



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Optimizing Cricket Team Selection: Machine Learning Strategies for National and Franchise Leagues Building Championship Teams with AI

Hardeek Khadilkar, Rohit Ghate, Piyanshu Zade, Purva Bomanwar, Deepali Baghel

B. Tech Scholar, Department of Computer Science & Engineering, Ballarpur Institute of Technology, Ballarpur, India

**ABSTRACT:** Selecting cricket performers is a crucial job for both international squads and trade teams. Normally, recruiters depend on their expertise and know-how to gauge a player's mental readiness, batting, and bowling performance. Nevertheless, with the evolution of machine learning algorithms, there is an opening to automate and enhance the selection procedure. This analysis proposes a Machine Learning (ML)-established approach for cricket player selection. The approach exploits physical fitness figures, batting, and bowling data, and other pertinent metrics to craft exhaustive player profiles. Three ML algorithms - linear regression, bolster vector regression, and accidental forests - are utilized to spot the most encouraging performers. Figures on an ample quantity of cricket players and their appearances in international and trade leagues are assembled from ESPNCRICINFO and CRICBUZZ, the two notable cricket websites. The proposed strategy is then coached and tested on this data, and its correctness and performance are assessed. Based on the performance gradings procured from these models, two teams are selected for the international team and one squad for each trade league group. The outcomes prove that the proposed strategy considerably enhances the selection method and identifies performers with high potency. Besides, the bolster vector regression algorithm surpasses other ML models in prophecy and player selection.

**KEY WORDS:** cricket player selection, machine learning, franchise league, national team selection.

## I. INTRODUCTION

In the contemporary sports background, the costs in the sports sector reach into the billions of dollars, with a broad assortment of professionals such as instructors, consultants, physiotherapists, nutritionists, and coaches encouraging players and teams on an international proportion. Among the most prominent sports globally is cricket, a bat and ball sport engaged in 106 constituent nations of the International Cricket Council (ICC), which has branched into a multi-billion-dollar industry. The sport is performed in three core formats: Twenty-Twenty Cricket (T20), One Day Universal (ODI) cricket, and Experiment cricket. With the advancement of both the ODI and T20 formats, cricket has promptly garnered universal recognition as a team sport, bringing about the formation of trade T20 leagues in numerous nations worldwide.

Efficient teams substantially depend on data examination to achieve deeper insights into player achievement and the game itself. This method is predominantly beneficial to stakeholders within the sport, encompassing players, technical experts, coaches, and growing talents. With the speedy development of cricket, leaders consistently look for progressive resolutions to improve team performance and procure a competitive edge. Maximizing player performance while minimizing the danger of injury is a chief apprehension of player performance supervision, and cricket data scrutiny plays a vital function in this procedure.

Nevertheless, traditional statistical methods may not effectively meet the needs of sports data scrutiny owing to varied constraints. Regular statistical methods are typically constructed on beliefs about the data, and classic cricket data and scrutiny may not habitually adhere to these beliefs. Consequently, the selection of data scrutiny methodologies is of preeminent importance. Cricket-linked undertakings are intricate and commonly impacted by multiple human variables, rendering it puzzling to employ conventional metrics. Additionally, the modern game comprises numerous interconnected variables that traditional metrics may clash to include.

Team selection is a critical aspect influencing cricket performance, generally contingent on coaches and selectors' expertise, which may be vulnerable to prejudices and faults. Machine Learning (ML) procedures offer a data-driven alternative, examining player data to recognize trends and predict outcomes. ML models trained on varied data sources like player information and prowess data can appraise key factors affecting performance.

Sports analytics benefit from ML algorithms, with advancements in Internet of Things (IoT) technology facilitating swift and more exact data collection. ML strategies have become more prevalent due to increased computation reserves and data availability, intensifying data collection, extraction, and scrutiny.

This analysis aspires to shape well-balanced cricket teams by selecting performers grounded on their capabilities. An explicit strategy was gestated considering varied parameters to scrutinize each player's appropriateness for positions like captain, wicketkeeper, batsman, and bowler. Data from notable cricket websites was employed to cruciate players' physical fitness and calculate fitness scores.

