 *b) Team B3*

Fig. 8. Team A3 and B3 Best Team Assesment

Three teams that were chosen by the model and three teams that were made as a baseline and compared to assess the correctness of the selection process see Fig. 9 - Fig.11. Quantifying the degree of improvement in the chosen team's success rate was made easier by comparison with the ground reference. The corresponding teams are listed below, along with a comparison between them and the optimal squad that was chosen following algorithm execution:

	Player	Ratings	Role		Player	Ratings	Role
0	Virat Kohli	0.63	BT	0	Parthiv Patel	0.61	BT
1	AB de Villiers	0.50	BT	1	Shimron Hetmyer	0.22	BT
2	Shimron Hetmyer	0.22	BT	2	AB de Villiers	0.50	BT
3	Parthiv Patel	0.61	BT	3	Virat Kohli	0.63	BT
4	Colin de Grandhomme	0.40	BT	4	Shivam Dube	0.18	BT
5	Yuzvendra Chahal	0.87	BW	5	Umesh Yadav	0.53	BW
6	Navdeep Saini	0.56	BW	6	Washington Sundar	0.17	BW
7	Umesh Yadav	0.53	BW	7	Tim Southee	0.88	BW
8	Washington Sundar	0.17	BW	8	Dale Steyn	0.80	BW
9	Tim Southee	0.88	BW	9	Marcus Stoinis	0.55	ALL
10	Moeen Ali	0.39	ALL	10	Moeen Ali	0.39	ALL

The contrast between the teams chosen by the algorithm and the teams used as a baseline for validation is shown in the above graph. The graph demonstrates how precisely the algorithm chose the best squad to face the various opponent combinations. The contrast between the teams chosen by the algorithm and the teams used as a baseline for validation is shown in the above graph. The graph demonstrates how precisely the algorithm chose the best squad to face the various opponent combinations.

a) Team B1 *b) Team GR1*

Fig. 9. Team B1 and GRI Assesment

	Player	Ratings	Role		Player	Ratings	Role
0	Virat Kohli	0.63	BT	0	AB de Villiers	0.50	BT
1	AB de Villiers	0.50	BT	1	Shimron Hetmyer	0.22	BT
2	Shimron Hetmyer	0.22	BT	2	Virat Kohli	0.63	BT
3	Parthiv Patel	0.61	BT	3	Colin de Grandhomme	0.40	BT
4	Colin de Grandhomme	0.40	BT	4	Moeen Ali	0.39	ALL
5	Gurkeerat Mann Singh	0.52	BT	5	Marcus Stoinis	0.55	ALL
6	Shivam Dube	0.18	BT	6	Tim Southee	0.88	BW
7	Yuzvendra Chahal	0.87	BW	7	Parthiv Patel	0.61	BT
8	Umesh Yadav	0.53	BW	8	Navdeep Saini	0.56	BW
9	Navdeep Saini	0.56	BW	9	Umesh Yadav	0.53	BW
10	Moeen Ali	0.39	ALL	10	Washington Sundar	0.17	BW

a) Team B2 *b) Team GR2*

Fig. 10. Team B2 and GR2 Assesment

	Player	Ratings	Role		Player	Ratings	Role
0	Virat Kohli	0.63	BT	0	Shimron Hetmyer	0.22	BT
1	AB de Villiers	0.50	BT	1	Virat Kohli	0.63	BT
2	Shimron Hetmyer	0.22	BT	2	Colin de Grandhomme	0.40	BT
3	Parthiv Patel	0.61	BT	3	Shivam Dube	0.18	BT
4	Colin de Grandhomme	0.40	BT	4	Moeen Ali	0.39	ALL
5	Gurkeerat Mann Singh	0.52	BT	5	Navdeep Saini	0.56	BW
6	Yuzvendra Chahal	0.87	BW	6	Tim Southee	0.88	BW
7	Navdeep Saini	0.56	BW	7	Umesh Yadav	0.53	BW
8	Umesh Yadav	0.53	BW	8	AB de Villiers	0.50	BT
9	Washington Sundar	0.17	BW	9	Yuzvendra Chahal	0.87	BW
10	Moeen Ali	0.39	ALL	10	Washington Sundar	0.17	BW

a) Team B3 *b) Team GR3*

Fig. 11. Team B3 and GR3 Assesment

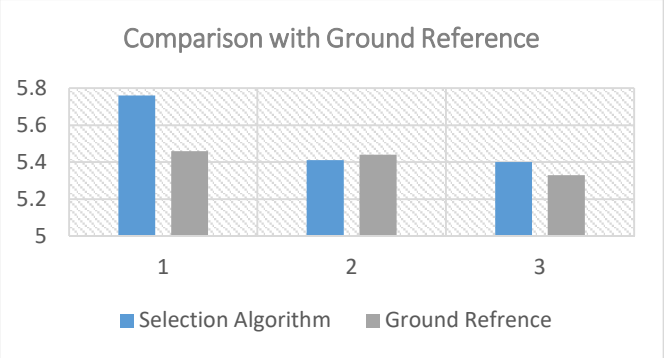

Fig. 4. Comparison with Ground Reference

III. CONCLUSION & FUTURE WORK

Multiple linear regression, linear regression, support vector regression, and polynomial regression approaches were all tested in this work, and polynomial regression came to be the best model for predicting score (MSE= 94.5). SVM, Gaussian Nave Byes classification, k-NN classification, and deep learning with 3-layered neural network are also tested, with the best model for predicting match outcome (Accuracy= 91.46%) being identified as 3-layered NN. This paper presents an AI alternative that takes into consideration all additional

