Volume 07 Issue 02 (2026) Pages 107-115

INSPIREE: INDONESIAN SPORT INNOVATION REVIEW

ISSN 2746-6965 (Online), 2774-2520 (Print)
Journal Homepage: https://inspiree.review/index.php/inspiree

Original Research Articles

OPEN ACCESS

A Performance Metrics–Based Model for Predicting Match Outcomes in the 2023 ICC Cricket World Cup

https://doi.org/10.53905/inspiree.v7i02.172

Ajay Kumar[*1abcde], **Anurodh Sisodia**[2c], **Yogesh Chander Takhur**[1ad]

[1]Government College Chinnor, Gwalior, Madhya Pradesh, India.
[2]LNIPE Gwalior, Madhya Pradesh, India.
[3]Sports University of Haryana, Sonipat, Haryana, India.

ABSTRACT

The purpose of the study. To develop and validate a logistic regression model for predicting match outcomes in the 2023 ICC Cricket World Cup using selected in-game performance indicators and to determine the relative contribution of each variable.

Materials and methods. Data from 47 matches of the 2023 ICC Men's Cricket World Cup were analysed (one match decided by the Duckworth-Lewis-Stern method was excluded). Independent variables included Toss outcome, Opening Partnership Score, Runs and Wickets Lost in Powerplay, Total Number of Fours and Sixes, and Total Wickets Lost in an Inning. Binary logistic regression was applied to predict match outcomes (win/loss). Model goodness-of-fit was evaluated using the Hosmer-Lemeshow test.

Results. Wickets Lost in an Inning was the strongest predictor (OR = 2.324, $p < 0.05$); each additional wicket lost increased the odds of losing by 132.4%. Each additional four reduced the odds of losing by 13.7% (OR = 0.863). Total sixes and other variables showed weaker or non-significant effects. Toss outcome and Opening Partnership Score were not statistically significant predictors. The final model demonstrated good fit (Hosmer-Lemeshow test, $p > 0.05$) and acceptable predictive accuracy.

Conclusions. Preserving wickets throughout the innings and maximising boundary scoring (especially fours) are the most critical factors influencing match outcomes in the 2023 ICC Cricket World Cup. The developed logistic regression model offers a reliable tool for performance analysis and strategic decision-making in limited-overs cricket.

Keywords: Cricket, 2023 World Cup, Match Outcome Prediction, Logistic Regression, Performance Indicators, Wickets Lost, Boundary Scoring.

ARTICLE INFO

EDITED BY
Dr. Samsurijal Hasan MM
Universitas Suryakancana, Indonesia.

Dr. Eneng Fitri Amalia, M.Pd
Universitas Suryakancana, Indonesia.

ARTICLE HISTORY
Received : November 20, 2025.
Accepted : January 24, 2026.
Published: May 27, 2026.

CITATION
Kumar, A., Sisodia, A., & Takhur, Y. C. (2025). A Performance Metrics–Based Model for Predicting Match Outcomes in the 2023 ICC Cricket World Cup. *INSPIREE: Indonesian Sport Innovation Review*, 7(02), 107-115. https://doi.org/10.53905/inspiree.v7i02.172

INTRODUCTION

The integration of data analytics into modern sports has revolutionized performance evaluation, tactical planning, and outcome prediction across disciplines. Cricket, being one of the most statistically rich sports in the world, has particularly benefited from this transformation. The availability of ball-by-ball data, advanced tracking technologies, and real-time statistical platforms has shifted analytical focus from traditional aggregate metrics to dynamic, phase-specific performance indicators that better reflect the evolving nature of limited-overs cricket (Mullick, 2024; Sanjaykumar et al., 2024).

The 2023 ICC Men's Cricket World Cup, contested between ten elite teams across eleven venues in India, generated an exceptionally rich dataset characterized by diverse playing conditions, high stakes, and varying pitch behaviors (Sanjaykumar et al., 2024). These factors make the tournament an ideal laboratory for testing predictive models based on in-game variables rather than pre-match attributes such as team rankings, historical head-to-head records, or player reputation, which often fail to capture real-time momentum shifts (McEwan et al., 2023).

[abcde]Authors'Contribution: a-Study design; b-Data collection; c-Statistical analysis; d-Manuscript preparation; e-Funds collection.
Corresponding Author: Ajay Kumar; e-mail: drajaydhankhar@gmail.com

Critical examination of existing literature

Existing literature has investigated multiple dimensions of cricket analytics. Early studies focused predominantly on pre-match predictors like home advantage, toss outcome, and team composition (Allsopp & Clarke, 2004; I.Y et al., 2025; McEwan et al., 2023). More recent work has shifted toward in-play variables: powerplay performance(Akl & Hasan, 2025; Bandyopadhyay & Bhattacharjee, 2025), partnership strength and boundary frequency(Ozay et al., 2025; Vora, 2025), and wicket-loss patterns across innings phases (Pedarla, 2025; Sravani et al., 2025). However, most published models either predict first-innings totals using linear regression(Luo et al., 2025) or employ machine learning algorithms on historical data spanning multiple tournaments and eras (Jayalath, 2017; Lansiaux et al., 2025; Tandililing et al., 2025). Few studies have applied logistic regression exclusively to a single high-profile tournament to isolate the true impact of contemporary in-game indicators under uniform competitive pressure (Burra, 2023; McEwan et al., 2023).