The research contribution includes a revised view of the scarcity of studies regarding squad predictions for national and franchise teams in cricket. The developing ML-based approach overcome this gap by considering many performance parameters, enhancing predictive accuracy. Detailed methodology for player selection and practical application of ML algorithms demonstrates the research's value in improving cricket squad predictions across diverse contexts. Finally, Performance values are used to select the best players for both national teams and franchise leagues. On the whole, the major contributions of this research are:

- Conduct a comprehensive literature review on predicting player performance in cricket The review shows limited works predict squads for national teams, and none include franchise league squad prediction.
- Address identified the research gap by developing an ML-based approach for predicting squads for both national teams and franchise leagues in cricket.
- Overcome limitations of previous studies by considering a substantial number of parameters for measuring player performance, enhancing the predictive accuracy of the model
- Develop a detailed methodology for selecting players for specific positions (captain, wicketkeeper, batsman, and bowler) based on individual strengths and capabilities
- Collect data from reputable cricket websites, ESPNCRICINFO, 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
- Utilize 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 method that considers multiple performance parameters and leveraging ML algorithms to improve squad predictions for diverse cricket contexts.

## II. LITERATURE SURVEY

Player selections are vital in sports, particularly in cricket, where selecting eleven players from squads of 15 to 20 is challenging for team officials, coaches, and captains. Factors such as recent performance, opposition, venue, and series type are typically considered when making those decisions, making player analysis complex. Researchers have increasingly turned to machine learning (ML) techniques for forecasting player performance across various aspects of sports. ML algorithms have been utilized in self-awareness, medical applications, performance assessment, outcome predictions, injury estimation, tracing fitness, analysis of smart sports, content online generation, team selection, and sports fantasy.

In cricket, ML algorithms have been extensively employed to predict match outcomes and assess players. Several works focus on player selection using methodologies. For instance, one study aimed to establish psychological profiles of teenage cricket players and compared these profiles in terms of cognitive assets and adverse mental states. Players with high mental strength were found to have cognitive assets and fewer adverse mental situations compared to those with medium mental strength.

Another study introduced approaches to assess players' performance, considering the characteristics of both bowlers and batsmen. Clustering techniques were used to forecast effective batsmen or bowlers for specific matches in the Indian Supreme League (ISL). Additionally, researchers proposed a methodology for evaluating a player's worth for the ISL auction, considering factors such as past bid pricing, expertise, and metrics performance like bowling rate and wicket-taking abilities. Artificial neural networks were employed in another study to predict bowlers' IPL performances based on past ODI and T20I game metrics. Players are ranked according to performance, facilitating the selection of ideal players for competitions and matches. Furthermore, a social analyzing network was used to evaluate bowler's and batter's performance, providing insights into team dynamics and player contributions.



### III. EXISTING SYSTEM

Predicting the players performances is nothing but selecting the top players for every match in sports. In cricket, precisely 11 players are chosen at the start and remain fixed unless an injury occurs. Performance needs to predict whether players are excellent contenders based on records and considerations. The selection squad considered a balance of batters, bowlers, and all-rounders. The team included a wicketkeeper with a number of wickets and impressive batting. The algorithm might utilize assorted methods to boost optimization. There are two hybrid types. One direction employs the mechanism to pick one optimization method and alternates between iterative optimization. Other methods enhance techniques by using the primary formula one algorithm. This method combined updating formulas and choosing formulas for the hybrid system.

### IV. PROPOSED SYSTEM

In this project, teams were chosen for cricket matches using ML algorithms. A balanced team assembles by selecting batsmen, bowlers, batters, bowling all-rounders, and wicketkeepers within regulations. Features pass with feature optimization algorithms, improving accuracy. Methods are drawn from limited input variables. Logistic Regression, Naive Bays, K Nearest Neighbors, Support Vector Mechanics, Gradient Boosting, Binary Tree, Random Forest, XG Boost, and Cat Boost are nine classifiers from techniques as input. To enhance optimization efficiency, a solo method combined with CS-PSO optimization. Forecasting cricket team used in this study.

Objective: The aim of the system is to analyze player data, historical trends, metrics recommendations selectors for decisions.

#### Collecting Data:

- Gathered historical data on players's performance across various formats of the game (Test, ODI, T20).
- Include batting average, bowler averages, strike rate, fielding statistics, and other irrelevant metrics.
- Collecting data on player fit, injury history, and other performance-unrelated factors.

