



An efficient team prediction for one day international matches using a hybrid approach of CS-PSO and machine learning algorithms

Manoj Ishi ^{a,*}, Jayantrao Patil ^a, Vaishali Patil ^b

^a Department of Computer Engineering, R. C. Institute of Technology, Shirpur, India

^b RCPET's Institute of Management Research and Development, Shirpur, India

ARTICLE INFO

Keywords:

Team formation
Player evaluation
Feature optimization
Metaheuristic algorithm
Nature inspired algorithm

ABSTRACT

Player classification is vital in cricket since it assists the coach and skipper in determining individual players' roles in the squad and allocating tasks appropriately. The performance statistics help to classify players as batsmen, bowlers, batting all-rounder, bowling all-rounder, and wicketkeeper. This research aims to correctly identify cricket teams in the one-day international format by categorizing players into five groups. Based on their previous and current performance, the players are rated as excellent, very good, good, satisfactory, or poor. An enhanced model for the game of cricket is presented in this study, in which an eleven-member team picked using an unbiased technique. Players should be selected based on their performance, batting average, bowling average, opposing team strength and weakness, etc. Nature-inspired algorithms are used for feature optimization to improve the accuracy of machine learning prediction models. The blending of Cuckoo Search and Particle Swarm Optimization is performed called CS-PSO, which successfully integrates the capabilities from both approaches to create reliable and suitable solutions in accomplishing global optimization efficiently. Using a hybrid of CS-PSO feature optimization and Support Vector Machine, batters, bowlers, batting all-rounders, bowling all-rounders, and wicketkeepers were picked with an accuracy of 97.14%, 97.04%, 97.28%, 97.29%, and 92.63%, respectively.

1. Introduction

Cricket's continual growth needs innovation to remain ahead of the competition and attract new fans or followers. The One-Day International (ODI) format is a prominent example of this as it is possibly the most significant alteration in any team sport. Batting and bowling are the two considerable abilities in all forms of cricket. Each ball bowled in cricket creates massive data. Individual players' batting and bowling performances are evaluated and averaged to define a team's overall performance. Batting average and strike rate are often used to assess batsmen's performance in cricket, whereas bowling average, economy rate, and strike rate are typically used to analyse bowler's performance. However, the majority of the present criteria on the scorecard are ineffective in determining a player's natural skill. Batting average, for example, tells us the number of runs scored by a batsman on average before losing his wicket. The batting average defines a player's potential to score runs. Though, it can't tell how efficient a batsman is in scoring rapidly. Similarly, looking at the economy rate, one understands the pace at which a bowler loses runs but not his ability to take wickets. As a

result, many performance metrics have been developed to quantify cricketers' batting and bowling performances by integrating standard performance data. The bowling team's dot-ball bowling and wicket-taking skills and the batting team's boundary-hitting proficiency and 50-plus partnerships are critical for ODI success. Forming a team to play a specific rival team is an arduous process since many things must be considered, including the weaknesses and strengths between both sides [1].

Predicting the players' performance is nothing more than selecting the top players for every match in any sport. In cricket, precisely 11 players are chosen at the start of the play and remain fixed for the entire game unless an injury occurs. The individual's performance needs to be predicted with a choice as to whether the player is an exceptional contender for participation in the squad based on past records and other considerations. The decision for selection of the squad considered an enormous balance of batters, bowlers, and all-rounders. The team should have included a wicketkeeper with remarkable numbers behind wickets and impressive batting statistics. Although fielding seems to be a crucial part of a play, bat and ball skills are valued more than fielding [2,

* Corresponding author.

E-mail addresses: ishimanoj41@gmail.com (M. Ishi), jbpatil@hotmail.com (J. Patil), vaishali.imrd@gmail.com (V. Patil).

<https://doi.org/10.1016/j.array.2022.100144>

Received 29 December 2021; Received in revised form 14 February 2022; Accepted 1 April 2022

Available online 12 April 2022

2590-0056/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

3].

Consistency is critical in the selection phase. Focus is given to each player's quality to get an impressive list of players for the Indian cricket squad in ODIs. In most sports, team selection is a subjective problem based on widely-accepted perceptions of what constitutes a good team [4]. We addressed the challenge of constructing a 'good' squad from a group of players based on their historical performance data. Because the entire market of players is large, determining the best team gets increasingly challenging, and typical logical procedures may fail to build a great team within rule constraints [5].

"Cricket is a game of beautiful uncertainties," as the phrase goes, it is the game's extremely unexpected character that piques the curiosity of its fans. The game is organized to determine the winning team mainly by the most potent team competing on that particular day. There have been instances where lower-scoring teams won by eliminating their opponents for a lower score. Simultaneously one cannot rule out the possibility of the other team actively chasing a huge score. Because of the way the game is played, the game's outcome is overturned by losing wickets or a streak of magnificent scoring strokes. Such comments emphasize the erratic character of cricket [6]. The decision-maker choices and ranking of contenders are used to select the team based on priorities [7].

Machine Learning algorithms and associated Artificial Intelligence technologies are helpful in various fields such as prediction and decision making. Machine learning technology may be used by team management to evaluate the efficiency of opposition team members. Before the game use of machine learning allows both the players and the coaches to analyse the areas where they can improve. Analysing each player's whole previous history manually is nearly difficult. As a result, an intelligent system that predicts player performance based on previous performance might benefit team management and selectors. Because most previous studies employ a short time frame, our primary focus is on using players' previous performance over a more extended period for better accuracy [8]. The quantitative components give information where statistics will be comparable for two players that played against different opponents and performed similarly. Still, they leave out specific crucial details: The player who scored against a stronger opponent should be given a higher rating. Dismissals of batters with a better career record should be given a higher rating than dismissals of batters with lower career records [9]. This work demonstrates building a team forecasting model using the classification and prediction approaches. Cricketers may be classified into several categories based on their evolutionary performance. Despite the vast number of potential classes, a player's performance can place them into one of five primary evolutionary classes. The upcoming performance of players is predicted by considering their initial performance. Any player that is not fruitful will be labelled a poor performer and may be dropped from the squad [10].

Feature optimization is critical in machine learning because high-dimensional datasets include duplicated, noisy, and irrelevant characteristics. Feature optimization reduces data dimensionality and chooses only the most significant features to enhance classification performance and reduce computation costs. Metaheuristic algorithms are recognized as a viable approach for addressing feature optimization problems. Metaheuristic algorithms have become popular due to their stochastic and non-deterministic character. The phrase "nature-inspired algorithm" refers to a class of metaheuristic optimization algorithms evolved from natural phenomena. Swarm intelligence is a type of metaheuristic optimization algorithms based on natural agent behaviour. A swarm structure represents social intelligence consisting of many homogenous, self-organized, and fragmented agents disseminated throughout the ecosystem, such as schools of fish, ant colonies, and flocks of birds. In SI, achieving the best solution necessitates the communication of knowledge among the members of the swarm system. SI has commonly been employed in the key to large-search-space optimization problems. In this work, we studied and implemented Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), Cuckoo Search (CS), Grey Wolf Optimizer (GWO), Moth-Flame Optimization (MFO), Whale

Optimization Algorithm (WOA), Bat Algorithm (BAT), and Firefly Algorithm (FFA). All algorithm uses a fitness function to increase classification accuracy while significantly reducing the number of shortlisted parameters [11].

Nowadays, hybrid algorithms are given more attention for improving the prediction models' accuracy. The hybrid algorithm might utilize an assortment of two or more approaches to boost the algorithm's optimization capabilities. There are two types of hybrid optimization techniques. One direction is to employ a mechanism to pick one of the two optimization methods and then alternate between the two algorithms in the iterative optimization process. The other method enhanced the different techniques by using the primary formula of one algorithm. This method combined the two optimization algorithms' position updating formulae and chose alternative position updating formulas for optimization with a specific mechanism to apply to the hybrid system [12]. The current study aims to develop a hybrid algorithm that combines CS and PSO to achieve higher global optimization outcomes and solve proposed player classification issues. It is vital to pick the most acceptable cricket line-up to win matches.

This paper's significant contributions are as follows

- Five algorithms are proposed based on the features that reflect their strengths to calculate the rating of batters, bowlers, batting all-rounders, bowling all-rounders, and wicketkeepers.
- CS-PSO hybridization is used as feature optimization strategies to eliminate redundant, irrelevant, and noisy features.
- The suggested method's performance is compared to standard PSO and CS methods with six other algorithms.
- The hybrid approach of CS-PSO with machine learning models is used to find the right set of team combinations from the group of players obtained after evaluating players' strength for one-day international matches.

This model would benefit the players, team, and the coaches since they would know where and how their squad should focus on getting better outcomes in the match [2,3]. These models are also helpful for sports analysis shows broadcasted on TV during the gameplay. News channels are also having reserved slots for a sports show. In that sports show, they discuss the key points of players, players' statistics, and teams' statistics. This model of players selection is also helpful for them. This forecasting system is also useful for other real-word problems like weather forecasting, share market prediction, etc. After studying the variables of these prediction models, our model will require few changes to run effectively.