Identification of research gaps

Despite these advances, significant research gaps persist. First, the relative importance of commonly discussed variables has not been systematically ranked within the same elite tournament using a single robust statistical framework. Second, many earlier models include variables from both batting and bowling teams simultaneously, introducing multicollinearity and reducing interpretability for practical coaching applications. Third, there remains limited empirical evidence on how these factors interacted specifically under 2023 World Cup conditions—characterized by high-scoring venues, new white-Kookaburra balls, and aggressive modern batting approaches (Mullick, 2024; Sanjaykumar et al., 2024).

Rationale for the research

Addressing these gaps is crucial because accurate identification of the most influential in-game predictors can directly inform real-time strategic decision-making: when to attack or defend in the powerplay, how aggressively to rotate strike in the middle overs, and how to balance risk versus wicket preservation in death overs (Puram et al., 2022). Such insights are invaluable not only for coaches and analysts but also for broadcasters, fantasy platforms, and betting markets that increasingly rely on data-driven probability estimates (Sadekar et al., 2024).

Objectives

Therefore, the objectives of the present study were: to develop and validate a binary logistic regression model capable of predicting match outcomes in the 2023 ICC Men's Cricket World Cup using exclusively in-game performance indicators collected from one team's perspective per observation; to quantify the statistical significance and practical effect size of each predictor; and to rank the selected variables—Toss outcome, Opening Partnership Score, Runs and Wickets Lost in Powerplay, Total Number of Fours and Sixes, and Total Wickets Lost in an Inning—according to their relative contribution to match result prediction (Jayalath, 2017).

MATERIALS AND METHODS

Participants

The initial sample comprised all 48 completed matches of the 2023 ICC Men's Cricket World Cup played between 5 October and 19 November 2023 across ten Indian venues. One match (England vs. South Africa) affected by rain and decided via the Duckworth-Lewis-Stern method was excluded because of truncated second innings, leaving 47 matches (94 completed innings) available for analysis. Each innings was treated as an independent observation from the perspective of the batting team.

Study Organization

This was a retrospective, observational analytical study utilizing secondary data extracted from official International Cricket Council (ICC) scorecards and the ESPNcricinfo statistical database (accessed December 2023 – February 2024). Data extraction followed a predefined protocol to ensure consistency and reproducibility. The following seven independent variables were recorded for every innings:

Table 1. Operational Definition of Study Variables

Variable Category	Variable	Type	Measurement Scale	Operational Definition	Coding / Unit
Independent Variables	Toss	Categorical	Nominal (Binary)	Result of the coin toss prior to the match	1 = Won the toss; 0 = Lost the toss
	Opening Partnership Score	Continuous	Ratio	Total runs scored before the fall of the first wicket	Runs
	Runs Scored in Powerplay	Continuous	Ratio	Total runs scored during overs 1–10	Runs
	Wickets Lost in Powerplay	Discrete	Ratio	Number of wickets lost during overs 1–10	0–10 wickets
	Total Number of Fours	Discrete	Ratio	Total boundaries (4 runs) scored in the innings	Count
	Total Number of Sixes	Discrete	Ratio	Total six-hit boundaries scored in the innings	Count
	Total Wickets Lost	Discrete	Ratio	Total wickets lost in the innings (excluding run-outs, where applicable,	Count

				to focus on bowling dismissals)	
Dependent Variable	Match Outcome	Categorical	Nominal (Binary)	Final result of the match from the recorded team's perspective	1 = Win; 0 = Loss

Statistical Analysis

Binary logistic regression was performed to model the probability of winning as a function of the seven independent variables. Forward stepwise selection was used to identify significant predictors. Odds ratios (OR) with 95% confidence intervals were calculated for each variable. Model goodness-of-fit was assessed using the Hosmer-Lemeshow test, and predictive accuracy was evaluated through classification tables and ROC curve analysis. All analyses were conducted using SPSS version 27.0. A p-value < 0.05 was considered statistically significant.

Ethical considerations

This study used only publicly available, secondary match data with no human participants involved. Therefore, ethical approval was not required.

RESULTS

The Hosmer and Lemeshow test was the first output of this logistic regression analysis, as shown in Table 2.

Table 2. Hosmer and Lemeshaw Test

Step	Chi-square	df	Sig.
1	8.473	5	.132
2	5.216	8	.734

The Hosmer-Lemeshow test is applied to determine whether the logistic regression model accurately predicts the dependent variables. This test evaluates the model's goodness of fit. According to Hosmer and Lemeshow, the model's effectiveness is assessed using chi-square statistics, with the model considered highly efficient if the chi-square value is small (Kumar, 2023). For the first and second models, the p-values associated with the chi-square are 0.132 and 0.734, respectively, both of which are greater than 0.05. This indicates that the models are statistically efficient. The third model, with a chi-square value of χ^2 (8) = 5.126 and a p-value of 0.940, also shows that the model is efficient, as the p-value is larger than 0.05 and thus insignificant.

