A Markov Decision Process (MDP)-Based Q-Learning Framework for Optimizing Bowling Strategies in T20 Cricket

Abdurrahman Sabir[1*], Muhammad Irfan Ud Din[2], Syed Habib Shah[3], Gohar Ayub[4], Mujeeb[1], Shahid Iqbal[5], Qamruz Zaman[1]

1. Department of Statistics, University of Peshawar
2. Department of Computer Science, University of Swabi
3. Institute of Numerical Sciences, Kohat University of Science and Technology
4. Department of Mathematics and Statistics, University of Swat
5. CDPM/IER/Assistant DAS University of Peshawar. Email: shahidiqbalkhan@uop.edu.pk

Corresponding Author*: abdurrahmansabir27@gmail.com

Abstracts

Optimizing bowling strategies in T20 cricket is a complex sequential decision-making problem due to the dynamic, stochastic, and context-dependent nature of the game. Traditional approaches rely on expert judgment, intuition, or aggregated performance statistics, which fail to capture ball-by-ball tactical adaptability and phase-specific strategic requirements. This study develops a Markov Decision Process (MDP)-based Q-learning framework to optimize bowling strategies in T20 cricket using real ball-by-ball data. The framework models each delivery as a state transition, defining states by match phase (Powerplay, Middle over, Death over), delivery length, and bowling line, with actions representing bowling variations including yorkers, slower balls, bouncers, and spin deliveries. The Q-learning algorithm learns optimal policies directly from historical match data without requiring pre-specified transition probabilities, and a reward function based on runs conceded and wickets taken guides policy optimization.

The methodology comprises four stages: data collection and preprocessing from nineteen T20I bowlers using ball-by-ball records from January to December 2024, state-action space design with 105 possible states, MDP formulation with context-aware reward engineering, and Q-learning policy optimization followed by hybrid bowler ranking. Results demonstrate that optimal bowling strategies are highly context-dependent and bowler-specific. In the Powerplay, Nathan Ellis (Mean-Q = 18.86), Josh Hazlewood (18.01), and Arshdeep Singh (16.87) achieved the highest Q-values through disciplined line-length combinations and swing variations. During the middle overs, spin bowlers dominated with Adil Rashid (26.03), Rashid Khan (23.78), and Axar Patel (23.27) recording the highest Mean Q-values. In Death overs, Rashid Khan (30.33), Gudakesh Motie (30.28), Arshdeep Singh (24.67), and Jofra Archer (21.59) demonstrated exceptional effectiveness through yorkers, slower balls, and variation-based strategies. The hybrid ranking model integrates classical metrics (wickets, economy rate, strike rate, bowling average, dot ball percentage, boundary percentage) with normalized Q-scores using a 65:35 weighting scheme. Arshdeep Singh ranked first overall (0.7844) with the highest classical sum (0.6683) and maximum Q-score (1.000), followed by Rashid Khan (0.7817) and Gudakesh Motie (0.6935). Adil Rashid benefited most from RL inclusion, rising from 10th in classical sum to 6th in hybrid rank due to a strong Q-score (0.7805), while Anrich Nortje showed the largest divergence with a respectable classical sum (0.5105) but very low Q-score (0.0793), dropping to 16th. The findings confirm that strategic intelligence measured via Q-learning reveals value not captured by traditional statistics, advocating for reinforcement learning-based evaluation in cricket analytics.

Keywords: Reinforcement Learning, Q-learning, Markov Decision Process, T20 Cricket, Bowling Strategy Optimization, Hybrid Ranking, Sports Analytics.

Introduction

In cricket, player selection is a crucial task for both national teams and franchise leagues. Traditionally, selectors rely on their experience and domain knowledge to evaluate players' physical fitness, batting ability, and bowling performance [1]. Player rankings are also of major importance to sports authorities, athletes, and fans, as they influence team composition, tournament strategies, and commercial value. With the increasing commercialization of sports, ranking systems have gained even greater significance for investors and other stakeholders. In this context, Premkumar et al. introduced new variables and refined existing performance measures based on specific Key Performance Indicators (KPIs) to improve player evaluation in cricket. Their work highlighted several overlooked variables absent from earlier ranking models, including the widely used International Cricket Council ranking system [2].

While ranking systems are essential for evaluating player quality, modern cricket increasingly requires optimization beyond static performance measurement. In particular, the rapid and dynamic nature of T20 cricket demands real-time strategic decision-making, where captains and bowlers must continuously adapt bowling strategies and field placements according to match conditions. Since each delivery directly affects runs, wickets, and remaining balls, this problem naturally aligns with sequential decision-making frameworks such as Markov Decision Processes (MDPs) and Reinforcement Learning (RL).

Within this setting, each ball can be modeled as a state transition, making Markov models highly suitable for representing the stochastic and sequential structure of cricket. Reinforcement Learning has already been applied in sports analytics to model batting strategies, optimize run chases, and simulate match scenarios [3]–[6]. Among RL methods, Q-learning is particularly attractive because it is a model-free algorithm capable of learning optimal policies directly from interaction data without requiring explicit transition probabilities, which are often unavailable in complex sports environments.

Building on these developments, the optimization of bowling strategies in T20 cricket has become increasingly data-driven through advancements in artificial intelligence (AI), machine learning, and reinforcement learning. In this context, MDPs provide a strong mathematical foundation for modeling sequential decisions under uncertainty, which is highly relevant in sports analytics where outcomes depend on chains of interdependent actions [7]–[9].

More generally, MDPs are widely recognized as fundamental tools for modeling sequential decision-making problems across large-scale real-world systems [10]–[12]. When system dynamics are fully observable, MDPs can be solved using dynamic programming methods [13]. However, in many practical domains—including cricket—the environment is only partially understood, and transition dynamics are difficult to estimate explicitly. Consequently, traditional dynamic programming approaches become infeasible, motivating the use of model-free RL algorithms such as Q-learning to approximate value functions and derive optimal policies directly from observed interactions [14].

In this research, Reinforcement Learning is employed as a machine learning framework positioned between supervised and unsupervised learning. RL enables an agent to learn optimal sequential decisions from delayed and limited feedback within a formal MDP structure, where states, actions, rewards, policies, and value functions collectively determine behavior. These methods have been extensively studied in optimization and control problems, particularly in uncertain and stochastic environments, making them highly suitable for sports analytics.

Recent literature further demonstrates the growing use of MDPs and Q-learning in sports strategy optimization, including cricket. These approaches allow analysts to process large-scale ball-by-ball datasets, model player behavior, and simulate tactical scenarios to identify strategies that optimize objectives such as minimizing runs conceded or maximizing wicket-taking opportunities [15][16].

Although foundational research has primarily focused on RL theory and applications in sports such as soccer and badminton [17] [18], an emerging body of work has begun addressing cricket-specific challenges, including contextual field placements, bowler–batsman matchups, and phase-specific strategies [19] [20].

This study aligns with a broader research trend that utilizes Markov models and reinforcement learning to optimize sequential decisions in sports. For example, in soccer, offensive behavior has been modeled as an MDP learned from event data and analyzed using probabilistic model checking to evaluate shot selection and optimize defensive responses [21]. Similarly, time-dependent MDP formulations have been used in basketball to model plays under shot-clock constraints, simulate seasons, and evaluate alternative decision policies [22]. These studies collectively demonstrate the effectiveness of Markov frameworks in uncovering strategy efficiency from sequential sports data.

More broadly, reinforcement learning—particularly Q-learning—has been widely adopted for strategy optimization in games and sport-like environments because it learns action values directly from experience without requiring explicit environmental models. Existing surveys categorize Q-learning variants, including deep and multi-agent extensions, and highlight their suitability for complex, high-dimensional decision problems [23], [24]. In sports contexts, RL methods such as Q-learning and SARSA have also demonstrated improved accuracy and faster convergence compared with Monte Carlo approaches when optimizing dynamic strategies in simulated environments [25].

Specialized sport-specific MDP formulations further reinforce the applicability of this approach. For example, Sport Strategy Optimization MDPs in beach volleyball explicitly maximize win probability under action failure risk while providing two-scale (coarse-to-fine) recommendation models for coaches [26], [27]. These studies indicate that MDP-based optimization can effectively bridge theoretical decision models and practical coaching strategies.

Parallel to these sequential decision frameworks, cricket analytics has also advanced through machine learning-based performance indices and ranking systems for T20 players. Recent studies on role-aware deep performance indices and network-based ranking models for Indian Premier League batsmen demonstrate the value of integrating multi-metric statistical evaluation with learned models for strategic planning [28], [29]. Therefore, a hybrid bowling score that combines normalized traditional bowling metrics with an RL-derived Q-score aligns naturally with two complementary research directions: (i) Markov and reinforcement learning approaches for strategy optimization in sports, and (ii) multi-criteria, data-driven player performance benchmarking in cricket and related domains [30]–[32].

A research gap exists in applying Q-learning to optimize T20 bowling strategies using real ball-by-ball data. No prior study has defined a match-phase-aware MDP, learned delivery-type policies from real data, evaluated strategies separately across powerplay/middle/death overs, or integrated Q-scores with traditional metrics. This study addresses the gap by proposing an MDP-based Q-learning framework that models each delivery as a state transition and learns context-aware bowling policies from historical match data.