The rest of the paper is organised as follows: The literature review section contains references to previously published work. The next part will go through the data selection procedure including descriptive statistics for the collected data. The methods section discusses machine learning methodologies. The findings of this study are then presented in the next section. Finally, the discussion and conclusion section goes through the research summary.

2. Literature review

In cricket, several academics have focused on predicting the team. A few of them are discussed in the following section.

I. Wickramasinghe categorizes all-rounders in his study using the mean of their batting and bowling averages [2]. The author classified all-rounders into four types: genuine all-rounders, batter all-rounders, bowling all-rounders, and ordinary all-rounders. Naive Bayes (NB), K-Nearest Neighbors (KNN), and Random Forest (RF) was used to make the prediction. The k-fold cross-validation approach is used to test all of the analytical outcomes. As per the experiment results, RF has a significantly greater prediction accuracy.

V.S. Vetukuri et al. proposed a hybrid methodology that selects efficient players by combining the concepts of recurrent neural networks

(RNN) and genetic algorithms (GA) [3]. Individual players' historical statistics are adequately pre-processed, and an initial feature grid is generated with each individual for the mathematical function utilized in GA. This enhanced feature matrix is then sent into RNN, which computes a final score for each participant. This suggested approach generates a concurrent rank table that team selectors may use to make quick and accurate player selections for the coming match.

A. Khot et al. put forward the concept of co-players to identify rising stars and improve team selection [4]. A discriminative classifier trained for classification is a support vector machine (SVM). Plotting hyperplane divides a point into two groups based on its coordinates: (1) rising star (2) not a rising star. The RS score is calculated by identifying which coplayer characteristics are favorably and adversely linked. The RS score is used to compile the final list of rising stars for the batting domain with an accuracy of 60%, 70% for the bowling domain, and an all-rounder assessment with 40%.

F. Ahmed et al. present a multi-objective strategy based on the NSGA-II algorithm to discover team members and optimise overall batting and bowling strength [5]. They used a player's batting average and bowling average in international T-20 cricket to measure their batting and bowling performance. After issue formulation, they employ the elite non-dominated sorting genetic algorithm (NSGAI) to perform multi-objective genetic optimization across the team. They use the feasible solution created through knee region analysis to determine their fitness score on all such metrics to account for such issues.

A. Balasundaram et al. used K-means clustering, decision trees, support vector machine, and random forest for player classification [6]. WEKA is the data mining program utilized to execute operations upon that provided dataset. Cross-fold validation is used to build training and testing data from class-labelled data. The prediction accuracy of the classifier is 91.87% while using the decision tree, 93.46% for SVM, and 95.78% with random forest, showing that the constructed model was effective in anticipating the best player option for the team.

M. Bello et al. introduced a revised approach to team formation wherein two organizations form teams by selecting persons out of a shared pool of applicants [7]. This study proposes the Ant Colony Optimization (ACO) and the Max-Min Ant System to Team Formation (MMAS-TS) metaheuristics. Every structure occurs throughout the solutions discovered by the best pair of ants in the iteration. The pair of ants that have obtained the most effective solution from the beginning of execution receives a pheromone deposit. Each decision-maker assigns a ranking to the applicants based on his preferences for forming teams while retaining the overall quality.

C. Kapadiya et al. tried to predict how many runs a batter would score and how many wickets a bowler would take in a given game on a particular day [8]. Machine learning techniques such as decision trees, SVM, naïve Bayesian, and random forest are used for prediction. They presented a method that uses a weather dataset with cricket match information to forecast player performance. A novel weighted random forest classifier including hyperparameter tuning is employed in their model with an accuracy of 92.25%.

P. Chhabra et al. proposed modeling players into embeddings using a semi-supervised statistical technique for building a team selection [9]. The 'Quality Index of Player' grading system is developed in this article that considers both qualitative and quantitative aspects of evaluating players. CRICTRS is a semi-supervised team suggestion framework that requires player embeddings to advise a team focused on the opponent's strengths and weaknesses. This method is developed from collaborative filtering and the Bernoulli experiment that ranks players based on the quality of their runs and wickets.

H. Ahmad et al. uses supervised machine learning models to predict Star Cricketers through the batting and bowling domains [10]. Predictions are made using Bayesian rule functions and decision-tree-based frameworks with a cross-validation approach to validate the performance. The contribution of each feature to the prediction challenge was determined using state-of-the-art metrics such as information gain, gain

ratio, and chi-squared statistics. SVM achieved the most remarkable results in the batting domain with an accuracy of 86.59%. In the bowling dataset, NB and CART beat all other models with an accuracy of 88.9%.

P. Agrawal and T. Ganesh outline the Indian cricket team's player selecting procedure using integer optimization programming [13]. Batting statistics are based on two variables to determine a player's efficiency for a match: batters average and batters consistency. A bowler's performance is evaluated using his bowling average, consistency, and economy rate. Spearman's rank correlation coefficient is applied to match the typical technique of varied values. The decision variable of a unit function type has been utilized to pick the cricket team using integer programming. The player's consistency was a significant factor in the rating.

Z. Mahmood et al. presented the prediction of rising stars in basketball as a machine learning issue using Classification and Regression Trees, Support Vector Machines, Maximum Entropy Markov Model, Bayesian Network, and Naïve Bayes [14]. Their goal is to create a function that can predict whether a class is a rising star or not based on a collection of features. They use the h-index to compensate for the lack of an average efficiency score for each co-player. To calculate the efficiency of co-players, they employ the Hollinger linear formula. A 10-fold cross-validation approach is employed to train and validate them using all three datasets to measure the strength of the classifiers.

I. Kumarasiria and S. Perera used a genetic algorithm to identify the best cricket team [15]. The players' fitness is evaluated using batting average, batting strike rate, bowling average, a total of wickets in each match, win percentage, and experience. The fitness values of every chromosome are then calculated to indicate the total fitness of each player. The fitness values for each team are obtained using the mutation approach. The stopping condition for this task is to repeat the process until the gap between the best and poorest player's fitness scores is less.

M. Ishi and J. Patil extensively survey team formation and winner prediction methods in cricket [24,25]. They studied every parameter and method used for the evaluation of players. Based on the study, parameters and procedures are found for team formation. Every parameter has its impact, and based on that; the parameter weightage needs to be finalized.

3. Methodology

This work uses machine learning algorithms to pick teams for one-day international cricket games. Within the limits of the rules, batters, bowlers, batting all-rounder, bowling all-rounder, and wicketkeeper are chosen to form a balanced squad. With the help of the feature optimization algorithm, the player's features make their way through to increase and improve prediction accuracy dynamically. The number of input variables is reduced using methods inspired by Nature. Nine classifiers use the selected features from Nature Inspired methods as input: Logistic Regression, Naive Bayes, K Nearest Neighbors, Support Vector Machine, Gradient Boosting Algorithm, Decision Tree, Random Forest, XGBoost, and CatBoost. The CS-PSO hybrid optimization approach is used with a standalone algorithm to improve the feature optimization performance for selecting relevant features. Fig. 1 depicts

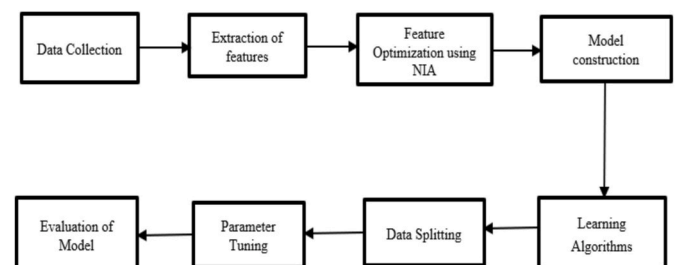


Fig. 1. Steps for team prediction.

the machine learning framework processes for predicting the team of a cricket match for this research.

3.1. Data collection and interpretation

This investigation's data is gathered from a publicly available source. Web scraping is a method for extracting data from websites. The second alternative is to copy and paste the data manually. However, this is not technically feasible due to the time it takes. Instead of copying data manually, online scraping automates the process allowing data to be accessible fast and without wasting time. A dataset including 101 batters, 101 bowlers, 101 batting all-rounders, 101 bowling all-rounders, and ten wicketkeepers with player-related performance variables is developed to form the Indian team. Each player's information is contained in the linked dataset. For 1989–2021 data was gathered from ESPN Cricinfo for all four domains, namely batsmen, bowlers, all-rounders, and wicketkeepers [26]. The raw data is pre-processed before designing the prediction model. Many essential features for cricket prediction will be extracted as subsets from this primary data using data pre-processing techniques. Standard prediction methods are utilized to generate a model for these selected feature sets.