Table 3. Classification Table of Binary Logistic Regression Models: Model 1 (Step 1) Cut-off value = 0.50

Observed Outcome	Predicted Winning	Predicted Losing	Percentage Correct
Winning	35	12	74.5%
Losing	6	41	87.2%
Overall Accuracy			80.9%

Table 4. Classification Table of Binary Logistic Regression Models: Model 2 (Step 2) Cut-off value = 0.50.

Observed Outcome	Predicted Winning	Predicted Losing	Percentage Correct
Winning	38	9	80.9%
Losing	8	39	83.0%
Overall Accuracy			81.9%

The classification table provides a summary of how accurately the logistic regression model predicts cricket match outcomes (Win/Loss). The table is divided into two steps, representing different iterations of the model. In Step 1, the model correctly predicted 35 matches as wins (true positives) and incorrectly predicted 12 matches as losses (false negatives), resulting in a 74.5% accuracy for predicting winning matches. For losing matches, the model correctly predicted 41 matches as losses (true negatives) and incorrectly predicted 6 matches as wins (false positives), achieving an 87.2% accuracy for predicting losing matches. The overall accuracy of the model at this step was 80.9%. In Step 2, the model improved its predictions, correctly identifying 38 winning matches and misclassifying only 9, increasing the accuracy for predicting winning matches to 80.9%. For losing matches, the model correctly predicted 39 matches and misclassified 8, resulting in an 83.0% accuracy for predicting losing matches. The overall model accuracy improved slightly to 81.9%. This indicates that as the model progressed through iterations, its ability to predict match outcomes improved. The cut-off value for classifying a match as a win or loss was set at 0.500.

Table 5. Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	79.382a	.418	.558
2	67.403a	.488	.651

Table 4 provides a summary of the model's performance across two iterations of Binary Logistic Regression. The "Step -2 Log likelihood" values indicate the goodness-of-fit of the model at each step, with lower values representing a better fit. In the first step, the -2 Log likelihood is 79.382, which decreases to 67.403 in the second step, suggesting that the model's fit improved after including additional variables or adjustments.

The "Cox & Snell R Square" and "Nagelkerke R Square" are pseudo-R-squared values that help to explain the proportion of variance in the dependent variable (match outcome: Win/Loss) accounted for by the independent variables. The Cox & Snell R Square in the first step is 0.418, increasing to 0.488 in the second step, indicating that the model explains 41.8% of the variance initially and 48.8% after improvements. Similarly, the Nagelkerke R Square, which adjusts the Cox & Snell R Square to provide a more interpretable value, starts at 0.558 and rises to 0.651, suggesting that the final model explains about 65.1% of the variance in match outcomes.

Table 6. Binary Logistic Regression Analysis Predicting Match Outcome: Model 1 (Step 1): Wickets Lost in an Inning

Variable	B	S.E.	Wald	df	p-value	Exp(B)	Interpretation
Wickets Lost in an Inning	0.833	0.178	22.028	1	< .001	2.300	Each additional wicket lost increases the odds of losing by 2.30 times*
Constant	-6.874	1.597	18.540	1	< .001	0.001	—

Table 7. Binary Logistic Regression Analysis Predicting Match Outcome: Model 2 (Step 2): Addition of Total Number of Fours

Variable	B	S.E.	Wald	df	p-value	Exp(B)	Interpretation
Total Number of Fours	-0.147	0.048	9.358	1	.002	0.863	Each additional four reduces the odds of losing by 13.7%
Wickets Lost in an Inning	0.843	0.174	23.491	1	< .001	2.324	Each additional wicket lost increases the odds of losing by 2.32 times
Constant	-3.445	1.639	4.418	1	.036	0.032	—

Note: Dependent variable coded as: 1 = Loss, 0 = Win (as implied by positive B for wickets lost); Model 1 entered Wickets Lost in an Inning; Model 2 added Total Number of Fours; Statistical significance set at $p < .05$.

The most important table, Table 6 and 7, provides the regression coefficient values (B), which are crucial for formulating the logistic regression equation to predict the dependent variable based on the independent variables. This table also presents the regression coefficients (B), Wald statistics, and odds ratios (Exp(B)) for each variable in two different models. The B coefficients are instrumental in developing a logistic regression equation in terms of log-odds, illustrating the relationship between the independent and dependent variables. The logistic regression equation for Step 2, incorporating "Wickets Lost in an Inning" and "Total No. of 4's in an Inning," is as follows:

Log$\frac{p}{1-p}$ **=** -3.445 +0.843 **(**Wickets Lost in an Inning**)**– 0.147 **(**Total No. of 4's in an inning**)**

Where p is the probability of losing a match (the reference variable), and the dependent variable, also known as the logit(p), is represented as:

Log$\frac{p}{1-p}$

This equation estimates how the independent variables—wickets lost and fours hit—relate to the probability of losing the match, with the dependent variable expressed on a logit scale. The coefficients describe the expected change in the log odds of "losing a match" with a one-unit change in each predictor, assuming all other variables remain constant.