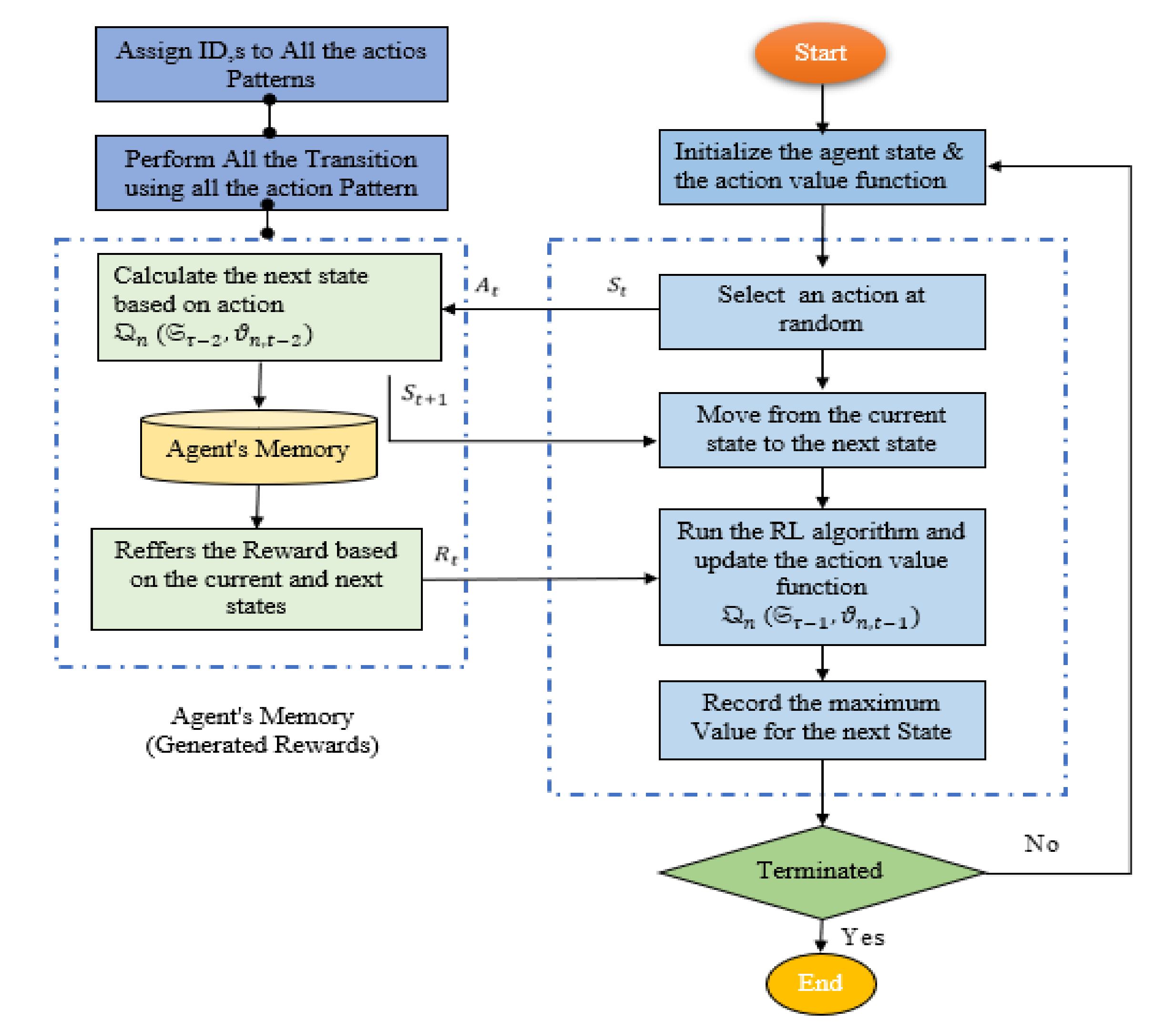


Figure # 1 Flowchart of the Proposed MDP-Based Q-Learning Framework for Bowling Strategy Optimization

Significance of the Study

This study bridges traditional cricket analytics with AI-based decision-making by applying MDPs and Q-learning to T20 bowling optimization. Unlike expert judgment or aggregated statistics, the framework enables data-driven, ball-by-ball strategy adaptation to match conditions. Practically, it equips coaches with phase-specific bowling recommendations and a hybrid Q-score for player evaluation. Academically, it extends RL methodologies to a real-world cricket environment—an area previously limited to conceptual studies—contributing to the growing literature on sequential decision-making in sports analytics.

Aim and Objectives:

The main objective of this study is to develop an MDP-based Q-learning framework for optimizing T20 bowling strategies using real ball-by-ball data. Specifically, the study aims to:

- Formalize T20 bowling as a sequential decision-making problem using the MDP framework by defining states, actions, and rewards.

- Learn optimal bowling policies by applying the Q-learning algorithm directly to historical ball-by-ball data.
- Evaluate the effectiveness of learned bowling policies across powerplay, middle overs, and death overs.
- Integrate traditional bowling performance metrics with RL-derived Q-scores into a hybrid bowler ranking framework.

3. Methodology

3.1. Research Design

This study adopts a quantitative and computational research design to optimize bowling strategies in T20 cricket using Reinforcement Learning (RL), specifically the Q-learning algorithm within a Markov Decision Process (MDP) framework. The proposed framework models bowling as a sequential decision-making problem where each delivery represents a transition between match states. Historical ball-by-ball cricket data is used to train the RL agent and evaluate optimal bowling strategies under different match conditions. The research framework models bowling as a sequential decision-making problem where each delivery represents a transition from one game state to another. The objective is to learn optimal bowling actions that maximize wicket-taking opportunities while minimizing runs conceded under varying match contexts.

The methodology consists of four major stages: data collection and preprocessing, state-action space design, MDP formulation with reward engineering, and Q-learning policy optimization, followed by hybrid bowler ranking.

3.2. Dataset Overview

The dataset comprises ball-by-ball records from T20 International matches, including bowler name, delivery variation, line, length, match phase (Powerplay, Middle, Death), and outcome (dot ball, run, boundary, wicket). Nineteen T20I bowlers (both pace and spin) were selected, with approximately 30% of their matches from January to December 2024 randomly sampled to ensure balance and reduce overfitting.

3.3. Data Preprocessing

Raw data were cleaned by removing incomplete records. Categorical variables (bowling type, line, length, phase) were encoded, and performance metrics (economy rate, strike rate, bowling average, dot ball percentage, boundary percentage) were derived for player ranking.

3.4. Classical Bowling Performance Formulas

3.4.1. Economy Rate (ER)

Measures the average number of runs conceded per over.

$$Economy\ Rate = \frac{Total\ Runs\ Conceded}{Total\ Overs\ Bowled}$$

3.4.2. Bowling Strike Rate (BSR)

Measures the average number of balls required to take one wicket.

$$Bowling\ Strike\ Rate = \frac{Total\ Balls\ Bowled}{Total\ Wickets\ Taken}$$

3.4.3. Bowling Average (BA)

Measures the average runs conceded per wicket taken.

$$Bowling\ Average = \frac{Total\ Runs\ Conceded}{Total\ Wickets\ Taken}$$

3.4.4. Dot Ball Percentage (DB%)

Represents the percentage of deliveries on which no runs are scored.

$$Dot\ Ball\ Percentage = \left(\frac{Total\ Dot\ Balls}{Total\ Balls\ Bowled}\right) \times 100$$

3.4.5. Boundary Percentage (BP%)

Represents the percentage of deliveries resulting in boundaries (4s or 6s).

$$Boundary\ Percentage = \left(\frac{Total\ Boundary\ Balls}{Total\ Balls\ Bowled}\right) \times 100$$

3.5. Markov Decision Processes (MDPs) and Reinforcement Learning Framework

A Markov Decision Process (MDP) is a mathematical framework used to model sequential decision-making problems in which outcomes are partly random and partly controlled by a decision-maker (or agent). The framework was formally introduced by Richard Bellman (1957) and provides the theoretical foundation of reinforcement learning [33] [35].

In the context of cricket analytics, the agent represents the bowler, who makes decisions regarding bowling actions such as line, length, and variation. The environment represents the match situation, including batter behavior, pitch conditions, and game context. At each time step (i.e., each delivery), the agent interacts with the environment by selecting an action, which influences the next state of the game.

3.6. Reward Function

Table # 2: Reward Function $\mathbf{R(outcome, phase)}$

Outcome	Base R	Powerplay +Bonus	Deathover +Bonus	Final R
Wicket (Out)	+15	+3 (=+18)	+5 (=+20)	Max +20
Dot Ball	+3	+1 (=+4)	+2 (=+5)	Max +5
1 / 2 / 3 runs	−3	—	—	−3
Boundary (Four)	−6	—	—	−6
Six	−10	—	—	−10
Extra / Wide	−2	—	—	−2

3.7. Policy and Objective

The objective of a Markov Decision Process (MDP) agent is to learn a policy $\pi: S \rightarrow A$, which maps each state to an action to maximize cumulative reward over time. Policies can be deterministic, where a single action is assigned to each state ($\pi(s) = a$), or stochastic, where a probability distribution over actions is defined for a given state ($\pi(a \mid s) = P(a \mid s)$). The optimal policy $\pi^*$is one that maximizes the expected discounted return, defined as the cumulative sum of future rewards: $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, where $\gamma$is the discount factor controlling the importance of future rewards.

3.8. Value Functions

Value functions are used to estimate the expected return of states and actions under a given policy. The state-value function, $V^\pi(s)$, represents the expected return obtained by starting from state $s$and following the policy π, while the action-value function, $Q^\pi(s,a)$, measures the expected return of taking action $a$in state $s$and then following the policy π. In Q-learning, the action-value function is the primary quantity learned by the agent. Once the optimal Q-values, $Q^*(s,a)$, are obtained, the optimal policy is determined by selecting the action with the maximum Q-value in each state, given by $\pi^*(s) = arg\ max_a\, Q^*(s,a)$.

The Bellman Optimality Equations define the recursive relationship for optimal decision-making in a Markov Decision Process (MDP). The optimal state-value function is given by

$$V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

and the optimal action-value function is defined as

$$Q^*(s,a) = \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')\right]$$

where immediate rewards are combined with discounted future returns to maximize long-term cumulative reward, Q-learning iteratively approximates these optimal Q-values directly from experience without requiring prior knowledge of transition probabilities.

3.9. Q-Learning Update Rule

In Q-learning, at each time step t, the agent observes a state s_t, selects an action a_t, receives a reward r_{t+1}, and transitions to the next state s_{t+1}. The Q-value is then updated using the following rule:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left[r_{t+1} + \gamma \max_{a'} Q(s_{t+1},a') - Q(s_t,a_t)\right]$$

The term inside the brackets is called the Temporal Difference (TD) error, defined as:

$$\delta_t = r_{t+1} + \gamma \max_{a'} Q(s_{t+1},a') - Q(s_t,a_t)$$

The TD error measures the difference between the current Q-value estimate and the target value based on observed reward and estimated future returns. The learning rate αcontrols how much the Q-value is adjusted toward this target. Over repeated updates, the Q-values gradually converge toward the optimal action-value function Q^*[36].