3.2. Features extraction

Domain expertise is necessary to extract the required features. Batting statistics are derived from the number of games played, the total of not out innings, runs scored, maximum score, batting average, strike rate, number of half-centuries and centuries, and the number of fours and sixes hit by batsmen. For bowlers, the number of overs bowled, wickets taken, maiden overs, bowling average, strike rate, economy rate, and 4/5 wicket haul taken considered. Both batsman and bowler traits are used to evaluate batting and bowling all-rounders. The difference between a player's batting and bowling averages also influences his success as a batting or bowling all-rounder. Batting-related factors, the number of catches taken, and stumpings completed are used to assess wicketkeeper quality. The following characteristics are studied year wise, opponent wise, venue wise, and inning wise to determine a player's strength. Binarization is used to transform these data. Thresholds for each of the characteristics are defined. The threshold values are just the sum of all players' values for a given attribute. Table 1 lists the

metrics for batters and bowlers along with descriptions, and Table 2 lists the characteristics of a batting all-rounder, bowling all-rounder, and wicketkeeper.

3.3. Feature optimization

Optimization is the process of adjusting a framework to ensure some aspects work more effectively or offering alternative outcomes under given restrictions as efficiently as possible by enhancing required parameters while eliminating unpleasant parameters. It is accomplished through the use of metaheuristic algorithms. Physical processes, animal behaviours, and evolutionary ideas are familiar sources of inspiration for metaheuristic algorithms. Researchers can easily understand metaheuristics and apply them to their problems due to their simplicity. The assessment of a suitable combination of feature optimization methodologies and machine learning algorithms is undertaken to obtain optimal accuracy. The search begins with a random beginning population in metaheuristic algorithms, which is then improved over time through iterations. Several optimal solutions teach us about solution space, resulting in spontaneous leaps towards the most plausible solution. Various applicant system works together to avoid finding the best solution locally. For this research, we employed the feature optimization approaches using SI algorithms which generally reproduce the social behaviour of swarms, herds, flocks, or schools of insects in nature [7,10,11,16–23]. For efficient feature optimization, we used the hybrid approach of CS-PSO algorithm. Performance is compared with standard CS, PSO method with other six algorithms: Ant Colony Optimization (ACO) [7], Grey Wolf Optimizer (GWO) [19], Whale Optimization Algorithm (WOA) [20], Bat-Inspired Algorithm (BBA) [21], Firefly Algorithm (FFA) [22], and Moth-Flame Optimization (MFO) [23].

3.3.1. Particle swarm optimization (PSO)

The PSO method is a stochastic optimization approach based on swarms that mimic animal social behaviour like insects, cattle, fish, and birds. These swarms follow a collaborative food-finding strategy, with each swarm member altering the search pattern in response to its own and other member's learning experiences. Particles in PSO may adjust their locations and velocities in response to changes in the environment to meet proximity and quality criterion. Furthermore, the swarm does not limit his mobility with PSO but instead seeks the optimal solution in

Table 1
Batsmen and bowlers' features.

Batsmen	Description	Bowlers	Description
run_score	number of runs scored	no_of_maiden_overs	no of maiden overs bowled
no_of_notouts	not outs in inning	no_of_runs_given	runs conceded by a bowler
batting_avg	average of batsmen	bowling_avg	average of bowlers
batting_strikerate	strike rate of batsmen	bowling_strikerate	strike rate of bowler
no_of_100's_50's	a weighted average of 100's and 50's	eco_rate	economy rate
no_of_0's	innings in which batsmen out for zero	no_wickets_taken	no of wickets taken
no_of_4's_6's	a weighted average of 4's and 6's hit	no_of_4_5_wicket_haul	weighted average of 4 and 5 wicket hauls
highest_score	the highest individual score for batsmen	no_of_max_wickets	maximum wickets in a single match
batting_score	a weighted average of batsmen score using all features reflecting batter's strength	bowling_score	a weighted average of bowling score using all features
best_batting_position_score	a weighted average of batsmen score using position wise strength	inningwise_bowling_score	a weighted average of bowler score using inning wise bowler's strength
inningwise_batting_score	a weighted average of batsmen score using inning wise strength	home_away_bowling_score	weighted average of bowler score using venue wise bowler's strength
home_away_batting_score	a weighted average of batsmen score using venue wise strength	opponent_bowling_score	a weighted average of bowler score using opponent wise strength
opponent_batting_score	a weighted average of batsmen score using opponent wise strength	yearwise_bowling_score	a weighted average of bowler score using year wise strength
yearwise_batting_score	a weighted average of batsmen score using year wise strength	wickets_taken_performance	a weighted average of bowler score depends upon wickets taken quality.
captaincy_point	batsmen is captain or not	captaincy_point	bowler is captain or not
overall_batting_score	a weighted average of all features to reflect batsmen strength in every aspect	overall_bowling_score	a weighted average of all features to reflect bowling strength in every aspect
player_rating	depending on performance players are assigned ratings as excellent, very good, good, satisfactory, poor	player_rating	depending on performance players are assigned ratings as excellent, very good, good, satisfactory, poor

Table 2

Batting all-rounder, bowling all-rounder and wicketkeeper features.

Batting all-rounder	Description	Bowling all-rounder	Description	Wicketkeeper	Description
All batting features	All batsmen feature described in Table 1	All batting features	All batsmen feature described in Table 1	All batting features	All batsmen feature described in Table 1
All bowling features	All bowlers feature described in Table 2	All bowling features	All bowlers feature described in Table 2	no_of_catches_wk	number of catches taken behind the wicket
bat_bowl_avg_diff	batting and bowling average difference for a player	bat_bowl_avg_diff	batting and bowling average difference for a player	no_of_stumpings	number of stumpings
overall_batting_all_rounder_score	weighted overall batting all-rounder score	overall_bowling_all_rounder_score	weighted overall bowling all-rounder score	wicketkeeping_ability	weighted overall wicketkeeper ability using a number of catches taken and stumpings done
player_rating_batting_allrounder	depending on performance players are assigned ratings as excellent, very good, good, satisfactory, poor	player_rating_bowling_allrounder	depending on performance players are assigned ratings as excellent, very good, good, satisfactory, poor	overall_wicketkeeper_score	a weighted score of batting and wicketkeeper features
				player_rating_wicketkeeper	depending on performance players are assigned ratings as excellent, very good, good, satisfactory, poor

the given solution space. In PSO, each individual is referenced as a particle and characterised as a reasonable alternative to optimization issues in the solution space. It can memorize the swarm's optimum places as well as its velocity. Each generation combines the particle's data to modify the movement of each dimension, which is then used to calculate the particle's new position. The particle has faith in its existing state of motion and moves inertia according to its velocity due to its own experiences. The "social" factor differentiates between the particle's present state and the swarm's global (or local) ideal position. It uses the social learning factor to imitate the movement of positive particles. It is believed that inertia weight is used in PSO to equalize global and local search with a higher inertia weight favouring global search and a lower inertia weight favouring local search [16].

When executing the method, it is imperative to accurately choose the particle population size N , the maximum number of repetitions M , inertia weight w , and other parameters. The following two equations are used to update the location and velocity of all particles.

$$v_i^{k+1} = wv_i^k + c1r1(pb_i^k - x_i^k) + c2r2(gb^k - x_i^k) \quad (1)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2)$$

In the above equation, x_i^k is the position of a particle. v_i^k indicates velocity, w is inertia weight, learning factors $c1$ and $c2$, and $r1$ and $r2$ are random numbers having values between 0 and 1. pb_i^k is the personal best of particle and gb^k represents the global best of the swarm.

3.3.2. Cuckoo search (CS)

This methodology is built on the brood parasitism of some cuckoo species and the random movements of Levy flights. Some cuckoo species deposit their eggs in host bird nests and may destroy other eggs to enhance the likelihood of their hatching. If the host birds do not locate and kill the eggs, they will hatch into a full-grown cuckoo. Cuckoo migration and environmental factors should ideally cause them to converge and choose the optimal location for reproduction and breeding. If the host birds find the eggs aren't theirs, they'll either discard them or abandon their nests and start again. Parasitic cuckoos

typically choose nests where the host bird has just laid eggs. Cuckoo eggs sprout somewhat sooner than host eggs. When the first cuckoo offspring hatches, his instinct is to evict the host eggs from the nest by forcing them out. As a result of this behaviour the cuckoo chick receives a larger part of the food given by its host bird. Similarly, a cuckoo chick may replicate a host chick's call to obtain more feeding opportunities. The host bird's young die of starvation, leaving just the cuckoo baby in the nest. The cuckoo's breeding behaviour may be used for various optimization issues [17]. A cuckoo egg indicates a fresh occurring solution, and also each egg in a nest signifies a solution. The objective is to eliminate less-than-ideal solutions with new and maybe improved ones in the nests. Selection of the fittest and adaptability to the environment are two essential qualities. These may be translated practically into two key aspects of contemporary metaheuristics: intensification and diversity. Diversification guarantees that the algorithm can successfully explore the solution space, whereas intensification attempts to explore the best existing solutions and select the best candidates or solutions [18]. The next solution is found with the levy flight strategy expressed with equation (3).