To make these B coefficients more interpretable, they are converted into odds ratios (Exp(B)). These values, presented in the leftmost column of Table No. 4, show the multiplicative change in the odds of the dependent variable occurring (i.e., losing the match) with each unit increase in the predictor variable.

In this model, Wickets Lost in an Inning emerges as the most significant predictor, with an odds ratio of Exp(B) = 2.324. This means that for each additional wicket lost, the odds of losing the match increase by approximately 2.324 times. The Wald statistic for this variable is highly significant (Wald = 23.491, p = 0.000), reinforcing its importance. Since the odds ratio is greater than one, it indicates that increasing the number of wickets lost increases the likelihood of losing the match. Specifically, if the number of wickets lost increases by one unit, the likelihood of losing increases by 132.4% (2.324 - 1 = 1.324).

On the other hand, the Total No. of 4's in an Inning has an odds ratio of Exp(B) = 0.863. The negative coefficient (-0.147) and the odds ratio less than one indicate that hitting more boundaries decreases the likelihood of losing the match. With each additional four hit, the odds of losing decrease by approximately 13.7% (1.00 - 0.863 = 0.137). The Wald statistic for this variable is also significant (Wald = 9.358, p = 0.002), meaning it has a meaningful influence on the match outcome.

Teams that managed to keep wickets intact during the powerplay were more likely to post a strong total or chase down targets. This finding aligns with previous research by Shao et al. (2025), which suggests that preserving wickets during the initial overs is crucial for ensuring stability throughout the innings. The significant influence of wickets lost in the powerplay highlights the importance of a balanced approach during this phase, where teams must weigh the risks of aggressive play against the need for stability.

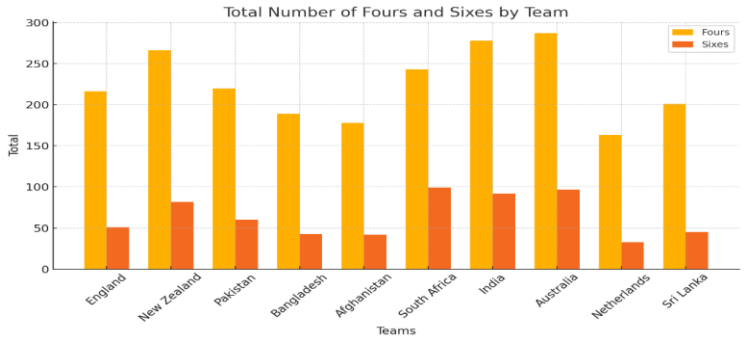

Figure 1. Total number of Four's and Sixes by Team

Overall, the analysis shows that Wickets Lost in an Inning is the most impactful predictor of match results, as losing more wickets significantly increases the probability of losing. On the contrary, hitting more boundaries (fours) decreases the chances of losing. These findings underscore the importance of maintaining wickets and scoring boundaries to influence match outcomes positively.

Teams that relied too heavily on clearing the boundary without a solid base, such as England in several matches, often collapsed under pressure (Singh et al., 2006). While sixes can quickly change the momentum of a game, they are not as reliable a predictor as other variables such as wickets lost and fours hit. Teams that balanced aggression with caution, like Australia, were more successful in maintaining consistency across their innings.

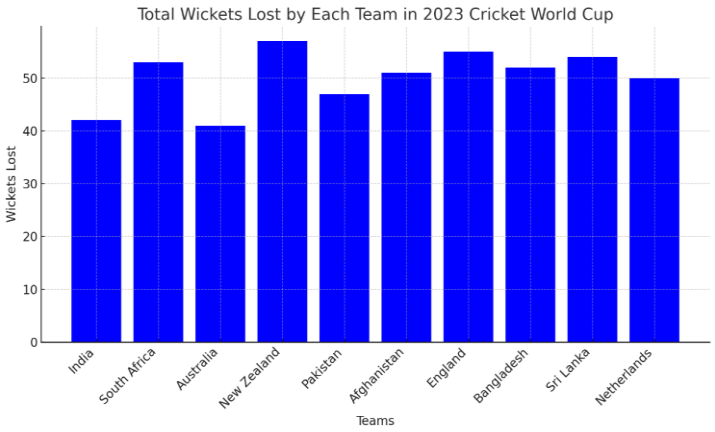

Figure 2. Total wickets lost by each team in 2023 cricket world cup

The number of wickets lost in an inning has proven to be one of the most significant predictors of a team's success in the 2023 Cricket World Cup. As data illustrated in Fig. No. 2, New Zealand, Pakistan, and Bangladesh were among the teams that lost over 50 wickets throughout the tournament. This higher number of wickets lost likely hindered their ability to consistently build partnerships and capitalize on crucial overs, including the death overs.

.

DISCUSSION

Cricket has always been a dynamic and unpredictable game, especially in formats like One Day Internationals (ODIs) where teams have to balance aggression with caution. The 2023 ICC Cricket World Cup exemplified this complexity, and our study delves into the key variables that determined match outcomes using a logistic regression model. In this discussion, we not only focus on wickets lost and boundary hitting but also examine other important variables such as the toss, opening partnership score, runs scored in powerplay, and wickets lost in powerplay.