3.10. Hybrid Ranking Model		
3.10.1	**Normalized Classical Score** $C_i = \omega_1(wkts) + \omega_2(econ) + \omega_3(sr) + \omega_4(avg) + \omega_5(dot) + \omega_6(bdry)$ Where weights are: $\omega_1 = 0.28$, $\omega_2 = 0.25$, $\omega_3 = 0.20$, $\omega_4 = 0.12$, $\omega_5 = 0.10$, $\omega_6 = 0.05$	
	3.10.1.1.	**Wickets Contribution:** $\omega_1(wkts) = 0.28 \times \left(\frac{W_i}{W_{max}}\right)$
	3.10.1.2.	**Economy Contribution:** $\omega_2(econ) = 0.25 \times \left(1 - \frac{E_i}{E_{max}}\right)$
	3.10.1.3.	**Strike Rate Contribution:** $\omega_3(sr) = 0.20 \times \left(1 - \frac{SR_i}{SR_{max}}\right)$
	3.10.1.4.	**Bowling Average Contribution:** $\omega_4(avg) = 0.12 \times \left(1 - \frac{AVG_i}{AVG_{max}}\right)$
	3.10.1.5.	**Dot Ball Contribution:** $\omega_5(dot) = 0.10 \times \left(\frac{DOT_i}{DOT_{max}}\right)$
	3.10.1.6.	**Boundary % Contribution:** $\omega_6(bdry) = 0.05 \times \left(1 - \frac{BDRY_i}{BDRY_{max}}\right)$
3.10.2.	**Reinforcement Learning Metric**	
	3.10.2.1.	**Average Q-Score (RL)** $Average\ Q\text{-}Score = \frac{\sum_{i=1}^{n} Q(s_i,a_i)}{n}$ Where: • $Q(s_i,a_i)$= Q-value of action $a_i$taken in state s_i • n= total number of state-action pairs for the bowler
	3.10.2.2.	**Q-Learning Score Normalization** $\boldsymbol{Q_i}^{\boldsymbol{norm}} = \left(\frac{Q_i - Q_{min}}{Q_{max} - Q_{min}}\right)$
3.10.3.	**Final Score** $\boldsymbol{F_i} = \mathbf{0.65}\, C_i + 0.35\, \boldsymbol{Q_i}^{\boldsymbol{norm}}$	

3.11. Markov Property

The defining characteristic of an MDP is the Markov Property, which states that the future state of a system depends only on the current state and the action taken, and is independent of the sequence of past states and actions that led to it. Formally:

$$P(S_{t+1} \mid S_t, A_t, S_{t-1}, A_{t-1}, \dots) = P(S_{t+1} \mid S_t, A_t)$$

Here, S_t denotes the current state of the match, A_t represents the bowling action, and S_{t+1} is the resulting next state after interaction with the environment.

This property significantly simplifies modelling in cricket analytics, as it implies that predicting the outcome of the next delivery requires only the current match situation and the selected bowling strategy, rather than the full history of previous deliveries.

3.12. State Transition Probability

The probability of transitioning from a current state $S_t = s$ to a subsequent state $S_{t+1} = s'$ is defined as:

$$P_{ss'} = P(S_{t+1} = s' \mid S_t = s)$$

These transition probabilities represent how the environment responds to the agent's action in a given match situation (e.g., resulting in a dot ball, boundary, wicket, or change in pressure state).

The state transition probabilities for all possible states can be arranged into a transition probability matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{pmatrix}$$

where each element p_{ij} represents the probability of transitioning from state i to state j.

It is important to note that each transition probability satisfies $p_{ij} \geq 0$, and for every state i, the sum of all outgoing probabilities equals 1:

$$\sum_{k=1}^{n} p_{ik} = \sum_{k=1}^{n} P(S_{t+1} = k \mid S_t = i)$$

$$\sum_{k=1}^{n} p_{ik} = 1$$

Each row of the matrix, therefore, represents a valid probability distribution over all possible next states. Consequently, the transition matrix P is row-stochastic.

Symbol	**Name**	**Definition**
$\mathbb{S}$	State Space	Set of all possible match situations observed by the agent
$\mathbb{A}$	Action Space	Set of all possible bowling actions available to the agent
$T(s, a, s')$	Transition Function	$T(s, a, s') = P(s' \mid s, a)$, probability that the environment transitions to state s'
$R(s, a)$	Reward Function	Immediate reward received after taking action a in state s
$\gamma \in [0,1]$	Discount Factor	Controls importance of future vs immediate rewards

Figure # 2: Hierarchy Diagram of the 105 States ($Phase \times Length \times Line$)

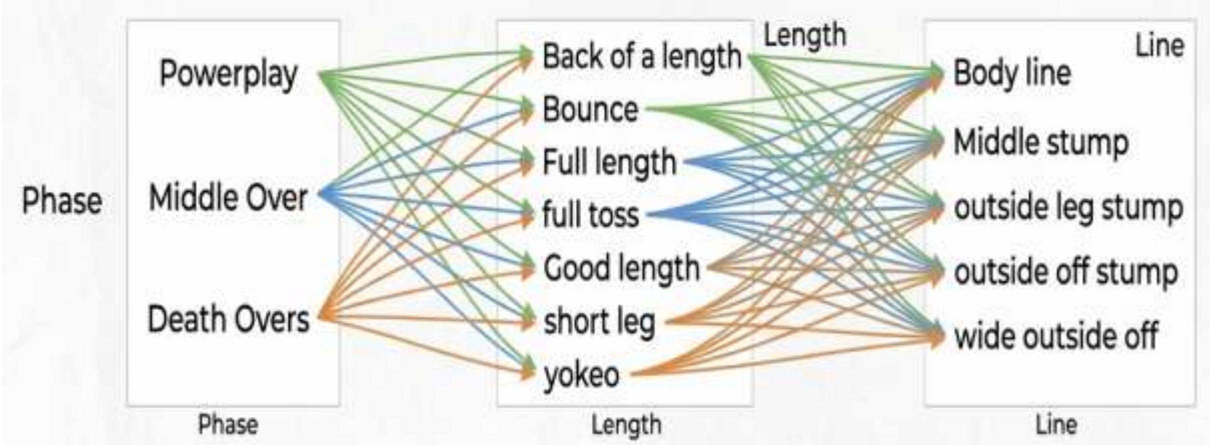
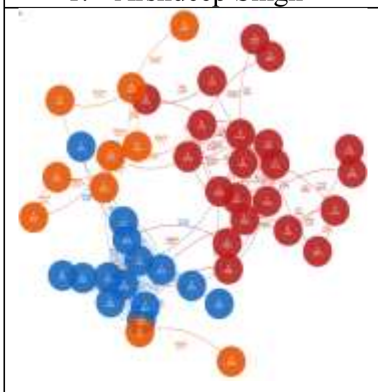
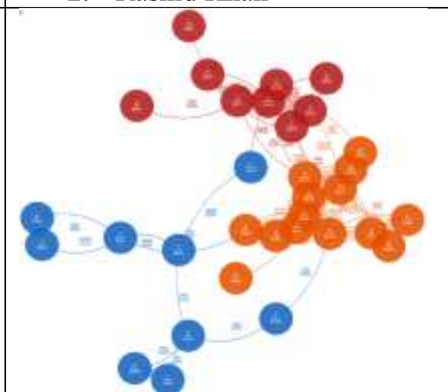
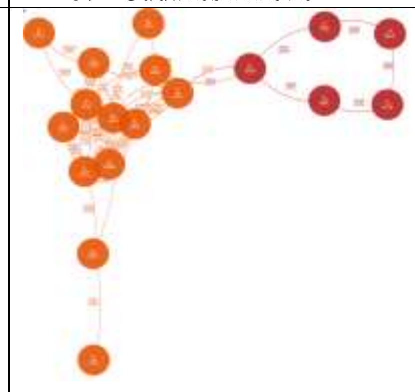
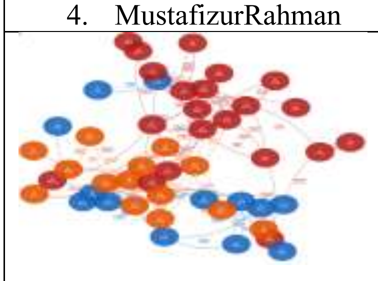
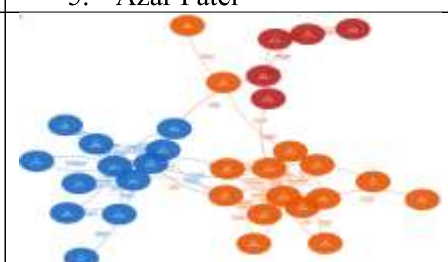
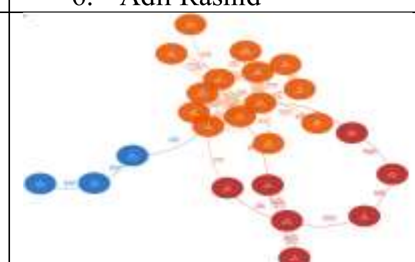
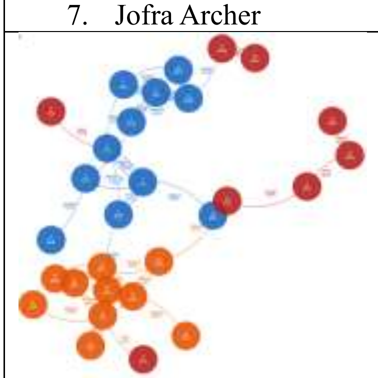
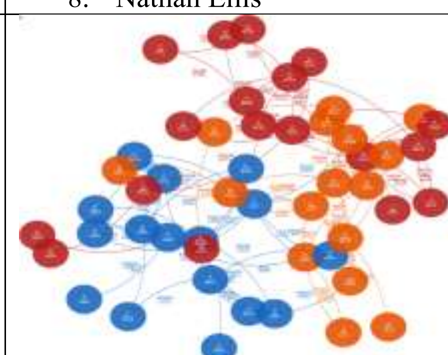
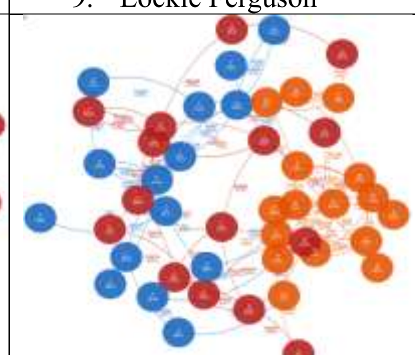
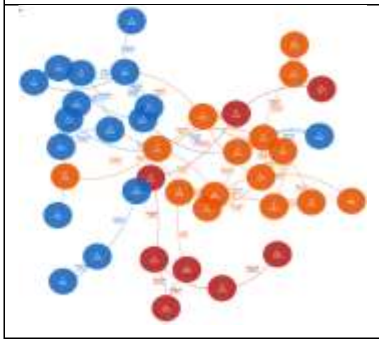