$$x_{i+1} = x_i^t + \alpha \oplus Levy(\beta) \quad (3)$$

Levy flight is a global randomized walking method whose future state x_i^{t+1} is determined only by the present state x_i^t and the transition probability levy (β). The random step length of a Levy flight is calculated from the Levy distribution with infinite variance and mean.

$$Levy \sim u = t^{-1-\beta} (0 < \beta < 2) \quad (4)$$

3.4. Hybrid Particle Swarm Optimization and Cuckoo Search algorithm

The well-known truth is that any population-based algorithm must explore and exploit to perform effectively. It is simple for the PSO algorithm to slip into the local optimum solution. The cuckoo algorithm's random walk technique, on the other hand, can boost the solution's variety in the search space. As a result, this research suggests incorporating the random walk approach into the PSO algorithm to create a new hybrid optimization method. The purpose of this hybridization approach

is to select relevant features from the data of batters, bowlers, all-rounders, and wicketkeepers to select optimal teams for ODI matches. The random numbers $r1$ and $r2$ are replaced with the Levy flight strategy to improve the searching for global optimum solutions. The stages of the CS-PSO hybrid optimization algorithm are as below and shown in Fig. 2:

Algorithm

1. Initialize the parameters for cuckoo search.
2. Divide the population into several groups.
3. Apply cuckoo search algorithm to find the local optimum solution for feature optimization using the fitness value of each individual.
4. Use local optimum solution obtained from cuckoo search as input population to PSO algorithm.
5. Initialize the particle from the input population for PSO.
6. For efficient searching of optimum solutions, formulas 1 and 3 are integrated into the PSO algorithm.

$$v_i^{k+1} = wv_i^k + (c1 \oplus \text{Levy}(\beta)) (pb_i^k - x_i^k) + (c2 \oplus \text{Levy}(\beta)) (gb^k - x_i^k) \quad (5)$$

It helps Particles in the PSO algorithm to obtain global solutions efficiently.

Because the local walk of particles is improved with levy flight.

7. The best solution obtained from the CS-PSO hybrid approach is an optimal output of the optimization algorithm.

The feature optimization algorithms take features from Table 1 and Table 2 as input to remove irrelevant features. The features that have more impact on defining players' strength are found using feature optimization algorithms. The optimized features obtained after feature optimization algorithms are provided as input to machine learning classifiers. After this, machine learning algorithms classify players into one of five class.

3.5. Learning algorithms/model selection

This study employs Logistic Regression, Nave Bayes, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting method, XGBoost, and CatBoost algorithms. Because of the classification problem, these algorithms are chosen. The winner

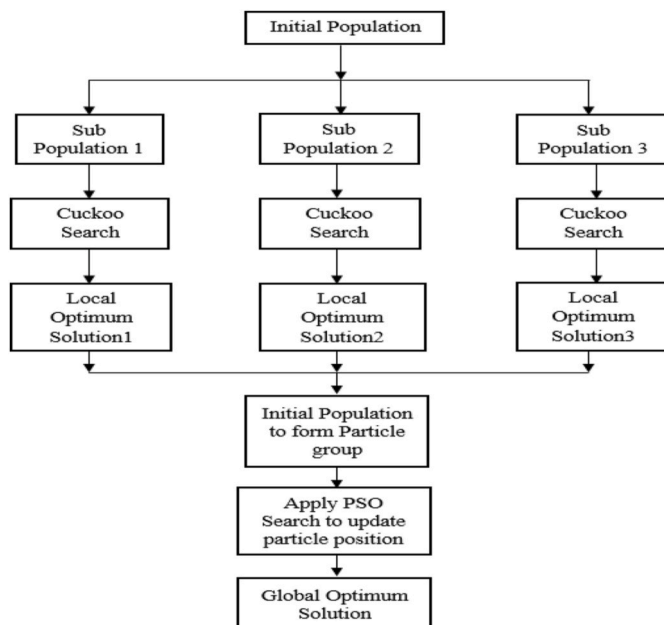


Fig. 2. PSO-CS Hybrid approach for feature optimization.

prediction classification problem is expressed as $y = f(x)$, at which x seems to be a single or set of independent variables and y is the dependent variable [2,4-6,8].

3.6. Training and testing

The data for estimating the team for cricket matches are split into 70-30% training and testing sizes in this investigation. This is because while attempting to anticipate match outcomes based on previous match records, it is important to preserve the chronological ordering of the sports results predictions data. The cross-validation method is inapplicable to sports prediction since it involves data shuffling which might change the order of events. As a result, manually splitting training and testing data to retain the chronological order of findings is preferable.

3.7. Parameter tuning and model evaluation

Models can contain a lot of parameters, and finding the appropriate mix of parameters can be a difficult task. In a machine learning method, parameter tuning is used to determine the best parameters for the model's training. Hyperparameters control the algorithm's learning process and improve model accuracy. The GridSearchCV method is used to tune the parameters for this work. The Grid search constructs and evaluates the model using any parameter combination given in the dictionary. After obtaining accuracy/loss for each variety of hyperparameters using the Grid Search technique, the parameters with the best performance are chosen. The prediction is made as a multi-class classification issue with the following classes: Excellent, Very Good, Good, Satisfactory, and Poor. The model evaluation is carried out with accuracy, precision, recall, and F1 score value [4-6].

4. Model formulation and feature construction

Five algorithms are proposed for evaluating batters, bowlers, batting all-rounder, bowling all-rounder, and wicketkeeper. The features that reflect players' batting strength are used to calculate batsmen's performance for teams using the batsmen strength algorithm. Similarly, the features showing the impact of bowlers, batting all-rounder, bowling all-rounder, and wicketkeeper are used to design algorithms for players according to their strengths. The player's performance is more critical concerned to form a balanced squad. The team's overall strength depends on players' strength and its impact on the team's performance. A team with quality players increases the chances of winning the matches. As players' statistics vary with the year, venue, opponent, etc., we studied every feature for these parameters. We analyse and evaluate each player's strength with just not only with their overall statistics but also with different conditions to select the quality player. The player's performance varies with the inning number. Some players are showing good show during the first inning, while some absorb the pressure of the second inning very well to deliver their strength. Some players are good at home conditions but cannot deliver away. Likewise, opponent wise performance is classified into two types: strong and weak. The teams with good players and high ICC ranking are considered strong opponents. While teams do not have a good ranking in ICC and players are not much consistent in performance, consider a weak opponent. The specific

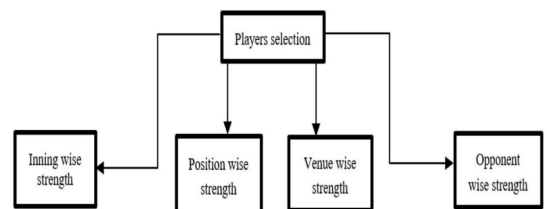


Fig. 3. Players classification conditionwise.

weights are assigned to each feature. The player's evaluation with different conditions is shown in Fig. 3.

We propose the first algorithm to evaluate the strength of batters. Firstly, in this algorithm, the strength of each player as a batsman (' α _batsmen_score') are calculated using run scored, number of not out innings, batting average, strike rate, milestone reaching ability in terms of 100's and 50's, number of 4's and 6's, and with individual high score. The batsmen score is calculated with the same parameters for batters' positions 1 to 10. Then a maximum of batsmen score is selected after evaluating each position of batsmen (' β _pos_score'). ' γ _inningwise_score' reflects the batsmen's performance inning wise with weights are assigned to first and second innings depending on the pressure value on players. Similarly, the performance of batsmen at home/away matches (' x _venue'), opponent wise (' y _opponent') is calculated. The current and previous five years' statistics are also analysed to reflect the quality of batsmen for selection (' w _yearwise'). The overall batting score ('batting_score') of batsmen is calculated using all the parameters mentioned above with specific weights assigned after studying each parameter's impact. At least four batsmen are required to form the quality side. A team with high-performing batters increases the chances of winning the match. So, it is necessary to select good batters. Algorithm 1 helps to find the batsmen with good performance for every parameter.