One of the first variables analyzed in our study was the toss, which has long been seen as a crucial aspect of match strategy in cricket. Winning the toss allows teams to make key decisions regarding whether to bat or bowl first, based on weather, pitch conditions, and team composition. In the 2023 World Cup, several matches were influenced by the toss. For example, *The* (McEwan et al., 2023) highlighted how teams winning the toss in day-night games often chose to bowl first, taking advantage of the dew factor to make chasing easier under lights. However, our study did not find the toss to be a significant predictor of match outcomes in isolation. The results indicate that while winning the toss can give a team an initial advantage, it does not directly correlate with match success unless it is combined with other performance factors such as wickets and runs scored. This suggests that while the toss might shape early strategy, its impact on the outcome diminishes as the match progresses.

The opening partnership score, which reflects the number of runs scored by the first two batsmen, is another key factor in shaping the innings. Strong opening partnerships lay the foundation for a big total, as seen in the 2023 World Cup where successful teams consistently relied on stable starts. In India's dominant group-stage performances, Rohit Sharma and Shubman Gill often provided solid starts, allowing the middle order to build on their platform. However, when this foundation faltered, as in the final against Australia where India lost early wickets, the team struggled to post a competitive total (Sadekar et al., 2024; Sanjaykumar et al., 2024) . This study also indicated that while the opening partnership score is important, it was not as strong a predictor as wickets lost later in the innings. The reason is that even if an opening partnership is strong, the loss of wickets in middle and death overs can derail the momentum. Therefore, while a high opening partnership score is desirable, it must be complemented by stability throughout the innings for a team to secure a win.

The powerplay overs (the first ten overs in an ODI) offer teams a unique opportunity to score freely due to fielding restrictions. In the 2023 World Cup, teams that scored aggressively in the powerplay often set the tone for the rest of their innings. For instance, England's aggressive approach in the powerplay against South Africa resulted in a high run rate, putting immense pressure on the bowlers and setting the stage for a high total (Khan & Sarkar, 2025). In this study researcher found that runs scored in the powerplay have a moderate influence on match outcomes. Teams that scored more runs during this phase were better positioned to build a strong total. However, the predictive power of this variable was somewhat limited, as teams that lost wickets in the powerplay often struggled

later in the innings. This suggests that while scoring runs quickly during the powerplay is beneficial, it must be balanced with wicket preservation.

Losing wickets in the powerplay can be detrimental to a team's chances of building a competitive total. The 2023 World Cup saw several instances where teams that lost early wickets were unable to recover. For example, Mullick (2024)& Austin et al. (2025) reported on England's early collapse against India in the group stages, where they lost key wickets in the powerplay and never recovered, ultimately posting a subpar total. In this study, wickets lost in the powerplay emerged as one of the stronger predictors of match outcomes.

The data on boundaries (fours and sixes) from the 2023 Cricket World Cup provides crucial insights into the strategies adopted by teams for accumulating runs. Teams like India, Australia, and New Zealand led in overall boundary counts, with India and Australia particularly excelling in hitting both fours and sixes. This data (Ref. Fig No. 1) reflects India's dominant performance during the group stages, where their ability to rotate strike and find regular boundaries allowed them to consistently build competitive totals without taking excessive risks (Jaiswal & Srivastava, 2025).

A closer look at the graph indicates that teams such as South Africa and Australia not only accumulated a significant number of fours but also hit a considerable number of sixes, showcasing their ability to accelerate the run rate during critical phases of the game. This combination of hitting both types of boundaries highlights a balanced approach, particularly in the middle and death overs. Conversely, teams like Pakistan and Bangladesh, which hit fewer sixes, were less capable of pushing the run rate when required. This supports previous research findings that teams who hit more boundaries, especially fours, have a higher probability of winning matches. In fact, our study found that the number of fours hit has a significant influence on match outcomes, with an odds ratio of 0.863, indicating that for every additional boundary hit, the chances of losing decrease (Nicholls et al., 2023; Roberts et al., 2023).

However, it is important to balance aggressive boundary hitting with wicket preservation. As (Jamil et al., 2023) noted, teams that rely too heavily on boundaries without keeping wickets in hand risk collapse, as seen in matches where aggressive batting strategies backfired after the loss of key players. This was evident in teams like Sri Lanka and Netherlands, which struggled to maintain competitive totals due to their lower boundary counts, both in terms of fours and sixes.

The presence of six-hitting capabilities, as seen in teams like India and South Africa, provides a clear advantage during the middle and death overs, where acceleration becomes crucial. The data from the World Cup supports the assertion that a well-balanced approach—combining boundary hitting with smart strike rotation—helps teams build and maintain a competitive edge.

In conclusion, the analysis of boundary counts reaffirms the findings of our study. Teams that excel in hitting boundaries, while simultaneously preserving wickets, have a greater chance of posting competitive totals or chasing down targets successfully. This trend underscores the evolving nature of modern ODIs, where boundary-hitting prowess must be coupled with strategic play to maximize the chances of success (Jain et al., 2024; Lohawala & Rahman, 2018).