parameters, such as pitch, toss, and first-inning score and raises issues about the usage of the traditional D/L approach for incomplete matches. This study uses only the IPL dataset, with T20 cricket showing the greatest outcomes. Previous research produced results with higher computing costs and fewer features that were less accurate; however, my proposed method with a wider feature set produced results with lower computation costs and higher accuracy. This study may be of particular benefit to cricket club managers, sports data analysts, and scholars interested in sports analysis. We can extend this research to different formats of cricket like the ODI, and Test, by utilizing additional attributes like the weather and pitch conditions.

The T20 format of cricket can have large randomness because any slight change of bowler or field can hugely impact the game, especially in power play overs. So, we can also try to introduce one randomness parameter to judge how that can impact the game. We can also predict the man of the match, the highest scorer, and the highest wicket-taker of upcoming matches. The entire research may be used to forecast comparable results in many sports, such as baseball, tennis, football, and so on.

REFERENCES

[1] Lemmer, H. H. (2011), Team selection after a short cricket series, European Journal
of Sport Science, DOI:10.1080/17461391.2011.587895.

[2] Lemmer, H. H. (2002), The combined bowling rate as a measure of bowling performance in cricket, South African Journal for Research in Sport, Physical Education and Recreation, 24, 37-44.

[3] Lemmer, H. H. (2004), A measure for the batting performance of cricket players, South African Journal for Research in Sport, Physical Education and Recreation, 26, 55-64.

[4] Somaskandhan, Pranavan, et al. "Identifying the optimal set of attributes that impose high impact on the end results of a cricket match using machine learning." 2017 IEEE International Conference on Industrial and Information Systems (ICIIS). IEEE, 2017.

[5] Lewis, A. J. (2005), Towards fairer measures of player performance in one-day cricket, Journal of the Operational Research Society, 56(7), 804-815.

[6] Farhana Siddiqui, Hasan Phudinawala, Chetan Davale, Soham Pawar," Innovative Idea for Playerelection usingSupport Vector Machine (SVM)" International Journal of Computer Sciences and Engineering, Vol. 7, Issue-4, pp.841-843, 2019

[7] Singla, Saurav, and Swapna Samir Shukla. "Integer optimisation for dream 11 cricket team selection." International Journal of Computer Sciences and Engineering 8.11 (2020): 1-6.

[8] Shah, S., Hazarika, P. J., & Hazarika, J. (2017). A study on performance of cricket players using factor analysis approach. International Journal of Advanced Research in Computer Science, 8(3), 656-660.

[9] Hossain, M. J., Kashem, M. A., Islam, M. S., & Marium, E. (2018, September). Bangladesh cricket squad prediction using statistical data and genetic algorithm. In 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT) (pp. 178-181). IEEE.

[10] Sharp, Gary & Brettenny, Warren & Gonsalves, John & Lourens, Michelle & Stretch, R. Integer optimisation for the selection of a Twenty20 cricket team. Journal of the Operational Research Society. 62. 10.1057/jors.2010.122, 2011.

[11] Kamble, A., Rao, R., Kale, A., and Samant, S. (2011), Selection of cricket players using analytical hierarchy process, International Journal of Sports Science and Engineering, 5, 207-212.

[12] Bretteny, W. (2010), Integer Optimization for the Selection of a Fantasy League Cricket Team, South Africa: M.Sc. Dissertation, Nelson Mandela Metropolitan University.

[13] Lemmer, H. H. (2011), Performance measure for wicket keepers in cricket, South African Journal for Research in Sport, Physical Education and Recreation, 33, 89-102.

[14] Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2011). A MCDM approach for evaluating bowlers performance in IPL. Journal of emerging trends in Computing and Information Sciences, 2(11).

[15] Chaudhary, R., Bhardwaj, S., & Lakra, S. (2019, February). A DEA model for selection of Indian cricket team players. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 224-227). IEEE.

[16] Prakash, C. D. (2016). A new team selection methodology using machine learning and memetic genetic algorithm for ipl-9. Int. Jl. of Electronics, Electrical and Computational System IJEECS ISSN.

[17] Prakash, C. D., Patvardhan, C., & Singh, S. (2016). A new machine learning based deep performance index for ranking IPL T20 cricketers. International Journal of Computer Applications, 137(10), 42-49.