Algorithm 1 Batsmen strength

```

1: for all players p do
2:  $\alpha$ _batsmen_score = 0.30*run_scored+0.05*notout_innigs+
  0.20*bat_avg + 0.15*bat_sr+0.15*milestone_reaching_ability+
  0.10*no_of_4's_6's+0.05*high_score-0.05*no_of_zeroes
3:  $\beta$ _pos_score = max ( $\alpha$ _batsmen_score_at_each_position)
4:  $\gamma$ _inningwise_score = 0.40* $\alpha$ _batsmen_score_first_inning+
  0.60* $\alpha$ _batsmen_score_secnd_inning
5:  $x$ _venue = 0.35* $\alpha$ _batsmen_score_home_matches+0.65*
   $\alpha$ _batsmen_score_away_matches
6:  $y$ _opponent = 0.70* $\alpha$ _batsmen_score_strong_opponent+0.30*
   $\alpha$ _batsmen_score_weak_opponent
7:  $w$ _yearwise = 0.20* $\alpha$ _batsmen_score_current_year+0.80*
   $\alpha$ _batsmen_score_last_five_year
8: batting_score = 0.25* $\alpha$ _batsmen_score+0.10* $\beta$ _pos_score+0.15*
   $\gamma$ _inningwise_score+0.10* $x$ _venue +0.15*
   $y$ _opponent+0.20* $w$ _yearwise+0.05*captain
9: endfor

```

Algorithm 2 is used to calculate the bowling strength of players. The bowling strength of players is measured with the number of wickets taken, bowling average, strike rate, economy rate, a total of 4/5 wicket haul, maximum wickets taken in one match, and count of maiden overs bowled (' α _bowler_score'). The bowler's efficiency is also measured concerning inning number (' β _inningwise_score'), venue (' γ _venue'), opponent (' x _opponent'), and yearwise (' w _yearwise') bowling score. The appropriate weights are assigned to calculate the value of every parameter for bowlers. The final bowling score ('bowling_score') combines all parameters with proper importance. The chances of winning

the match are also directly proportional to the team's bowling attack (see Table 5).

Algorithm 2 Bowler Strength

```

1: for all players p do
2:  $\alpha$ _bowler_score = 0.30*wickets_taken+0.20*bowl_avg + 0.10*
  bowl_sr+0.15*eco_rate +0.10*no_of_4_5_wicket_haul+0.05*
  max_wickets_taken+0.10*maiden_overs
3:  $\beta$ _inningwise_score = 0.40* $\alpha$ _bowler_score_first_inning+0.60*
   $\alpha$ _bowler_score_secnd_inning
4:  $\gamma$ _venue = 0.40* $\alpha$ _bowling_score_home_matches+0.60*
   $\alpha$ _bowling_score_away_matches
5:  $x$ _opponent = 0.80* $\alpha$ _bowling_score_strong_opponent+0.20*
   $\alpha$ _bowling_score_weak_opponent
6:  $w$ _yearwise = 0.20* $\alpha$ _bowling_score_current_year+0.80*
   $\alpha$ _bowling_score_last_five_year
7: bowling_score = 0.30* $\alpha$ _bowling_score+0.15* $\beta$ _inning
  wise_score+0.10* $x$ _venue+0.15* $x$ _opponent
  +0.10* $w$ _yearwise+0.15*wicket_taken_performance+0.05*captain
8: endfor

```

Algorithm 3 is used to select a batting all-rounder for the team. A batting all-rounder is a player who is good at batting and bowling. The batting all-rounder is best in batting performance as compared to bowling. Algorithms 1 and 2 are used to assess batting all-rounders by measuring the player's batting and bowling strength. The overall batting and bowling score are calculated. The difference between average batting and bowling score is obtained. At last, the batting all-rounder score is calculated with more weight assigned to the batting score than the bowling score. The teams require at least one batting all-rounder to form a balanced squad. Using our algorithm, the batting all-rounder can be found.

Algorithm 3 batting all-rounder

```

1: for all players p do
2: calculate overall batting score
3: calculate overall bowling score
4:  $x$ _diff = batting_score-bowling_score
5: batting_allrounder_score = 0.50*batting_score+0.30*
  bowling_score+0.20* $x$ _diff
6: end for

```

Algorithm 4 is used to choose the team's bowling all-rounder. A bowling all-rounder is a player who succeeds at both bowling and batting. In comparison to batting, the bowling all-rounder is the finest in bowling. Algorithms 1 and 2 are utilized to estimate a player's batting and bowling strength to evaluate a bowling all-rounder. After calculating the total batting and bowling scores, the difference between the average batting and bowling scores is computed. Finally, the bowling all-rounder score is determined, receiving greater weight than the batting score. Each side needs at least one bowling all-rounder to build a balanced line-up which can be determined using our method (see

Table 5
Accuracy for selection of bowler.

Classifier	Accuracy								
	Without Feature Optimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
Logistic Regression	80.64	84.28	83.43	82.31	86.04	86.91	87.79	91.56	91.86
Naïve Bayes	77.85	79.54	82.85	81.21	86.64	91.08	90.35	90.57	92.76
KNN	78.92	80.30	87.87	89.87	89.97	89.61	90.56	91.87	93.89
SVM	87.09	89.28	88.29	89.31	90.33	90.74	93.46	94.33	97.04
Decision Tree	79.28	82.35	85.17	84.02	88.42	89.62	90.39	92.36	92.71
Random Forest	86.62	87.53	87.34	88.60	89.57	90.64	90.85	95.53	95.92
GBM	85.07	86.25	88.17	89.33	88.37	89.24	91.27	94.47	96.05
XGBoost	83.28	84.47	85.49	85.24	89.07	92.53	93.12	93.73	95.67
CatBoost	85.46	86.33	87.02	90.69	90.64	91.64	91.34	93.63	96.22

Table 7

Accuracy for Selection of Bowling all-rounder.

Classifier	Accuracy								
	Without Feature Optimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
Logistic Regression	70.96	71.78	75.57	72.21	76.5	72.63	73.47	77.28	94.35
Naive Bayes	76.82	79.33	82.25	83.49	83.92	84.23	86.60	94.93	95.28
KNN	78.71	80.23	80.71	81.78	83.14	83.85	85.5	92.73	93.14
SVM	85.70	83.57	84.13	85.35	82.64	82.11	93.56	95.52	97.29
Decision Tree	75.76	79.59	80.35	84.71	84.64	87.15	88.91	93.31	96.32
Random Forest	76.64	77.55	78.16	79.78	86.31	86.12	85.25	95.77	96.52
GBM	82.95	84.78	84.42	85.05	92.23	90.24	90.52	94.76	96.43
XGBoost	83.20	84.63	87.17	86.73	85.89	87.36	90.46	95.05	96.20
CatBoost	80.85	87.21	88.47	86.94	90.68	85.64	92.30	96.24	96.69

Table 7).

Algorithm 4 bowling all-rounder

- 1: for all players p do
- 2: calculate overall batting score
- 3: calculate overall bowling score
- 4: $x_diff = \text{batting_score} - \text{bowling_score}$
- 5: $\text{bowling_allrounder_score} = 0.35 * \text{batting_score} + 0.45 * \text{bowling_score} + 0.20 * x_diff$
- 6: end for

The assessment of the wicketkeeper is performed with algorithm 5. In the first step, the wicketkeeper's overall batting score is calculated. The player's wicketkeeping ability (' β _wicket_keeping_ability') is calculated using the number of catches taken behind the wicket and the number of stumpings performed. The final wicketkeeper score is calculated using a weighted batting score and wicketkeeping ability. The team compulsory requires a wicketkeeper.

Algorithm 5 wicket keeper performance

- 1: for all players p do
- 2: calculate overall batting score
- 3: $\beta_wicket_keeping_ability = 0.70 * \text{no_of_catches} + 0.30 * \text{no_of_stumpings}$
- 4: $\text{wicket_keeper_score} = 0.45 * \text{batting_score} + 0.55 * \beta_wicket_keeping_ability$
- 5: end for

5. Results

Various binary and categorical characteristics are employed to create a team prediction model for one-day international cricket. The data are converted into a consistent format for experimentation. Some features are derived from the weighted combination of existing features. The batsmen's strength is calculated using 25 features, bowlers with 23, 45 for batting/bowling all-rounder, and 23 for a wicketkeeper described in Tables 1 and 2. The Cross-validation method of model selection is not used to preserve the chronological order of data. The training-testing data splitting method is used for model selection as the future match result is based on the outcome of previous matches. Five algorithms are proposed for the selection of players from each category. The players are categorised into five classes according to their final score obtained from

Table 3
Players class.

Player score	Category
41–50	Excellent
31–40	Very Good
21–30	Good
11–20	Satisfactory
0–10	Poor

algorithms as output. Depending on their performance, the player evaluations are excellent, very good, good, satisfactory, and poor. The players are divided into these five classes according to their score given below in Table 3:

The batsmen, bowlers, batting all-rounder, bowling all-rounder, and wicketkeeper are classified into five classes described in Table 3. After that, players are chosen. The results are compared with and without the use of feature optimization techniques. The best prediction accuracy is obtained by combining a machine learning model with a feature optimization approach. The hybrid approach of CS-PSO is used to select a good team, as it helps to select optimum features for the input of the machine learning classifier. If the features having more impact are provided as input to the ML algorithm, it improves algorithm efficiency. The performance of the hybrid approach is also compared with another feature optimization algorithm. This section discusses the results for all models using classifier assessment measures.