The number of sixes hit during an inning reflects a team's aggressive approach, aiming to maximize scoring in limited overs. In the 2023 World Cup, teams like England and South Africa adopted this approach, regularly clearing the boundary to post formidable totals. However, this study found that the number of sixes, while valuable, did not emerge as a significant predictor of match outcomes. The reason could be that six-hitting often involves greater risk, leading to more wicket losses.

Teams like India and Australia, on the other hand, lost fewer wickets overall (around 40-45), which allowed them to stabilize their innings, especially during the middle overs, and effectively capitalize on the final overs. Australia's World Cup-winning campaign, in particular, stands out for their ability to preserve wickets, especially in high-pressure situations, leading to strong finishes in their matches (Budhrani, 2024). This ability to maintain a steady flow of partnerships throughout the innings was a key factor in their success.

Our study supports these observations, demonstrating that wickets lost was the strongest predictor of match outcomes, with an odds ratio of 2.324. This means that for every additional wicket lost, the likelihood of losing the match increases significantly. This finding further underscores the importance of preserving wickets, as it directly correlates to a team's ability to build partnerships, maintain a competitive run rate, and effectively use the final overs.

For teams like New Zealand and Pakistan, the high number of wickets lost meant they often struggled to capitalize on powerplay runs and death-over accelerations, leading to several underwhelming performances. Meanwhile, teams like Afghanistan and England, which also saw a high number of wickets lost, faced similar issues in maintaining stability throughout their innings, contributing to their inconsistent performances during the tournament.

In conclusion, the 2023 World Cup data reaffirms the critical role that wicket preservation plays in determining match outcomes. Teams that can keep wickets intact, particularly through the middle overs, are better positioned to post competitive totals or successfully chase targets. The significant odds ratio observed in our study highlights the need for teams to focus on building long-lasting partnerships and minimizing unnecessary dismissals throughout the innings (Brown, 2017; Jamil et al., 2023).

CONCLUSION

This study aimed to develop a predictive model for cricket match outcomes based on key performance indicators from the 2023 ICC Cricket World Cup. Using a logistic regression framework, we analyzed variables including Toss, Opening Partnership Score, Runs Scored in Powerplay, Wickets Lost in Powerplay, Total Number of Fours, Total Number of Sixes, and Wickets Lost in an Inning to determine their influence on match results. The findings of this research offer several important insights into the dynamics of One Day International (ODI) cricket and highlight the critical role of specific game phases in determining success. One of the most significant conclusions from the analysis is that Wickets Lost in an Inning is the strongest predictor of match outcomes. The logistic regression model revealed an odds ratio of 2.324 for this variable, meaning that for each additional wicket lost, the likelihood of losing the match increases by approximately 132.4%. This result underscores the importance of wicket preservation in ODI cricket, particularly as teams approach the middle and death overs, where maintaining partnerships and stabilizing the innings can be crucial to posting a competitive

total or successfully chasing a target. Teams that lost fewer wickets, such as India and Australia, were better positioned to capitalize on scoring opportunities, demonstrating the need for a balanced approach between aggression and wicket retention.

The analysis also highlighted the Total Number of Fours in an Inning as another key variable, with an odds ratio of 0.863. This suggests that hitting more boundaries reduces the likelihood of losing the match, as each additional four decreases the odds of losing by 13.7%. Boundary hitting, especially fours, reflects a team's ability to accumulate runs steadily while minimizing risk, providing an advantage in building pressure on the opposition. Teams like India, South Africa, and Australia, which consistently scored boundaries while preserving wickets, saw greater success in the tournament.

Interestingly, the Toss, which is often perceived as a decisive factor in cricket, did not emerge as a significant predictor of match outcomes in this study. While winning the toss can influence strategic decisions like whether to bat or bowl first, the data suggests that the toss alone does not guarantee success. Its impact diminishes over the course of the game, as other factors such as powerplay performance, wicket preservation, and boundary scoring play a more significant role in determining the final result. This finding aligns with prior research, suggesting that the toss can shape early strategy but does not dictate match outcomes unless supported by strong in-game performances.

Powerplay performance—both in terms of runs scored and wickets lost—was another focus of this study. Teams that scored freely during the powerplay without losing wickets were more likely to establish a strong foundation for the rest of their innings. However, teams that lost early wickets in this phase often struggled to recover, as seen in several matches throughout the tournament. The moderate influence of runs scored in the powerplay suggests that while it is important to capitalize on fielding restrictions during the first ten overs, excessive risk-taking without considering wicket preservation can backfire, leading to collapses later in the innings.

The study also examined the impact of Total Number of Sixes, which, while reflecting a team's aggressive intent, did not emerge as a significant predictor of match outcomes. The reason may lie in the greater risk associated with six-hitting, which often leads to higher wicket losses. Teams that relied too heavily on clearing the boundary, such as England in several matches, frequently collapsed under pressure when wickets fell quickly, reducing the overall reliability of this strategy.