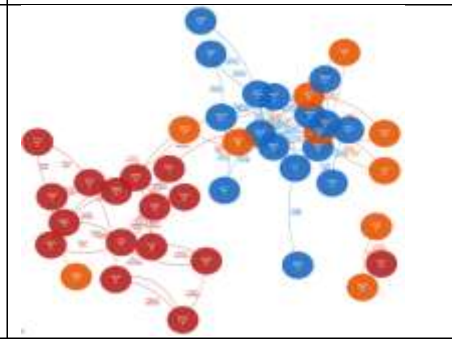
Figure # 3: Markov Decision Process State Transition Diagram of All Nineteen Bowlers, Powerplay (Blue), Middle over (Orange), Death Over (Red)

1. Arshdeep Singh	2. Rashid Khan	3. Gudakesh Motie
		
4. MustafizurRahman	5. Azar Patel	6. Adil Rashid
		
7. Jofra Archer	8. Nathan Ellis	9. Lockie Ferguson
		

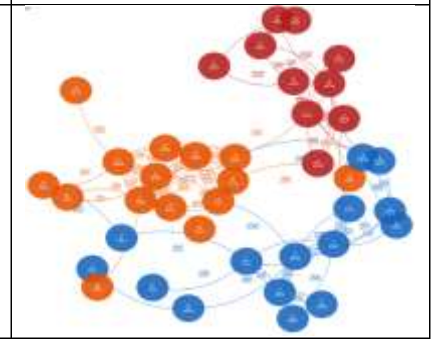
10. Maheesh Theekshana



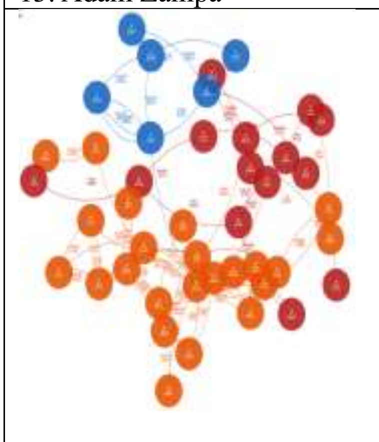
11. Shaheen Shah Afridi



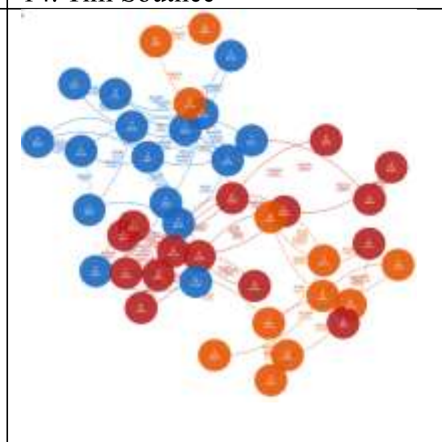
12. Ravi Bishnoi



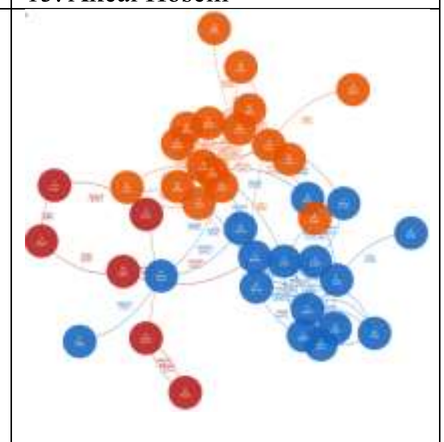
13. Adam Zampa



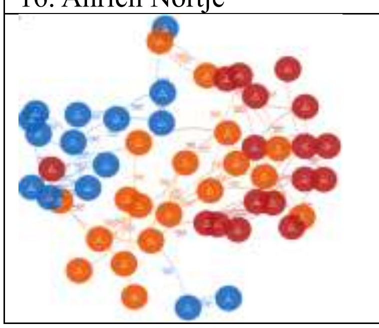
14. Tim Southee



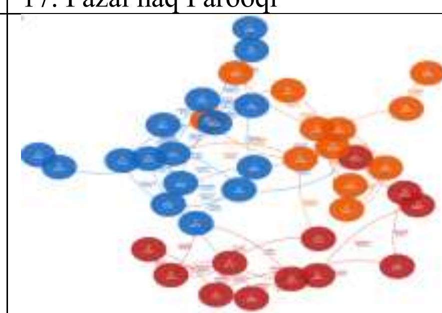
15. Akeal Hosein



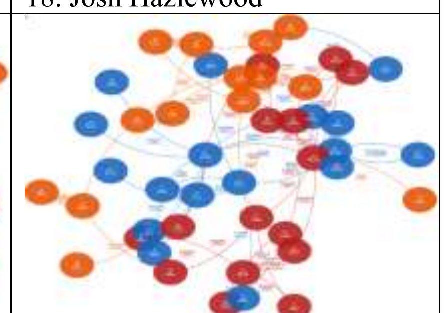
16. Anrich Nortje



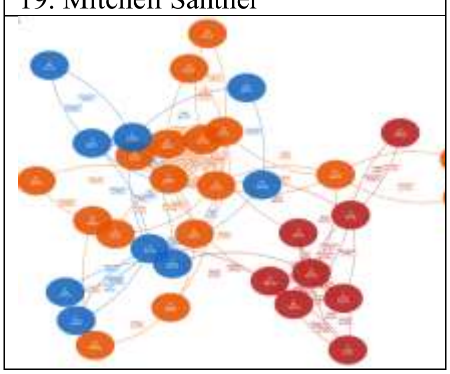
17. Fazal haq Farooqi



18. Josh Hazlewood



19. Mitchell Santner



The Markov Decision Process (MDP) state transition diagrams of all nineteen bowlers illustrate the sequential decision-making patterns followed by elite bowlers in T20 cricket. These diagrams model how bowlers transition between different tactical states defined by match phase, bowling length, and line, thereby capturing the probabilistic nature of bowling strategies. Fast bowlers such as Shaheen Shah Afridi, Arshdeep Singh, Jofra Archer, Josh Hazlewood, Tim Southee, and Anrich Nortje generally exhibited more structured and phase-specific transition patterns, indicating clearly defined tactical roles across powerplay, middle overs, and death overs. For instance, swing bowlers frequently transitioned between full and good-length deliveries outside off stump during the powerplay before shifting toward yorkers and fuller deliveries in the death overs. Bowlers such as Shaheen Shah Afridi and Arshdeep Singh demonstrated dual-role transition structures, reflecting their effectiveness both as new-ball swing bowlers and death-over specialists.

In contrast, variation-based pacers including Mustafizur Rahman, Nathan Ellis, and Fazalhaq Farooqi displayed transition pathways characterized by frequent switching between standard deliveries, slower balls, cutters, and yorkers. These patterns suggest a stronger reliance on deception and pace variation rather than pure speed or line-and-length consistency. Such bowlers often occupied states associated with tactical unpredictability, particularly in middle and death overs where variation becomes strategically important. Spin bowlers such as Rashid Khan, Adil Rashid, Ravi Bishnoi, Adam Zampa, Mitchell Santner, Akeal Hosein, Maheesh Theekshana, and Gudakesh Motie exhibited denser and more interconnected transition networks, indicating greater tactical adaptability and more diverse state movement. Their diagrams suggest repeated transitions among good-length and middle-stump states with variations in line, pace, and spin type, reflecting the strategic complexity of spin bowling. Attacking leg-spinners such as Rashid Khan and Adam Zampa demonstrated more dynamic state transitions, while defensive spinners like Mitchell Santner and Akeal Hosein showed clustered transitions around containment-oriented states, emphasizing economy and pressure building.