#### Data Preprocessing:

- Cleanse the data by handling missing valuing, outstanding, and inconsistencies.
- Normalize or scour numerical features to ensure they are on the similar.
- Encoding categorical variables like player roles (batswoman, bowler, all-under), playing positions, and team.

#### Feature Selection:

- Used techniques like correlation analysis, feature importance, or domain dumbness to select the more irrelevant features for squad selections.
- Features might conclude recent performance, fatness levels, playing conditions (e.g., spin-friendly pitches), and opponent strengths.

#### Model Building:

- Choose diversity machine learning model for squad selections. Ensemble methods like Random Forest or Gradient Boosting can't work well for this task.
- Split the data into training and testing sets to evaluate model performance.
- Training the model using the training data, optimizing hyper parameters using technique like cross-validation.

#### Model Evaluation:

- Evaluate the model performance using metrics such as accuracy, precision, recall, and F1-score.
- Use technics like confusion matrices and ROC curves to assess the model's ability to correct predicting squad selections.

#### Predict Squad Selection:

- Once the model is training and evaluated, use it to predict the best squad composition for upcoming matches or series.
- Consider constraints such as squad side, balance between batting and bowling strengths, and player roles within the team.

#### **Fine-Tuning and Validation:**

- Continuously update and fine-tune the model based on newdata and feedback from team selectors and coaches.
- Validate the model's predictions against actual squad selections to assess it's real-world effectiveness.

#### **Deployment:**

- Implementation the model into a user-friendly application or platform where selectors can input player data and receive squad recommendations.
- Ensure the deployment environment is secure, scalable, and accessibly to authorized users only.

#### **Monitoring and Maintenance:**

- Monitoring the model performance overtime, retriggering it periodically with updated data to maintain accuracy and relevant.
- Address any issues or biases that sin may arise during model deployment and usage.

### **V. ADVANTAGES OF PROPOSED SYSTEM**

**Data-Driven Decisions:** Machine learning algorithms can analyze extensive datasets, including player statistics, match performance, fitness levels, and other relevant metrics. This approach enables more objective and informed decisions compared to traditional subjective methods.

**Predictive Analytics:** Machine learning models can predict player performance based on historical data. This analysis helps selectors identify players likely to excel in upcoming matches or tournaments, leading to more competitive and effective squad selection.

**Efficiency:** Automated machine learning systems streamline squad selection by quickly processing and analyzing data, reducing the time and effort required by selectors. This efficiency allows selectors to focus more on strategy and team dynamics rather than manual analysis.

**Fairness and Transparency:** Machine learning algorithms promote fairness and transparency in squad selection by objectively considering various factors such as performance metrics, form, and fitness levels. This reduces biases and favoritism inherent in traditional selection processes.

**Optimized Performance:** Data-driven insights ensure the selection of the most suitable players, leading to better overall team performance. A well-balanced and optimized squad improves results in matches and tournaments.

**Adaptability:** Machine learning models evolve over time with more data, ensuring the squad selection process remains up-to-date and reflective of current player capabilities and sport trends.

**Cost-Effectiveness:** While requiring initial investment in technology and expertise, using machine learning for squad selection can be cost-effective. It reduces human errors, enhances selection quality, and optimizes team performance, resulting in better returns for cricket organizations.

### **VI. APPLICATIONS**

Machine learning can significantly enhance the process of cricket squad selection by providing data-driven insights and predictive analysis. Here are some key applications and benefits of using machine learning in cricket squad selection:

#### **Player Performance Prediction:**

- Machine learning algorithms can analyze historical performance data of players, including batting averages, bowling strike rates, fielding statistics, etc., to predict their future performance.
- This predictive analysis can help selectors identify players who are likely to perform well in upcoming matches or tournaments.

#### **Injury Prediction and Management:**

- By analyzing players' injury history, workload data, and other relevant factors, machine learning models can predict the likelihood of a player getting injured.

- This information can be used to manage players' workloads, schedule rest periods, and make informed decisions about squad selection to minimize the risk of injuries.

#### Opposition Analysis:

- Machine learning algorithms can analyze data related to opponents' strengths, weaknesses, strategies, and player performance trends.
- This analysis can help selectors formulate game plans, choose the right combination of players, and make tactical decisions during matches based on real-time data.