Overall, the logistic regression model developed in this study proved to be a statistically efficient predictor of match outcomes, as confirmed by the Hosmer and Lemeshow test. The model's findings indicate that Wickets Lost in an Inning and Total Number of Fours are the most significant factors in predicting the outcome of a cricket match. Other variables, such as the Toss and Powerplay performance, while influential, did not carry the same predictive weight. This underscores the importance of a comprehensive approach in cricket, where teams must balance boundary scoring with wicket retention to optimize their chances of success. The implications of this study are valuable for both cricket analysts and teams. From an analytical perspective, the findings highlight the need to focus on micro-level variables—such as specific phases of play and key game events—instead of relying solely on broader indicators like team strength or historical records. For teams and coaches, the study emphasizes the importance of building strategies that prioritize wicket preservation and consistent boundary hitting, particularly in high-pressure situations. Teams that can maintain this balance are more likely to achieve consistent success in ODI cricket, as demonstrated by the top-performing teams in the 2023 ICC Cricket World Cup. Moreover, the predictive model developed in this study has practical applications for decision-making during matches. Teams can use the insights from this model to adapt their strategies dynamically, focusing on minimizing wicket losses in critical phases and maximizing boundary opportunities without compromising stability. Additionally, analysts can apply similar models to future tournaments, improving the accuracy of match predictions and helping teams fine-tune their game plans.

In conclusion, this study contributes to the growing body of literature on cricket analytics by identifying the key variables that influence match outcomes and developing a robust predictive model based on data from the 2023 ICC Cricket World Cup. The insights gained from this research have the potential to transform how teams approach strategy and performance analysis in ODI cricket, emphasizing the importance of balancing aggressive play with smart, data-driven decision-making.

ACKNOWLEDGMENTS

Researchers extend their sincere gratitude to the Department of Physical Education at Government College Chinnor, LNIPE Gwalior, and the Sports University of Haryana for their unwavering support and facilities which greatly contributed to the research.

CONFLICT OF INTEREST

Author declares no conflict of interest.

REFERENCES

Akl, A., & Hasan, R. (2025). A Comprehensive Survey on Pipeline Monitoring Technologies: Advancements, Challenges, Market Opportunities and Future Directions. *Journal of Pipeline Science and Engineering*, 100353. https://doi.org/10.1016/j.jpse.2025.100353

Allsopp, P. E., & Clarke, S. R. (2004). Rating Teams and Analysing Outcomes in One-Day and Test Cricket. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *167*(4), 657. https://doi.org/10.1111/j.1467-985x.2004.00505.x

Austin, A., Rahman, I. U., & Rana, Z. (2025). Understanding the ability of households to cope with economic shocks: an empirical study of Pakistan during the COVID-19 pandemic. *Journal of Health, Population and Nutrition*, *44*, 80. https://doi.org/10.1186/s41043-025-00772-y

Bandyopadhyay, R., & Bhattacharjee, D. (2025). Applications of higher order Markov models and Pressure Index to strategize controlled run chases in Twenty20 cricket. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2505.01849

Brown, P. (2017). *Optimising Batting Partnership Strategy in the First Innings of a Limited Overs Cricket Match*. https://doi.org/10.26686/wgtn.17065379

Budhrani, T. (2024). Excellence in Cricket leveraging AI/ML and Automation. *International Journal for Research in Applied Science and Engineering Technology*, *12*(1), 509. https://doi.org/10.22214/ijraset.2024.57992

Burra, S. L. (2023). Using Logistic Regression for Cricket Analysis: Predicting Match Outcomes and Player Performances. *International Journal for Research in Applied Science and Engineering Technology*, *11*(4), 2963. https://doi.org/10.22214/ijraset.2023.50839

I.Y, A., J.O, U., Ito, S., & M.E, N. (2025). Development of An Ai-Driven Adaptive Learning Management System Using Data Analytics. *International Journal of Scientific Research in Modern Science and Technology*, *4*(7), 52. https://doi.org/10.59828/ijsrmst.v4i7.349

Jain, V., Banerjee, S., & Kothari, P. (2024). Trade-off between fitness and experience: insights from historical data on ODI cricket. *Frontiers in Behavioral Economics*, *3*. https://doi.org/10.3389/frbhe.2024.1492735

Jaiswal, A., & Srivastava, N. (2025). Posture-Driven Action Intent Inference for Playing style and Fatigue Assessment. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2507.11642

Jamil, M., Kerruish, S., Mehta, S., Phatak, A., Memmert, D., & McRobert, A. P. (2023). Identifying which factors impact bowling and batting performances during the “death” phase of an innings in international men’s 50-over cricket. *International Journal of Performance Analysis in Sport*, *23*(2), 111. https://doi.org/10.1080/24748668.2023.2193030

Jayalath, K. P. (2017). A machine learning approach to analyze ODI cricket predictors. *Journal of Sports Analytics*, *4*(1), 73. https://doi.org/10.3233/jsa-17175