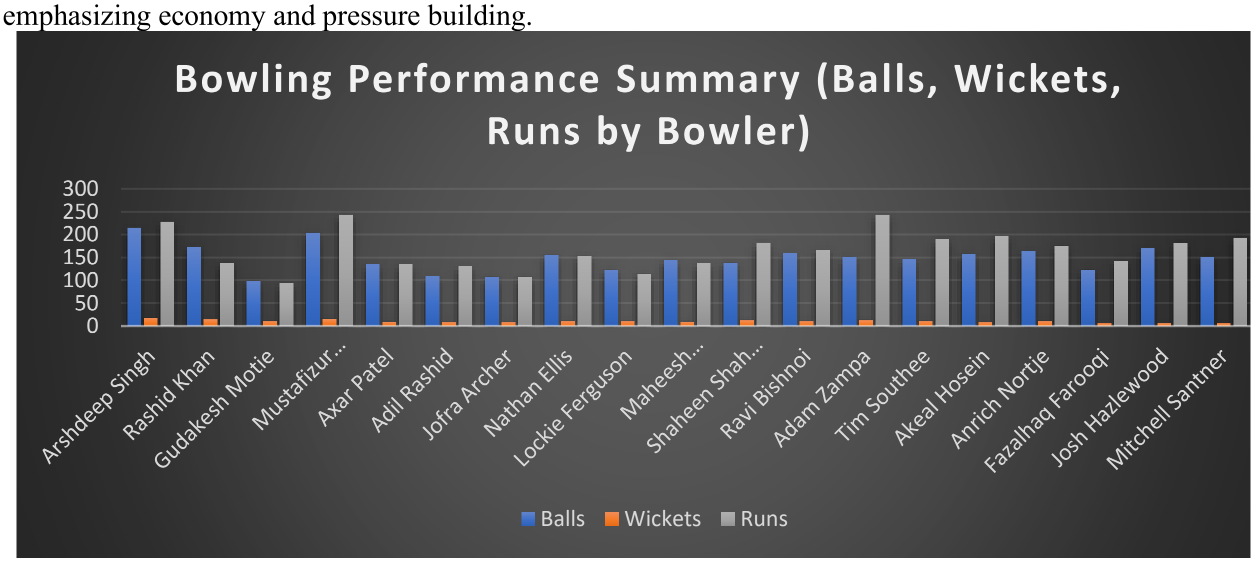


Figure # 4: Bowling Performance Summary (Balls, Wickets, Runs by Bowler)

Figure 4 presents the bowling performance summary of nineteen T20 bowlers by comparing total balls bowled, wickets taken, and runs conceded. Arshdeep Singh recorded the highest workload (214 balls) and wickets (18), indicating strong wicket-taking ability despite conceding 227 runs. Rashid Khan and Gudakesh Motie showed efficient performances with good wicket returns while maintaining relatively lower runs conceded, highlighting better bowling economy. In contrast, bowlers such as Mustafizur Rahman and Adam Zampa took wickets but also conceded comparatively higher runs, suggesting a more aggressive yet expensive bowling approach. Overall, the figure reflects variation in workload, effectiveness, and economy across bowlers in T20 matches.

Table # 3: Performance Ranking of T20 Bowlers Using Traditional Bowling Metrics and Reinforcement Learning-Based Average Q-Scores

Rank	Bowler Name	Economy Rate	Bowling Strike Rate	Bowling average	Dot Ball %	Boundary %	Average Q score
1	Arshdeep Singh	6.364	11.89	12.61	36.9	15	21.48474
2	Rashid Khan	4.786	12.36	9.86	34.7	6.4	21.40458
3	Gudakesh Motie	5.753	9.7	9.3	32	9.3	19.7649
4	Mustafizur Rahman	7.182	13.53	16.2	31.5	15.3	17.28018
5	Axar Patel	6	14.89	14.89	36.6	10.4	17.20308
6	Adil Rashid	7.222	13.5	16.25	36.1	15.7	17.96552
7	Jofra Archer	6	13.38	13.38	37.4	13.1	15.25731
8	Nathan Ellis	5.923	15.5	15.3	37.4	12.9	14.69685
9	Lockie Ferguson	5.512	12.3	11.3	35.8	8.1	12.9199
10	Maheesh Theekshana	5.748	15.89	15.22	32.9	9.8	13.2988
11	Shaheen Shah Afridi	7.87	11.5	15.08	35.5	20.3	12.04703
12	Ravi Bishnoi	6.264	15.9	16.6	36.5	13.2	10.85666
13	Adam Zampa	9.656	12.58	20.25	21.9	21.9	12.11634
14	Tim Southee	7.821	14.5	18.9	27.6	16.6	10.55913
15	Akeal Hosein	7.529	19.62	24.62	31.8	17.8	11.2186
16	Anrich Nortje	6.366	16.4	17.4	42.1	15.2	6.719518
17	Fazalhaq Farooqi	6.992	24.2	28.2	35.5	16.5	8.061513
18	Josh Hazlewood	6.391	28.17	30	43.2	15.4	7.035112
19	Mitchell Santner	7.629	30.2	38.4	27.2	13.9	5.44853

Table 3 ranks 19 T20 bowlers by combining traditional metrics with an Average Q-score that measures strategic decision-making quality. Arshdeep Singh ranks first with the highest Q-score (21.48) and strong traditional numbers, followed closely by Rashid Khan (21.40 Q-score, best economy of 4.79) and Gudakesh Motie (19.76 Q-score, best average of 9.3). Key divergences emerge: Adil Rashid ranks sixth with a solid Q-score (17.97) despite mediocre traditional stats, indicating he is undervalued by conventional metrics. Conversely, Anrich Nortje (16th) and Josh Hazlewood (18th) have excellent dot ball percentages (42.1% and 43.2%) but very low Q-scores (6.72 and 7.04), suggesting their predictable bowling lacks phase-adaptive tactical variation. Mitchell Santner ranks last with the lowest Q-score (5.45). The table confirms that Q-score captures strategic intelligence as a distinct performance dimension not fully reflected in traditional bowling statistics.

Algorithm 1 Pseudocode for Calculating Transition Probability Using MDP for T20 Cricket Bowling

initialize: states = (Phase, Length, Line), actions = (Bowling Variations), transition probability
define: transition probability matrix
trans_prob = {}
initialize: transition probabilities
for state **in** states:
$trans_{Prob}$[state] = {}

for action **in** actions:
$trans_{-Prob}$[state][action] = {}
assign: transition probabilities from observed deliveries
$trans_{-Prob}$['Powerplay, Good Length, Middle Stump'] ['Yorker']
['Powerplay, Yorker, Middle Stump'] = 0.35
$trans_{-Prob}$['Middleover, Short, Outside Off'] ['Bouncer']
['Middle over, Short, Outside Off'] = 0.42
$trans_{-Prob}$['Death over, Full Length, Wide Outside Off'] ['Slower Ball']
['Death over, Full Length, Wide Outside Off'] = 0.58
compute: transition probability for state-action pair
define $Comp\_trans_{-Prob}$(state, action, next_state):
if next state in $trans_{-Prob}$[state][action]:
return $trans_{-Prob}$[state][action][next_state]
else:
return 0
calculate:
result ← $Comp\_trans_{-Prob}$(state, action, next_state)
end

Algorithm 2 Q-Learning Algorithm for T20 Cricket Bowling Strategy Optimization

initialize: Agent ← Bowler
initialize: Environment ← Cricket match situation
(Phase, Length, Line, Bowling Variation, Outcome)
initialize: Q-table with **0**
initialize: states s_t, actions a_t
while training do
for $t \rightarrow 1$**to** T**do**
extract state s_t
choose action $a_t$using ϵ-greedy policy with probability $1 - \epsilon$, or explore random action with probability ϵ
execute bowling action (a_t)
move into the subsequent state (s_{t+1})
compute reward R_t
compute Q-value using Bellman technique:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[R_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right]$$

update transition count:

$$C(s_t, a_t, s_{t+1}) = C(s_t, a_t, s_{t+1}) + 1$$

set $s_t \leftarrow s_{t+1}$
end for
decay exploration rate:

$$\epsilon \leftarrow max\ (\epsilon_{min}, \epsilon \times 0.85)$$

end while
compute transition probabilities:

$$T(s, a, s') = \frac{C(s, a, s')}{\sum_{s'} C(s, a, s')}$$

extract optimal policy:

$$\pi^*(s) = arg\ \max_{a}\ Q(s, a)$$

return: Q-table, optimal policy π^*, transition model T

Algorithms 1 and 2, together with the flowchart in Figure 1, present a complete MDP-based Q-learning framework for optimizing T20 bowling strategies. The process begins by initializing the agent's state and action-value function, then assigning IDs to all possible bowling actions (variations). The agent selects an action randomly or via an ε-greedy policy, moves from the current state to the next state based on transition probabilities derived from observed ball-by-ball data (Algorithm 1), and computes rewards based on runs conceded and wickets taken. The Q-value is updated using the Bellman equation (Algorithm 2), and the maximum value for the next state is recorded. The agent's memory stores both transitions and generated rewards, enabling iterative learning. This cycle repeats until termination, with the framework ultimately returning an optimal policy $\pi^*(s) = arg\ max_a\ Q(s,a)$. In essence, Algorithm 1 defines the transition logic ("what happens next?"), Algorithm 2 drives learning through reward-based updates ("what action is best?"), and the flowchart visualizes the sequential decision-making loop—from state initialization to policy extraction—providing a practical blueprint for data-driven, phase-specific bowling optimization.

Table # 4: Bowling State-Action Combinations with Highest Q-Values for All the Nineteen Bowlers

		State (Phase, Length, Line)	Action (Bowling Variation)	Q_Value
1	Arshdeep Singh	Deathover, Full Length, Wide Outside Off	Slower Ball	43.3443
		Deathover, Full Length, Outside Leg Stump	Standard Delivery	42.5831
		Deathover, Good Length, Outside Off Stump	Slower Ball	41.1738
		Deathover, Short, Outside Leg Stump	Slower Ball	39.4611
		Deathover, Short, Wide Outside Off	Bouncer	39.1541
		Middleover, Good Length, Wide Outside Off	Standard Delivery	37.8347
		Deathover, Full Length, Wide Outside Off	Seam	36.9148
		Deathover, Full Length, Outside Off Stump	Standard Delivery	36.8268
		Deathover, Good Length, Outside Off Stump	Standard Delivery	35.2004
		Deathover, Good Length, Wide Outside Off	Seam	34.7248
2	Rashid Khan	Deathover, Good Length, Outside Off Stump	Googly	64.7692
		Deathover, Good Length, Wide Outside Off	Legbreak	60.5707
		Deathover, Good Length, Middle Stump	Googly	54.4641
		Deathover, Short, Outside Off Stump	Googly	50.7449
		Deathover, Good Length, Middle Stump	Flatter	43.0779
		Deathover, Full Length, Outside Off Stump	Googly	42.0718
		Middleover, Short, Middle Stump	Legbreak	40.6830
		Deathover, Full Length, Outside Off Stump	Flatter	39.2960
		Deathover, Good Length, Wide Outside Off	Googly	37.2090
		Middleover, Good Length, Middle Stump	Flatter	33.1324
3	Gudakesh Motie	Deathover, Good Length, Outside Leg Stump	Orthodox	34.5266
		Deathover, Full Length, Middle Stump	Orthodox	33.2170
		Deathover, Full Length, Wide Outside Off	Orthodox	29.0828
		Deathover, Good Length, Middle Stump	Orthodox	27.3108
		Deathover, Full Toss, Wide Outside Off	Orthodox	27.2467
		Middleover, Short, Wide Outside Off	Standard Delivery	26.9713