#### Formulating Optimal Squad Combinations:

- Machine learning can optimize squad selection by considering various factors such as player roles (batsman, bowler, all-rounder), playing conditions (pitch type, weather conditions), and opposition strengths.
- By using advanced algorithms, selectors can identify the most effective squad combinations for different match scenarios, maximizing the team's chances of success.

#### Talent Identification and Development:

- Machine learning can assist in identifying talented players at grassroots levels by analyzing performance metrics, playing styles, and potential for growth.
- This helps in nurturing young talent, providing them with opportunities, and shaping the future of cricket by selecting promising players for development programs and national squads.

## VII. CONCLUSION

For institutions aiming to enhance performance, selecting the right personnel is a crucial challenge, particularly in team sports like cricket where team composition significantly affects outcomes. This study focuses on team selection in one-day international cricket, aiming to classify players into specific roles using machine learning techniques and feature optimization.

A dataset of 414 Indian ODI players was analyzed using nine machine learning techniques to categorize players into five roles: batters, batting all-rounders, bowlers, bowling all-rounders, and wicketkeepers. Initially, Support Vector Machine (SVM) achieved prediction accuracies of 93.54% for batters, 87.29% for batting all-rounders, 87.09% for bowlers, 85.70% for bowling all-rounders, and 84.21% for wicketkeepers.

Further optimization, combining SVM with CS-PSO (Cat Swarm Particle Swarm Optimization), improved prediction accuracies significantly. For instance, the hybrid technique achieved accuracies of 97.14% for batters, 97.04% for bowlers, 97.29% for bowling all-rounders, and 97.28% for batting all-rounders. Wicketkeeper selection accuracy reached 92.63%.

The study's findings are valuable for cricket officials, coaches, captains, and player selection committees, providing insights into identifying the best players for specific roles. Future research could incorporate additional performance data, including opponents' details, to further refine the classification model and potentially extend its application to predicting outcomes in Twenty20 matches.

## REFERENCES

- [1] Sanderson, J., & Brown, K. (2020). COVID-19 and youth sports: Psychological, developmental, and economic impacts. *International Journal of Sport Communication*, 13(3), 313-323. doi: 10.1123/IJSC.2020-0236.
- [2] Deep Prakash, C., & Verma, S. (2022). A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket. *Decision Analytics Journal*, 2, 100025. doi: 10.1016/J.DAJOUR.2022.100025.
- [3] Kampakis, S., & Thomas, W. (2015). Using Machine Learning to Predict the Outcome of English County Twenty over Cricket Matches. Retrieved from arXiv:1511.05837. doi: 10.48550/arXiv.1511.05837.
- [4] Bandulasiri, A., Brown, T., & Wickramasinghe, I. (2016). Factors affecting the result of matches in the one day format of cricket. *Operations Research and Decisions*, 26(4), 21-32.
- [5] Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213-222. doi: 10.1007/S41060-017-0093-7.
- [6] Sarlis, V., & Tjortjis, C. (2020). Sports analytics — Evaluation of basketball players and team performance. *Information Systems*, 93, 101562. doi: 10.1016/J.IS.2020.101562.

- [7] Xu, T., & Tang, L. (2021). Adoption of machine learning algorithm-based intelligent basketball training robot in athlete injury prevention. *Frontiers in Neurorobotics*, 14, 117. doi: 10.3389/FNBOT.2020.620378/BIBTEX.
- [8] Horvat, T., & Job, J. (2019). Importance of the training dataset length in basketball game outcome prediction by using naive classification machine learning methods. *Elektrotehniški vestnik*, 86(4), 197-202. Retrieved from <https://ev.fe.uni-lj.si/4-2019/Horvat.pdf>.
- [9] Yang, Y., Qin, T., & Lei, Y. H. (2016). Real-time eSports match result prediction. In *Conference on Neural Information Processing Systems*. Spain, Barcelona. Retrieved from <http://www.forbes.com/sites/leighsteinberg/2015/08/18/changingthe-game-the-rise-of-sports-analytics/>.
- [10] Gowda, M., Dhekne, A., Shen, S., Choudhury, R. R., Yang, X., Yang, L., et al. (2018). IoT platform for sports analytics. *GetMobile: Mobile Computing and Communications*, 21(4), 8-14. 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, Verstockt 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.





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details