Algorithm 1: Batsmen selection

The batsmen evaluation is performed using Algorithm 1. Then machine learning algorithms are applied to classify batters into one of the five class/categories, and the classification accuracy is calculated. After that, Nature Inspired algorithms are applied for feature optimization. Feature optimization selects the more essential features or has more weight in classification. The accuracy of the batter's selection is compared with or without a feature optimization algorithm.

The following findings are noted from Table 4:

- Without feature optimization, the SVM obtained a high accuracy of 93.54%.
- The Naive Bayes algorithm has the lowest accuracy of 71.42%.
- The use of feature optimization approaches utilising a nature-inspired methodology improved the accuracy of machine learning models.
- With CS-PSO Logistic Regression method achieved a maximum accuracy of 94.28%, whereas Naive Bayes earned a maximum accuracy of 93.93%.

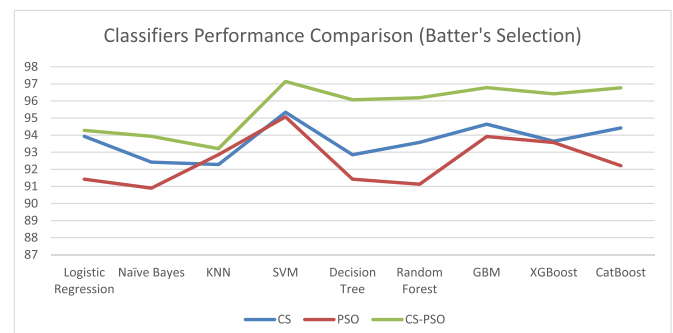
**Fig. 4.** Classifiers Performance Comparison for batter's selection.

Table 4

Accuracy for selection of batsmen.

Classifier	Accuracy								
	Without Feature Optimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
Logistic Regression	78.57	81.68	83.33	84.85	85.71	86.36	91.42	93.93	94.28
Naive Bayes	71.42	83.87	92.55	88.57	82.85	85.28	90.90	92.42	93.93
KNN	77.41	89.39	85.11	80.25	90.90	86.58	92.86	92.28	93.21
SVM	93.54	95.07	95.93	94.28	94.55	95.32	95.07	95.34	97.14
Decision Tree	87.09	88.82	89.83	92.33	92.86	92.50	91.42	92.85	96.07
Random Forest	86.62	90.35	89.02	90.50	91.14	94.23	91.13	93.58	96.19
GBM	90.71	92.40	91.21	92.14	94.29	93.35	93.92	94.64	96.78
XGBoost	91.42	92.15	91.60	91.96	92.87	93.07	93.57	93.64	96.42
CatBoost	90.32	91.42	93.54	94.28	95.45	90.28	92.21	94.42	96.77

- The maximum accuracy of SVM with CS-PSO is 97.14% for batter's selection. The CS-PSO accuracy is better as compared to individual CS and PSO algorithm.
- The performance comparison of all classifiers with individual CS, PSO and Blended CS-PSO is shown in Fig. 4. From the graph, it is clear that the performance of the CS-PSO algorithm is better for all the classifiers during batter's selection.

Algorithm 2: Bowler's selection

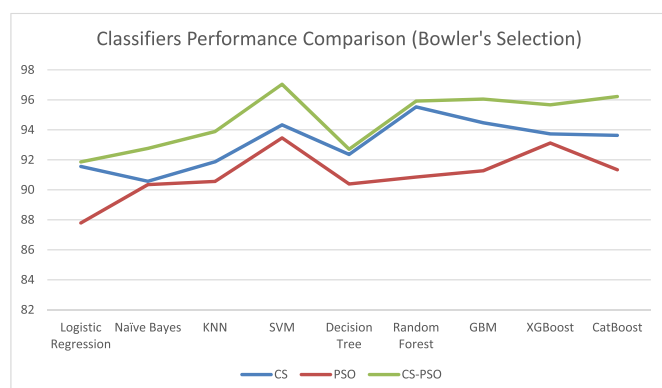
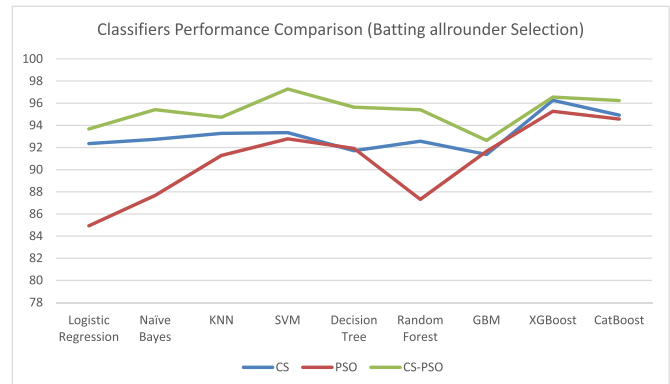
The bowler is evaluated using Algorithm 2. The classification accuracy is then determined using machine learning algorithms to classify bowlers into five classes/categories. After that, algorithms inspired by nature are used. Bowler selection accuracy is compared with and without using a feature optimization method. From the above Table 5 following observations are made:

- Instead of using a feature optimization procedure, the SVM method produces an overall accuracy of 87.09%.
- With an accuracy of 77.85%, the Naive Bayes algorithm is the least accurate.
- The accuracy of Logistic Regression improved to 91.86% after applying the CS-PSO hybrid approach.
- SVM with a hybrid CS-PSO has a maximum accuracy of 97.04%.
- After using a Nature-Inspired algorithm, all machine learning algorithm's accuracy improves significantly.
- The bowler's selection performance comparison for all classifiers is shown in Fig. 5. The bowlers are also important pillars of the team, and CS-PSO selects the bowlers with maximum accuracy.

Algorithm 3: Batting all-rounder selection

The batting all-rounder is evaluated using Algorithm 3. Machine learning algorithms are then used to determine the accuracy of the categorization, which categorizes batting all-rounders into one of five classes/categories.

The following is a description of the results using Table 6:

**Fig. 5.** Classifiers Performance Comparison for bowler's selection.**Fig. 6.** Classifiers Performance Comparison for batting allrounder selection.

- SVM achieves the highest accuracy of 87.29%, while Logistic Regression achieves the lowest accuracy of 71.64%.
- The accuracy of the SVM and CS-PSO pair is 97.28%.
- All machine learning models obtained good accuracy after applying hybrid CS-PSO algorithm.
- The accuracy of a machine learning classifier adopting the Nature Inspired method has improved significantly.
- The comparative performance of CS, PSO and Blended CS-PSO is shown in Fig. 6. The CS-PSO achieved maximum accuracy for selection of the batting allrounder. Batting allrounder is having better batting capability as compared to bowling capability.

Algorithm 4: Bowling all-rounder selection

The bowling all-rounder is assessed using Algorithm 4. Bowling all-rounders are players who are more effective at bowling than at batting. From the above Table 7 following observations are made:

- If the machine learning models are assessed without utilising feature optimization, the SVM method achieves an accuracy of 85.70%, whereas Logistic Regression achieves 70.96%.
- The maximum accuracy for bowling all-rounder selection is achieved with a combination of CS-PSO and SVM having an accuracy of 97.29%.
- The accuracy of a bowling all-rounder is improved using a hybrid method of Nature Inspired algorithms and Machine Learning models.
- The performance for bowling allrounder selection is shown in Fig. 7.

Algorithm 5: Wicketkeeper selection

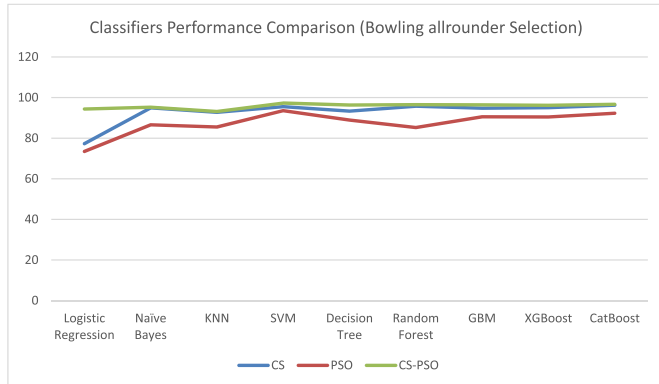
Due to cricket's regulatory constraints of cricket one wicketkeeper is required in the squad to complete the line-up for the side. Our suggested method 5 for selecting the wicketkeeper satisfies this criterion. Table 8 shows the accuracy attained using machine learning models and the feature optimization approach.

The following observations are drawn from the table above:

Table 6

Accuracy for Selection of batting all-rounder.