		Middleover, Short, Outside Off Stump	Slower Ball	22.9712
		Middleover, Good Length, Outside Leg Stump	Standard Delivery	21.8827
		Middleover, Yorker, Outside Leg Stump	Flatter	20.7005
		Middleover, Good Length, Wide Outside Off	Standard Delivery	17.4549
4	**Mustafizur Rahman**	Deathover, Good Length, Wide OutSide off	Slower Ball	35.9417
		Deathover, Good Length, Middle Stump	Off-cutter	34.8136
		Middleover,Good Length, Wide OutSide off	Off-cutter	34.2976
		Deathover, Full Length, Outside Off Stump	Yorker	33.3519
		Deathover, Back of a Length, Wide OutSide off	Off-cutter	33.3346
		Middleover, Bouncer, Wide OutSide off	Standard Delivery	31.6141
		Deathover, Good Length, Wide OutSide off	Off-cutter	31.4587
		Deathover, Back of a Length, Wide OutSide off	Slower Ball	27.7839
		Middleover, Full Length, Outside Off Stump	Standard Delivery	27.7697
		Middleover, Good Length, Body line	Slower Ball	26.8581
5	**Axar Patel**	Middleover, Good Length, Outside Off Stump	Outswinger	36.4138
		Middleover, Short, Outside Off Stump	Standard Delivery	35.1670
		Middleover, Short, Body line	Slower Ball	33.2900
		Middleover, Good Length, Middle Stump	Standard Delivery	31.0856
		Middleover, Full Length, Wide Outside Off	Slower Ball	30.4757
		Middleover, Short, Wide Outside Off	Standard Delivery	29.4709
		Middleover, Full Length, Outside Off Stump	Outswinger	28.8280
		Middleover, Good Length, Middle Stump	Slower Ball	28.6769
		Middleover, Short, Outside Leg Stump	Standard Delivery	27.7525
		Middleover, Full Length, Middle Stump	Slower Ball	27.2920
6	**Adil Rashid**	Middleover, Full Length, Wide Outside Off	Outswinger	36.9583
		Middleover, Good Length, Wide Outside Off	Googly	35.8379
		Middleover, Short, Outside Off Stump	Slower Ball	35.2457
		Middleover, Full Length, Outside Off Stump	Googly	32.7216
		Middleover, Good Length, Outside Leg Stump	Skidding	32.0868
		Middleover, Full Length, Outside Off Stump	Slower Ball	31.0832
		Middleover, Full Length, Middle Stump	Googly	30.2751
		Middleover, Short, Wide Outside Off	Skidding	30.1481
		Middleover, Full Length, Outside Leg Stump	Standard Delivery	29.9117
		Middleover, Full Length, Outside Leg Stump	Slower Ball	29.9090
7	**Jofra Archer**	Deathover, Full Length, Outside Off Stump	Yorker	46.0953
		Deathover, Short, Outside Off Stump	Slower Ball	35.4518
		Deathover, Full Length, Outside Off Stump	Standard Delivery	27.4145
		Deathover, Bouncer, Wide Outside Off	Bouncer	26.8033
		Powerplay, Good Length, Wide Outside Off	Slower Ball	24.2161
		Middleover, Full Length, Outside Off Stump	Standard Delivery	22.9577
		Middleover, Short, Body line	Standard Delivery	21.2367

		Powerplay, Good Length, Outside Leg Stump	Standard Delivery	20.2489
		Powerplay, Full Length, Outside Leg Stump	Standard Delivery	20.1460
		Deathover, Short, Outside Leg Stump	Bouncer	18.2279
8	**Nathan Ellis**	Powerplay, Short, Middle Stump	Slower Ball	39.9068
		Deathover, Short, Outside Off Stump	Slower Ball	30.3380
		Deathover, Short, Outside Leg Stump	Bouncer	30.2272
		Deathover, Full Length, Outside Leg Stump	Slower Ball	29.2588
		Powerplay, Good Length, Wide Outside Off	Seam	26.7648
		Powerplay, Full Length, Outside Leg Stump	Standard Delivery	25.3514
		Powerplay, Good Length, Body line	Slower Ball	25.3227
		Powerplay, Yorker, Outside Off Stump	Inswinger	24.6508
		Powerplay, Good Length, Middle Stump	Slower Ball	24.0230
		Powerplay, Good Length, Outside Off Stump	Slower Ball	23.8472
9	**Lockie Ferguson**	Deathover, Good Length, Outside Off Stump	Standard Delivery	30.1886
		Deathover, Good Length, Outside Leg Stump	Standard Delivery	29.3732
		Deathover, Yorker, Wide Outside Off	Standard Delivery	26.5362
		Deathover, Full Length, Wide Outside Off	Standard Delivery	25.4550
		Deathover, Full Length, Outside Off Stump	Standard Delivery	22.1419
		Middleover, Full Length, Body line	Standard Delivery	20.3121
		Powerplay, Full Length, Outside Off Stump	Outswinger	19.7927
		Deathover, Full Length, Middle Stump	Yorker	19.5282
		Middleover, Full Length, Middle Stump	Off-cutter	18.7130
		Middleover, Short, Body line	Bouncer	18.0944
10	**Maheesh Theekshana**	Middleover, Good Length, Wide Outside Off	Skidding	37.3874
		Middleover, Good Length, Outside Off Stump	Slower Ball	36.2614
		Middleover, Good Length, Outside Leg Stump	Standard Delivery	30.7382
		Middleover, Full Length, Middle Stump	Standard Delivery	28.2419
		Powerplay, Good Length, Middle Stump	Slower Ball	27.3513
		Middleover, Full Length, Outside Off Stump	Skidding	26.3472
		Middleover, Full Length, Outside Off Stump	Standard Delivery	25.0575
		Powerplay, Full Length, Middle Stump	Slower Ball	24.8503
		Middleover, Yorker, Outside Off Stump	Standard Delivery	23.8619
		Middleover, Full Length, Wide Outside Off	Slower Ball	23.0837
11	**Shaheen Shah Afridi**	Deathover,Yorker, Middle Stump	Yorker	54.3730
		Powerplay, Short, Outside Off Stump	Bouncer	25.9142
		Middleover, Yorker, Middle Stump	Yorker	24.4980
		Deathover, Full Length, Out side Leg Stump	Yorker	22.5003
		Powerplay, Short, Wide OutSide off	Yorker	20.4505
		Powerplay, Good Length, Body line	Standard Delivery	20.2200
		Deathover, Full Length, Middle Stump	Standard Delivery	19.2058
		Powerplay, Good Length, Outside Off Stump	Inswinger	19.1799

		Powerplay, Good Length, Wide OutSide off	Inswinger	18.1520
		Powerplay, Good Length, Wide OutSide off	Standard Delivery	17.8415
12	Ravi Bishnoi	Middleover, Good Length, Body line	Legbreak	20.9315
		Middleover, Good Length, Middle Stump	Slower Ball	20.4042
		Deathover, Short, Outside Off Stump	Legbreak	19.6781
		Middleover, Good Length, Outside Leg Stump	Slower Ball	19.2913
		Deathover, Good Length, Outside Leg Stump	Slower Ball	18.3169
		Powerplay, Short, Middle Stump	Legbreak	18.1694
		Deathover, Good Length, Middle Stump	Legbreak	18.1418
		Deathover, Good Length, Outside Off Stump	Googly	18.1385
		Middleover, Good Length, Outside Off Stump	Legbreak	17.0019
		Powerplay, Good Length, Wide Outside Off	Googly	16.6323
13	Adam Zampa	Middleover, Full Length, Outside Off Stump	Slower Ball	45.6651
		Middleover, Good Length, Wide Outside Off	Slower Ball	43.2711
		Middleover, Good Length, Wide Outside Off	Flatter	39.2235
		Middleover, Good Length, Outside Off Stump	Slower Ball	34.8255
		Middleover, Full Length, Outside Leg Stump	Standard Delivery	33.6637
		Middleover, Full Length, Outside Off Stump	Standard Delivery	32.5851
		Deathover, Full Length, Middle Stump	Standard Delivery	32.3854
		Middleover, Full Length, Wide Outside Off	Slower Ball	31.9947
		Deathover, Short, Middle Stump	Flatter	29.9449
		Deathover, Full Length, Outside Leg Stump	Standard Delivery	27.7609
14	Tim Southee	Middleover, Short, Middle Stump	Bouncer	31.6407
		Middleover, Full Length, Outside Off Stump	Slower Ball	28.1392
		Deathover, Full Length, Middle Stump	Slower Ball	25.9712
		Deathover, Short, Outside Leg Stump	Standard Delivery	24.9163
		Middleover, Short, Outside Off Stump	Slower Ball	22.9884
		Deathover, Short, Wide Outside Off	Standard Delivery	21.1636
		Deathover, Full Length, Wide Outside Off	Off-cutter	20.9351
		Powerplay, Yorker, Wide Outside Off	Outswinger	20.5614
		Deathover, Short, Wide Outside Off	Seam	20.2692
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	18.5826
15	Akeal Hosein	Powerplay, Back of a Length, Body line	Inswinger	31.9524
		Powerplay, Full Length, Body line	Outswinger	28.2971
		Powerplay, Full Length, Body line	Inswinger	27.0017
		Powerplay, Full Length, Outside Leg Stump	Inswinger	26.5550
		Powerplay, Full Length, Body line	Flatter	24.5421
		Middleover, Full Length, Body line	Inswinger	24.5286
		Powerplay, Back of a Length, Body line	Slower Ball	22.2816
		Powerplay, Good Length, Middle Stump	Standard Delivery	22.2100
		Powerplay, Good Length, Wide Outside Off	Standard Delivery	20.9932