Khan, Z., & Sarkar, I. (2025). Efficacy of Outcome-Based Education in Enhancing English Proficiency Among Engineering Students in Hyderabad. *Journal of Engineering Education/Journal of Engineering Education Transformations/Journal of Engineering Education Transformation*, *38*, 445. https://doi.org/10.16920/jeet/2025/v38is2/25054

Kumar, A. (2023). Non-Invasive estimation of muscle fiber type using ultra Sonography. *International Journal of Physical Education Sports and Health*, *10*(1), 89. https://doi.org/10.22271/kheljournal.2023.v10.i1b.2758

Lansiaux, É., Azzouz, R., Chazard, E., Vromant, A., & Wiel, É. (2025). *Development and Comparative Evaluation of Three Artificial Intelligence Models (NLP, LLM, JEPA) for Predicting Triage in Emergency Departments: A 7-Month Retrospective Proof-of-Concept*. 1. https://doi.org/10.1145/3773276.3774289

Lohawala, N., & Rahman, M. A. (2018). Are strategies for success different in test cricket and one-day internationals? Evidence from England-Australia rivalry. *Journal of Sports Analytics*, *4*(3), 175. https://doi.org/10.3233/jsa-180191

Luo, X., Chen, L., Jing, C., Yang, Y., & Lee, S. (2025). Exploring perceptual differences and constraints in esports among American university students. *Frontiers in Sports and Active Living*, *7*, 1669960. https://doi.org/10.3389/fspor.2025.1669960

McEwan, K., Pote, L., Nicholls, S., Radloff, S. E., & Christie, C. J. (2023). The role of selected pre-match covariates on the outcome of One-day International (ODI) cricket matches. *South African Journal of Sports Medicine*, *35*(1). https://doi.org/10.17159/2078-516x/2023/v35i1a15012

Mullick, P. (2024). Statistical Evolution of ODI Cricket: Analyzing Performance Trends and Effect Sizes. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.11652

Nicholls, S., Pote, L., Thomson, E., & Theis, N. (2023). The Change in Test Cricket Performance Following the Introduction of T20 Cricket. *Sports Innovation Journal*, *4*, 1. https://doi.org/10.18060/26438

Ozay, D., Jahanbakth, M., & Wang, S. (2025). Exploring the Intersection of Big Data and AI With CRM Through Descriptive, Network, and Contextual Methods. *IEEE Access*, *13*, 57223. https://doi.org/10.1109/access.2025.3554549

Pedarla, H. K. (2025). Encrypted Traffic Analytics (ETA): Machine Learning Approaches for Intrusion Detection Without Decryption. *International Journal on Science and Technology*, *16*(4). https://doi.org/10.71097/ijsat.v16.i4.9559

Puram, P., Roy, S., Srivastav, D., & Gurumurthy, A. (2022). Understanding the effect of contextual factors and decision making on team performance in Twenty20 cricket: an interpretable machine learning approach. *Annals of Operations Research*, *325*(1), 261. https://doi.org/10.1007/s10479-022-05027-1

Roberts, L., Little, D. R., Jackson, M., & Spittal, M. J. (2023). Test cricketers score quickly during the ‘nervous nineties’: Evidence from a regression discontinuity design. *PLoS ONE*, *18*(6). https://doi.org/10.1371/journal.pone.0287700

Sadekar, O., Chowdhary, S., Santhanam, M. S., & Battiston, F. (2024). Individual and team performance in cricket. *Royal Society Open Science*, *11*(7), 240809. https://doi.org/10.1098/rsos.240809

Sanjaykumar, S., Udaichi, K., Rajendiran, G., Creţu, M., & Козіна, Ж. (2024). Cricket performance predictions: a comparative analysis of machine learning models for predicting cricket player’s performance in the One Day International (ODI) world cup 2023. *Health Sport Rehabilitation*, *10*(1), 6. https://doi.org/10.58962/hsr.2024.10.1.6-19

Shao, Y., Wang, L., Jin, H., & Ye, H. (2025). The motivational factor of working to involve outdoor sport as mediated by health awareness and social influence: a case study for China and Malaysia. *BMC Public Health*, *25*(1), 1594. https://doi.org/10.1186/s12889-025-22661-z

Singh, G., Irfan, M., & Sasi, S. (2006). Real-time boundary detection for cricket game. *Australasian Conference on Interactive Entertainment*, 23. https://doi.org/10.5555/1231894.1231898

Sravani, K. G., Sricharitha, A., Jarpula, M., Radhakrishn, E., & Mr.B.Saida. (2025). Myocardial Infarction Risk Estimation Using Machine Learning. *International Research Journal on Advanced Engineering and Management (IRJAEM)*, *3*(5), 1794. https://doi.org/10.47392/irjaem.2025.0283

Tandililing, P., Napitupulu, B., Imawan, O. R., & Ismail, R. (2025). Predictive model of mathematical literacy ability, specifically university students' critical thinking skills: A case study on the inappropriateness of multilevel regression. *Journal of Education and E-Learning Research*, *12*(3), 507. https://doi.org/10.20448/jeelr.v12i3.7398

Vora, S. (2025). Cricket Data Analytics: Building the Best XI Players from T20 World Cup. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5180123