Classifier	Accuracy								
	Without Feature optimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
Logistic Regression	71.64	73.57	77.14	74.28	76.00	77.27	84.93	92.35	93.67
Naïve Bayes	78.21	84.92	85.22	82.84	84.30	86.48	87.68	92.74	95.42
KNN	77.57	85.48	86.07	86.45	88.27	89.29	91.28	93.28	94.73
SVM	87.29	86.78	88.03	88.36	89.83	92.07	92.78	93.34	97.28
Decision Tree	81.82	83.46	84.34	87.06	86.81	90.21	91.92	91.71	95.64
Random Forest	82.14	82.75	83.13	84.22	83.54	87.50	87.31	92.57	95.41
GBM	84.85	88.58	89.35	88.55	89.65	89.50	91.66	91.37	92.64
XGBoost	86.84	87.05	87.95	88.74	91.79	94.50	95.27	96.26	96.55
CatBoost	85.66	86.02	86.76	89.53	90.11	93.00	94.57	94.92	96.24

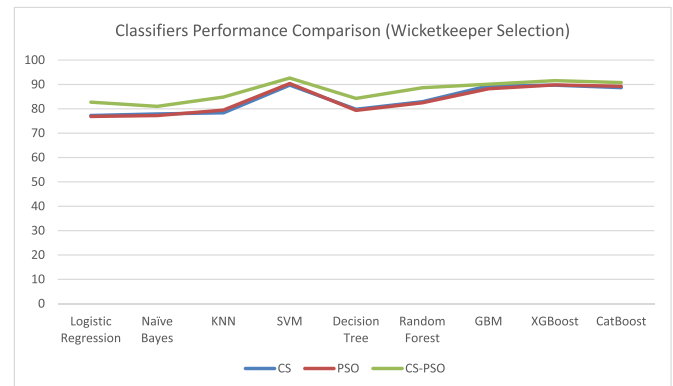
**Fig. 7.** Classifiers Performance Comparison for bowling allrounder selection.

- The accuracy of the SVM algorithm was 84.21%.
- With a 70.58% accuracy rate, KNN is not up to the task of selecting a wicketkeeper.
- With an accuracy of 92.63%, SVM and the CS-PSO are used to choose wicketkeepers.
- The wicketkeeper selection classifiers performance is shown with Fig. 8.

The classification report has the value of precision, recall, and F1-Score is shown in Table 9. Forming a good team is the main task of this work, and it is better shown with a classification report. From the classification report, our CS-PSO and SVM algorithms approach finds the team with quality players to improve the team's winning chances. We also used a paired *t*-test to see which of the nine classifiers showed more significant improvement than the others. To compare the results of measuring one group twice, a paired *t*-test is used. This statistical hypothesis technique estimates the *t*-value by taking the mean and variance of the differences between these two measures and running them several times. The probability that these two measurements are significantly different can be calculated using the *t*-value and the ideal

statistical significance (0.05) by consulting the *t*-distribution table [27]. The mean and standard deviation is calculated by selecting Logistic regression as the base classifier. The algorithms are run 15 times on the given dataset with an appropriate group to get the accurate value of mean and standard deviation for each category of player selection, and the best value is shown in Table 9.

In this work, players' performance is studied and evaluated considering all parameters. The parameters required for assessing batters are

**Fig. 8.** Classifiers Performance Comparison for wicketkeeper selection.**Table 9**

Classification report of player selection.

Player category	Precision	Recall	F1-Score	Mean	Standard Deviation
Batsmen	97	97	98	95.33	0.88
Bowler	96	96	97	91.59	3.03
Batting all-rounder	97	97	98	91.05	3.46
Bowling all-rounder	97	97	98	88.02	6.31
Wicketkeeper	92	92	93	86.98	3.36

Table 8

Accuracy of wicketkeeper selection.

Classifier	Accuracy								
	Without Feature Optimization	GWO	MFO	WOA	FFA	BAT	PSO	CS	CS-PSO
Logistic Regression	72.73	74.31	76.47	74.52	75.78	76.20	76.87	77.26	82.73
Naïve Bayes	71.84	72.42	72.84	73.05	73.68	79.14	77.21	77.89	81.01
KNN	70.58	73.26	74.10	74.73	75.15	78.48	79.46	78.34	84.81
SVM	84.21	82.70	86.03	85.58	83.81	87.80	90.38	89.74	92.63
Decision Tree	70.94	75.57	76.63	77.47	78.31	78.73	79.36	79.82	84.25
Random Forest	80.57	80.63	81.47	82.52	80.98	81.34	82.48	82.92	88.63
GBM	82.27	85.26	85.68	87.57	85.47	86.10	88.24	89.42	90.10
XGBoost	83.48	85.21	89.21	83.97	86.95	84.61	89.80	89.68	91.57
CatBoost	84.14	88.21	86.47	85.36	87.82	88.46	89.26	88.69	90.78

identified and analysed concerning every condition. For e.g. batter's performance is studied position-wise, weights are given according to the position is batted, maximum weight is assigned to a position where batters have a good average. So, batter's evaluation is performed for every parameter with the appropriate weight, and based on that, the batter's strength algorithm is proposed. Similarly, every parameter is studied and assessed carefully to evaluate bowlers, batting allrounders, bowling allrounders, and wicketkeepers. Based on the impact of parameters on the assessment of player weightage is decided. Then algorithms are proposed for bowlers, batting allrounder, bowling allrounder, and wicketkeeper. The Nature-Inspired algorithms are not used in literature to remove the irrelevant, redundant, and noisy features. We use the swarm intelligence algorithm to remove extra features from the dataset to get maximum accuracy for selecting the team. The blending of the CS-PSO algorithm is performed to improve the algorithm's efficiency, which is also one of the novel approaches used for the team prediction for this work.

The baseline models do not exist to compare work with others due to the availability of benchmark datasets. Nevertheless, the parameters used to predict the players of a cricket match are basically equivalent to forming a quality team. So, when we compromise the condition and evaluate our approach to the reference, we find that the accuracy gained is superior to the existing work. The maximum accuracy of 93.46% is achieved for team player selection from the literature [6]. Our approach of CS-PSO and SVM obtained a maximum average accuracy of 97% for the selection of players from each category as compared to the previous works [4,8,10]. Hence, the hybrid system of CS-PSO and SVM finds the well-balanced team for one day international matches with better accuracy.

6. Discussion

This work uses an ensemble of machine learning models in conjunction with feature optimization approaches to achieve the highest accuracy in predicting the team for an ODI match. The feature optimization approaches generate high accuracy with fewer characteristics as input for machine learning models. Feature optimization strategies are used to pick the features that have the most significant influence on player selection. The Nature-Inspired Metaheuristic method, which is inspired by natural creatures or swarms' behaviour, effectively selects feature subsets from a dataset. Every feature does not contribute equally to player assessment. Some characteristics have a greater influence on the machine learning classifier's result, whereas others do not. Feature optimization approaches identify characteristics with higher weights to enhance the classifier. We employ numerous Nature Inspired algorithms with a hybrid system of CS-PSO to pick the team for one-day international matches.

Batting average, strike rate, and milestone reaching ability are essential in batsman selection since they describe its consistency and scoring ability. Bowling average, strike rate, economy rate, and performance in away match significantly influence bowlers. Batting strength-related variables such as batting average and strike rate positively impact batting all-rounder selection, whereas bowling features affect bowling all-rounder selection. The number of catches behind the wicket and stumpings significantly contributes to the wicketkeeper selection compared to batting features. The wicketkeeper's glub is more important than his batting skill. After periodically estimating the model on training and testing data from the dataset, the correct blend of machine learning model and feature optimization approach is discovered. Combining the hybrid system of CS-PSO and SVM algorithm is better to select a team. The accuracy of models is also enhanced with less training time using all other feature optimization approaches [7,10,11,16–23].

The players with good scores and ratings are selected based on the CS-PSO hybrid approach and SVM algorithm. These players will only be considered for inclusion in the team by the selection member under separate abilities. We compare the team form with our approach to the

Table 10

Team Selection Comparison Ind vs Aus [28] & Ind Vs NZ Series 2019 [29].

Name of Player	Player Score	Player Rating	Role of Player	Member of team (Ind vs Aus)	Member of team (Ind vs NZ)
Virat Kohli	46	Excellent	Batsman	Yes	Yes
Rohit Sharma	44	Excellent	Batsman	Yes	Yes
Shikhar Dhawan	43	Excellent	Batsman	Yes	Yes
Ajinkya Rahane	38	Very Good	Batsman	No	No
Vijay Shankar	26	Good	Batting all-rounder	Yes	Yes
Ravindra Jadeja	38	Very Good	Bowling all-rounder	Yes	No
M S Dhoni	46	Excellent	Wicketkeeper	Yes	Yes
Bhuvneshwar Kumar	44	Excellent	Bowler	No	Yes
Mohammed shami	41	Excellent	Bowler	Yes	Yes
Kuldeep Yadav	37	Very Good	Bowler	Yes	Yes
Jasprit Bumrah	38	Very Good	Bowler	Yes	No

team selected for India-Australia Series [28] and the India-New Zealand Series [29] shown in Table 10. Table 10 display the effectiveness of the cricketers who have been actually selected for the team and the players from diverse expertise who were denied a spot on the national side. For India-Australia Series, Ajinkya Rahane and Bhuvneshwar Kumar were not selected, although they have a good player score and rating. In the India-New Zealand series, Ajinkya Rahane and Ravindra Jadeja were not part of the team after having a good record and chances of selection in the team. But most of the players selected from our models are part of the team playing eleven in the India-Australia and India-New Zealand series as per their skills.