		Powerplay, Back of a Length, Outside Leg Stump	Flatter	20.7482
16	**Anrich Nortje**	Middleover, Full Length, Out side Leg Stump	Bouncer	22.2918
		Middleover, Short, Middle Stump	Bouncer	21.2614
		Deathover, Good Length, Out side Leg Stump	Standard Delivery	20.8383
		Middleover, Good Length, Out side Leg Stump	Standard Delivery	20.0148
		Middleover, Good Length, Middle Stump	Slower Ball	17.9232
		Middleover, Yorker, Middle Stump	Standard Delivery	17.9214
		Middleover, Full Length, Body line	Standard Delivery	17.8244
		Middleover, Short, Middle Stump	Standard Delivery	17.6871
		Middleover, Full Length, Out side Leg Stump	Standard Delivery	17.3614
		Middleover, Short, Outside Off Stump	Bouncer	17.1733
17	**Fazalhaq Farooqi**	Deathover, Short, Outside Leg Stump	Bouncer	33.5300
		Deathover, Short, Outside Off Stump	Standard Delivery	27.3108
		Deathover, Full Length, Outside Off Stump	Standard Delivery	22.2387
		Powerplay, Full Length, Wide Outside Off	Slower Ball	21.2434
		Powerplay, Good Length, Wide Outside Off	Off-cutter	20.1094
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	19.4573
		Powerplay, Short, Middle Stump	Standard Delivery	18.8592
		Deathover, Full Length, Wide Outside Off	Slower Ball	18.1535
		Powerplay, Good Length, Middle Stump	Standard Delivery	17.1273
		Powerplay, Good Length, Middle Stump	Slower Ball	16.8311
18	**Josh Hazlewood**	Powerplay, Good Length, Middle Stump	Inswinger	33.7581
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	30.5986
		Powerplay, Good Length, Wide Outside Off	Inswinger	29.7886
		Powerplay, Back of a Length, Outside Leg Stump	Off-cutter	27.2483
		Powerplay, Good Length, Body line	Standard Delivery	25.7336
		Powerplay, Good Length, Wide Outside Off	Standard Delivery	24.9845
		Powerplay, Full Length, Outside Off Stump	Standard Delivery	24.3686
		Powerplay, Full Length, Outside Off Stump	Slower Ball	23.1724
		Powerplay, Short, Middle Stump	Slower Ball	21.7028
		Powerplay, Good Length, Wide Outside Off	Bouncer	20.5339
19	**Mitchell Santner**	Middleover, Full Length, Outside Leg Stump	Bouncer	20.3325
		Middleover, Full Length, Outside Off Stump	Standard Delivery	15.5920
		Middleover, Short, Wide Outside Off	Bouncer	14.1893
		Middleover, Good Length, Wide Outside Off	Outswinger	13.4143
		Middleover, Full Length, Middle Stump	Off-cutter	12.2421
		Middleover, Good Length, Wide Outside Off	Flatter	11.7901
		Middleover, Short, Wide Outside Off	Standard Delivery	11.7425
		Middleover, Full Length, Outside Leg Stump	Outswinger	11.4536
		Middleover, Full Length, Wide Outside Off	Outswinger	11.4299
		Deathover, Good Length, Wide Outside Off	Slower Ball	11.2882

Table # 4 presents the top ten bowling state-action combinations with the highest Q-values for each of the nineteen T20 bowlers under the proposed Markov Decision Process (MDP)-based Q-learning framework. Each state is defined by the match phase (Powerplay, Middleover, or Deathover), delivery length, and bowling line, while the corresponding action represents the optimal bowling variation learned by the reinforcement learning agent. The results indicate that bowling effectiveness is highly context-dependent, as different bowlers exhibit distinct optimal strategies across match situations. A strong pattern is observed in the Deathover phase, where most pace bowlers achieve their highest Q-values through defensive and wicket-taking deliveries such as slower balls, yorkers, bouncers, and seam variations. For example, Arshdeep Singh recorded his highest Q-value (43.3443) with a slower ball delivered at full length wide outside off during the death overs, highlighting the strategic importance of pace variation and wide lines in limiting scoring opportunities. Similarly, Jofra Archer and Shaheen Shah Afridi showed strong preference for yorkers in death-over conditions, with Q-values of 46.0953 and 54.3730 respectively, emphasizing yorkers as an effective wicket-taking and run-containment option under high-pressure situations. Spin bowlers demonstrated different optimal patterns, primarily in the Middleover and Deathover phases. Rashid Khan achieved the highest overall Q-value (64.7692) with a googly bowled on a good length outside off stump during death overs, indicating exceptional strategic reward from deceptive spin variations. Likewise, Gudakesh Motie and Ravi Bishnoi showed high-value actions through orthodox spin and legbreak deliveries, suggesting that variation, flight, and line control are critical for spinners in restricting runs and inducing dismissals.

Several bowlers displayed phase specialization. Nathan Ellis, Josh Hazlewood, and Akeal Hosein generated their highest Q-values in the Powerplay, reflecting their effectiveness with disciplined line-length combinations and swing-based variations early in the innings. For instance, Josh Hazlewood's optimal action was an inswinger on a good length targeting middle stump (Q = 33.7581), which aligns with his real-world strength of exploiting movement with the new ball. The analysis also reveals that repeated appearance of certain states among top Q-values suggests robust strategic consistency. Wide outside-off lines, full-length deliveries, and slower-ball variations frequently appear across multiple bowlers, particularly in death overs, indicating that these tactical choices maximize expected rewards under pressure situations. Overall, the table demonstrates that the proposed Q-learning model successfully identifies bowler-specific optimal strategies by capturing the interaction between match phase, delivery characteristics, and bowling variation.

Table # 5: Optimal Bowling State-Action Combinations of All Nineteen Bowler

		State (Phase, Length, Line)	**Action (Bowling Variation)**	**Q_Value**
1	**Arshdeep Singh**	Deathover, Full Length, Wide Outside off	Slower Ball	43.34
		Deathover, Full Length , Out side Leg Stump	Standard Delivery	42.58
		Deathover, Good Length, Outside Off Stump	Slower Ball	41.17
		Deathover, Short, Out side Leg Stump	Slower Ball	39.46
		Deathover, Short, Wide OutSide off	Bouncer	39.15
		Middleover, Good Length , Wide OutSide off	Standard Delivery	37.15
		Deathover, Full Length, Outside Off Stump	Standard Delivery	36.83
		Deathover, Good Length, Wide OutSide off	Seam	34.72
		Deathover, Short, Middle Stump	Seam	31.58
		Deathover, Good Length, Middle Stump	Standard Delivery	28.03
2	**Rashid**	Deathover, Good Length, Outside Off Stump	Googly	64.77
		Deathover, Good Length, Wide Outside off	Legbreak	60.57

		Deathover, Good Length, Middle Stump	Googly	54.46
		Deathover, Short, Outside Off Stump	Googly	50.74
		Deathover, Full Length, Outside Off Stump	Googly	42.07
		Middleover, Short, Middle Stump	Legbreak	40.68
		Middleover, Good Length, Middle Stump	Flatter	33.13
		Middleover, Full Length, Middle Stump	Slower Ball	30.87
		Middleover, Full Length, Wide Outside off	Inswinger	29.34
		Middleover, Good Length, Wide Outside off	Flatter	28.94
3	Gudakesh Motie	Deathover, Good Length, Out side Leg Stump	Orthodox	34.53
		Deathover, Full Length, Middle Stump	Orthodox	33.22
		Deathover, Full Length, Wide Outside off	Orthodox	29.08
		Deathover, Good Length, Middle Stump	Orthodox	27.31
		Deathover, Full Toss, Wide Outside off	Orthodox	27.25
		Middleover, Short, Wide Outside off	Standard Delivery	26.97
		Middleover, Short, Outside Off Stump	Slower Ball	22.97
		Middleover, Good Length, Out side Leg Stump	Standard Delivery	21.88
		Middleover, Yorker, Out side Leg Stump	Flatter	20.7
		Middleover, Good Length, Wide Outside off	Standard Delivery	17.45
4	Mustafizur Rahman	Deathover, Good Length, Wide Outside off	Slower Ball	35.94
		Deathover, Good Length, Middle Stump	Off-cutter	34.81
		Middleover, Good Length, Wide Outside off	Off-cutter	34.3
		Deathover, Full Length, Outside Off Stump	Yorker	33.35
		Deathover, Back of a Length, Wide Outside off	Off-cutter	33.33
		Middleover, Bouncer, Wide Outside off	Standard Delivery	31.61
		Middleover, Full Length, Outside Off Stump	Standard Delivery	27.77
		Middleover, Good Length, Body line	Slower Ball	26.86
		Middleover, Full Length, Wide Outside off	Slower Ball	25.53
		Middleover, Short, Outside Off Stump	Standard Delivery	25.3
5	Axar Patel	Middleover, Good Length, Outside Off Stump	Outswinger	36.41
		Middleover, Short, Outside Off Stump	Standard Delivery	35.17
		Middleover, Short, Body line	Slower Ball	33.29
		Middleover, Good Length, Middle Stump	Standard Delivery	31.09
		Middleover, Full Length, Wide Outside off	Slower Ball	30.48
		Middleover, Short, Wide Outside off	Standard Delivery	29.47
		Middleover, Full Length, Outside Off Stump	Outswinger	28.83
		Middleover, Short, Out side Leg Stump	Standard Delivery	27.75
		Middleover, Full Length, Middle Stump	Slower Ball	27.29
		Middleover, Good Length, Out side Leg Stump	Standard Delivery	27.11
6	Ad	Middleover, Full Length, Wide Outside off	Outswinger	36.96