7. Conclusion and future work

Personnel selection is a critical issue for institutions seeking to improve their performance. Team selection is one example of this problem where the goal is to identify team members. In most cases, the rating of applicants is used to improve both the people and team selection processes. The appropriate player selection for every match has a massive impact on the game's result. Team organization members may pick each game's best players based on a fair prediction of how many runs a batter will score and how many wickets a bowler will take in a match. On the other hand, such conclusions are only be drawn from data gathered from diverse sources. Based on the players' data and qualities, we created a model for selecting an 11-member squad in this work. This article examined how to classify players in one-day international cricket. We studied putting players into one of the five categories using a dataset of 414 players from India playing ODI within a rule restriction for team selection, using the machine learning approaches and a feature optimization algorithm. We used nine machine learning approaches in this study to determine which class each player should fall. Following the initial implementation of the aforementioned algorithms, SVM has a prediction accuracy of 93.54% for selecting batters and 87.29% for selecting batting all-rounders. Bowlers with SVM have an accuracy of 87.09%, for bowling all-rounder is 85.70%, and wicketkeeper selection is 84.21%. With the correct parameter selection, we enhanced the forecast accuracy even further. We improved the prediction accuracy for batsmen selection using a combination of SVM and CS-PSO up to 97.14%. Similarly, employing a CS-PSO, the accuracy of SVM for bowler and bowling all-rounder selection was increased to 97.04% and 97.29%, respectively. The SVM with a CS-PSO provides a maximum prediction accuracy of 97.28% for batting all-rounder. The wicketkeeper is picked

with an accuracy of 92.63% using the hybrid approach, which selects acceptable values for the parameters. In conclusion, our results showed that the Nature Inspired algorithm beat machine learning approaches by a wide margin. The outcomes of this study assist cricket authorities and players in various ways. These results can be used by player selection committees, team coaches, and captains to find suitable players.

This work can be expanded in the future to include other parameters that affect the player's performance. Additional performance metrics like information regarding the opponent teams that the players play against should be included in future research. After adding additional characteristics to the data, the ultimate goal will be to enhance the classification model's accuracy. With appropriate data and feature changes, this technique may be applied to team prediction in Twenty 20 matches.

Author contribution statement

Manoj S Ishi (Corresponding author): Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Visualization **Jayantrao Patil** and **Vaishali Patil**: Writing - Resources, Review & Editing, Supervision, Project administration.

References

- [1] Saikia H. Quantifying the current form of cricket teams and predicting the match winner. *Management and Labour Studies* 2020;45(2):151–8. <https://doi.org/10.1177/0258042X20912603>.
- [2] Wickramasinghe I. Classification of all-rounders in the game of ODI cricket: machine learning approach. *Athens Journal of Sports* 2020;7(1):21–34. <https://doi.org/10.30958/ajspo.7-1-2>.
- [3] Vetukuri VS, Sethi N, Rajender R. Generic model for automated player selection for cricket teams using recurrent neural networks. *Evolutionary Intelligence* [Internet] 2021;14(2):971–8. <https://doi.org/10.1007/s12065-020-00488-4>.
- [4] Khot A, Shinde A, Magdum A. Rising star evaluation using statistical analysis in cricket. *Lecture Notes in Networks and Systems* 2021;127:317–26. https://doi.org/10.1007/978-981-15-4218-3_31.
- [5] Ahmed F, Jindal A, Deb K. Cricket team selection using evolutionary multi-objective optimization. *Studies in Computational Intelligence* 2021;938:71–8. https://doi.org/10.1007/978-3-030-63773-6_2.
- [6] Balasundaram A, Ashokkumar S, Jayashree D, Magesh Kumar S. Data mining-based classification of players in game of cricket. In: *Proceedings of the International Conference on Smart Electronics and Communication. ICOSSEC*; 2020. <https://doi.org/10.1109/ICOSSEC49089.2020.9215413>. 271–5.
- [7] Bello M, Bello R, Nowé A, García-Lorenzo MM. A method for the team selection problem between two decision-makers using the ant colony optimization. *Stud Fuzziness Soft Comput* 2018;357:391–410. https://doi.org/10.1007/978-3-319-60207-3_23.
- [8] Kapadiya C, Shah A, Adhvaryu K, Barot P. Intelligent cricket team selection by predicting individual players' performance using efficient machine learning technique. *Int J Eng Adv Technol* 2020;9(3):3406–9. <https://doi.org/10.35940/ijeat.c6339.029320>.
- [9] Chhabra P, Ali R, Pudi V. CRICTRS: embeddings based statistical and semi supervised cricket team recommendation system. In: *Proceedings of the International Conference on Machine Learning Techniques and NLP*; 2020. p. 67–77. <https://doi.org/10.5121/csit.2020.101207>. October.
- [10] Ahmad H, Ahmad S, Asif M, Rehman M, Alharbi A, Ullah Z. Evolution-based performance prediction of star cricketers. *Comput Mater Continua (CMC)* 2021;69(1):1215–32. <https://doi.org/10.32604/cmc.2021.016659>.
- [11] Khurma RA, Aljarah I, Sharihe A, Mirjalili S. EvoloPy-FS: an open-source nature-inspired optimization framework in Python for feature selection. *Lecture Notes in Algorithms for Intelligent System* 2020:131–73. https://doi.org/10.1007/978-981-32-9990-0_8.
- [12] Arora S, Singh H, Sharma M, Sharma S, Anand P. A new hybrid algorithm based on Grey Wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. *IEEE Access* 2019;7:26343–61. <https://doi.org/10.1109/ACCESS.2019.2897325>.
- [13] Agrawal P, Ganesh T. Selection of Indian cricket team in ODI using integer optimization. *J Phys Conf* 2020;1478(1). <https://doi.org/10.1088/1742-6596/1478/1/012001>.
- [14] Mahmood Z, Daud A, Abbasi RA. Using machine learning techniques for rising star prediction in basketball. *Knowl Base Syst* 2021;211:106506. <https://doi.org/10.1016/j.knsys.2020.106506> [Internet].
- [15] Kusumsiri I, Perera S. Optimal one day international cricket team selection by genetic algorithm. *Int J Sci Basic Appl Res* 2017;36(4):213–21. <https://gssrr.org/index.php/JournalOfBasicAndApplied/article/view/8097>.
- [16] Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. *Soft Comput* 2018;22(2):387–408. <https://doi.org/10.1007/s00500-016-2474-6>.
- [17] Al-abaji MA. A literature review of cuckoo search algorithm. *J Educ Pract* 2020:1–8. <https://doi.org/10.7176/jep/11-8-01>.
- [18] Yang X, Deb S, Behaviour ACB. Cuckoo search via levy flights. In: *Proceedings of the World Congress on Nature & Biologically Inspired Computing*, vols. 210–4. NaBIC; 2009. <https://doi.org/10.1109/NaBIC.2009.5393690>.
- [19] Mirjalili S, Mirjalili SM, Lewis A. Grey Wolf optimizer. *Adv Eng Software* 2014;69:46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007> [Internet].
- [20] Mirjalili S, Lewis A. The Whale optimization algorithm. *Adv Eng Software* 2016;95:51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008> [Internet].
- [21] Yang XS. A new metaheuristic Bat-inspired Algorithm. *Studies in Computational Intelligence* 2010;284:65–74. https://doi.org/10.1007/978-3-642-12538-6_6.
- [22] Yang XS. Firefly algorithm, stochastic test functions and design optimization. *Int J Bio-Inspired Comput* 2010;2(2):78–84. <https://doi.org/10.1504/IJBIC.2010.032124>.
- [23] Mirjalili S. Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl Base Syst* 2015;89(July):228–49. <https://doi.org/10.1016/j.knsys.2015.07.006> [Internet].
- [24] Ishi MS, Patil JB. A study on impact of team composition and optimal parameters required to predict result of cricket match. *Lect Notes Networks Syst* 2020;100:389–99. https://doi.org/10.1007/978-981-15-2071-6_32.
- [25] Ishi MS, Patil JB. A study on machine learning methods used for team formation and winner prediction in cricket. *Lect Notes Networks Syst* 2021;173. https://doi.org/10.1007/978-981-33-4305-4_12. LNNS:143–56.
- [26] <https://www.espnccricinfo.com/>.
- [27] Khorasgani RR. Comparison of different classification methods. *Heart Dis* 2010;78(15):83–4.
- [28] <https://www.cricbuzz.com/cricket-scores/21532/ind-vs-aus-1st-odi-australia-tour-of-india-2019>.
- [29] <https://www.cricbuzz.com/cricket-scores/20783/nz-vs-ind-1st-odi-india-tour-of-new-zealand-2019>.