		Middleover, Good Length, Wide Outside off	Googly	35.84
		Middleover, Short, Outside Off Stump	Slower Ball	35.25
		Middleover, Full Length, Outside Off Stump	Googly	32.72
		Middleover, Good Length, Out side Leg Stump	Skidding	32.09
		Middleover, Full Length, Middle Stump	Googly	30.28
		Middleover, Short, Wide Outside off	Skidding	30.15
		Middleover, Full Length, Out side Leg Stump	Standard Delivery	29.91
		Middleover, Good Length, Middle Stump	Standard Delivery	29.46
		Middleover, Short, Middle Stump	Googly	23.74
7	**Jofra Archer**	Deathover, Full Length, Outside Off Stump	Yorker	46.1
		Deathover, Short, Outside Off Stump	Slower Ball	35.45
		Deathover, Bouncer, Wide Outside off	Bouncer	26.8
		Powerplay, Good Length, Wide Outside off	Slower Ball	24.22
		Middleover, Full Length, Outside Off Stump	Standard Delivery	22.96
		Middleover, Short, Body line	Standard Delivery	21.24
		Powerplay, Good Length, Out side Leg Stump	Standard Delivery	20.25
		Powerplay, Full Length, Out side Leg Stump	Standard Delivery	20.15
		Deathover, Short, Out side Leg Stump	Bouncer	18.23
		Powerplay, Short, Out side Leg Stump	Standard Delivery	16.79
8	**Nathan Ellis**	Powerplay, Short, Middle Stump	Slower Ball	39.91
		Deathover, Short, Outside Off Stump	Slower Ball	30.34
		Deathover, Short, Out side Leg Stump	Bouncer	30.23
		Deathover, Full Length, Out side Leg Stump	Slower Ball	29.26
		Powerplay, Good Length, Wide Outside off	Seam	26.76
		Powerplay, Full Length, Out side Leg Stump	Standard Delivery	25.35
		Powerplay, Good Length, Body line	Slower Ball	25.32
		Powerplay, Yorker, Outside Off Stump	Inswinger	24.65
		Powerplay, Good Length, Middle Stump	Slower Ball	24.02
		Powerplay, Good Length, Outside Off Stump	Slower Ball	23.85
9	**Lockie Ferguson**	Deathover, Good Length, Outside Off Stump	Standard Delivery	30.19
		Deathover, Good Length, Out side Leg Stump	Standard Delivery	29.37
		Deathover, Yorker, Wide Outside off	Standard Delivery	26.54
		Deathover, Full Length, Wide Outside off	Standard Delivery	25.46
		Deathover, Full Length, Outside Off Stump	Standard Delivery	22.14
		Middleover, Full Length, Body line	Standard Delivery	20.31
		Powerplay, Full Length, Outside Off Stump	Outswinger	19.79
		Deathover, Full Length, Middle Stump	Yorker	19.53
		Middleover, Full Length, Middle Stump	Off-cutter	18.71
		Middleover, Short, Body line	Bouncer	18.09

10	Maheesh Theekshana	Middleover, Good Length, Wide Outside off	Skidding	37.39
		Middleover, Good Length, Outside Off Stump	Slower Ball	36.26
		Middleover, Good Length, Out side Leg Stump	Standard Delivery	30.74
		Middleover, Full Length, Middle Stump	Standard Delivery	28.24
		Powerplay, Good Length, Middle Stump	Slower Ball	27.35
		Middleover, Full Length, Outside Off Stump	Skidding	26.35
		Powerplay, Full Length, Middle Stump	Slower Ball	24.85
		Middleover, Yorker, Outside Off Stump	Standard Delivery	23.86
		Middleover, Full Length, Wide Outside off	Slower Ball	23.08
		Middleover, Full Length, Out side Leg Stump	Standard Delivery	21.23
11	Shaheen Shah Afridi	Deathover, Yorker, Middle Stump	Yorker	54.37
		Powerplay, Short, Outside Off Stump	Bouncer	25.91
		Middleover, Yorker, Middle Stump	Yorker	24.5
		Deathover, Full Length, Out side Leg Stump	Yorker	22.5
		Powerplay, Short, Wide Outside off	Yorker	20.45
		Powerplay, Good Length, Body line	Standard Delivery	20.22
		Deathover, Full Length, Middle Stump	Standard Delivery	19.21
		Powerplay, Good Length, Outside Off Stump	Inswinger	19.18
		Powerplay, Good Length, Wide Outside off	Inswinger	18.15
		Powerplay, Good Length, Out side Leg Stump	Inswinger	17.61
12	Ravi Bishnoi	Middleover, Good Length, Body line	Legbreak	20.93
		Middleover, Good Length, Middle Stump	Slower Ball	20.4
		Deathover, Short, Outside Off Stump	Legbreak	19.68
		Middleover, Good Length, Out side Leg Stump	Slower Ball	19.29
		Deathover, Good Length, Out side Leg Stump	Slower Ball	18.32
		Powerplay, Short, Middle Stump	Legbreak	18.17
		Deathover, Good Length, Middle Stump	Legbreak	18.14
		Deathover, Good Length, Outside Off Stump	Googly	18.14
		Middleover, Good Length, Outside Off Stump	Legbreak	17
		Powerplay, Good Length, Wide Outside off	Googly	16.63
13	Adam Zampa	Middleover, Full Length, Outside Off Stump	Slower Ball	45.67
		Middleover, Good Length, Wide Outside off	Slower Ball	43.27
		Middleover, Good Length, Outside Off Stump	Slower Ball	34.83
		Middleover, Full Length, Out side Leg Stump	Standard Delivery	33.66
		Deathover, Full Length, Middle Stump	Standard Delivery	32.39
		Middleover, Full Length, Wide Outside off	Slower Ball	31.99
		Deathover, Short, Middle Stump	Flatter	29.94
		Deathover, Full Length, Out side Leg Stump	Standard Delivery	27.76
		Middleover, Short, Wide Outside off	Slower Ball	27.15

		Middleover, Short, Middle Stump	Slower Ball	24.13
14	**Tim Southee**	Middleover, Short, Middle Stump	Bouncer	31.64
		Middleover, Full Length, Outside Off Stump	Slower Ball	28.14
		Deathover, Full Length, Middle Stump	Slower Ball	25.97
		Deathover, Short, Out side Leg Stump	Standard Delivery	24.92
		Middleover, Short, Outside Off Stump	Slower Ball	22.99
		Deathover, Short, Wide Outside off	Standard Delivery	21.16
		Deathover, Full Length, Wide Outside off	Off-cutter	20.94
		Powerplay, Yorker, Wide Outside off	Outswinger	20.56
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	18.58
		Deathover, Full Length, Outside Off Stump	Standard Delivery	18.52
15	**Akeal Hosein**	Powerplay, Back of a Length, Body line	Inswinger	31.95
		Powerplay, Full Length, Body line	Outswinger	28.3
		Powerplay, Full Length, Out side Leg Stump	Inswinger	26.56
		Middleover, Full Length, Body line	Inswinger	24.53
		Powerplay, Good Length, Middle Stump	Standard Delivery	22.21
		Powerplay, Good Length, Wide Outside off	Standard Delivery	20.99
		Powerplay, Back of a Length, Out side Leg Stump	Flatter	20.75
		Powerplay, Full Length, Wide Outside off	Inswinger	19.72
		Middleover, Full Length, Out side Leg Stump	Slower Ball	18.64
		Middleover, Full Length, Middle Stump	Flatter	18.29
16	**Anrich Nortje**	Middleover, Full Length, Out side Leg Stump	Bouncer	22.29
		Middleover, Short, Middle Stump	Bouncer	21.26
		Deathover, Good Length, Out side Leg Stump	Standard Delivery	20.84
		Middleover, Good Length, Out side Leg Stump	Standard Delivery	20.01
		Middleover, Good Length, Middle Stump	Slower Ball	17.92
		Middleover, Yorker, Middle Stump	Standard Delivery	17.92
		Middleover, Full Length, Body line	Standard Delivery	17.82
		Middleover, Short, Outside Off Stump	Bouncer	17.17
		Middleover, Good Length, Body line	Standard Delivery	17.16
		Middleover, Good Length, Outside Off Stump	Standard Delivery	17.04
17	**Fazalhaq Farooqi**	Deathover, Short, Out side Leg Stump	Bouncer	33.53
		Deathover, Short, Outside Off Stump	Standard Delivery	27.31
		Deathover, Full Length, Outside Off Stump	Standard Delivery	22.24
		Powerplay, Full Length, Wide Outside off	Slower Ball	21.24
		Powerplay, Good Length, Wide Outside off	Off-cutter	20.11
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	19.46
		Powerplay, Short, Middle Stump	Standard Delivery	18.86
		Deathover, Full Length, Wide Outside off	Slower Ball	18.15

		Powerplay, Good Length, Middle Stump	Standard Delivery	17.13
		Powerplay, Full Length, Out side Leg Stump	Standard Delivery	16.53
18	**Josh Hazlewood**	Powerplay, Good Length, Middle Stump	Inswinger	33.76
		Powerplay, Good Length, Outside Off Stump	Standard Delivery	30.6
		Powerplay, Good Length, Wide Outside off	Inswinger	29.79
		Powerplay, Back of a Length, Out side Leg Stump	Off-cutter	27.25
		Powerplay, Good Length, Body line	Standard Delivery	25.73
		Powerplay, Full Length, Outside Off Stump	Standard Delivery	24.37
		Powerplay, Short, Middle Stump	Slower Ball	21.7
		Powerplay, Good Length, Out side Leg Stump	Standard Delivery	20.38
		Powerplay, Back of a Length, Wide Outside off	Standard Delivery	20.11
		Deathover, Yorker, Outside Off Stump	Off-cutter	19.83
19	**Mitchell Santner**	Middleover, Full Length, Out side Leg Stump	Bouncer	20.33
		Middleover, Full Length, Outside Off Stump	Standard Delivery	15.59
		Middleover, Short, Wide Outside off	Bouncer	14.19
		Middleover, Good Length, Wide Outside off	Outswinger	13.41
		Middleover, Full Length, Middle Stump	Off-cutter	12.24
		Middleover, Full Length, Wide Outside off	Outswinger	11.43
		Deathover, Good Length, Wide Outside off	Slower Ball	11.29
		Middleover, Full Length, Body line	Standard Delivery	8.67
		Powerplay, Short, Outside Off Stump	Slower Ball	8.5
		Deathover, Full Length, Outside Off Stump	Standard Delivery	8.3

Table # 5 presents the top ten optimal bowling state-action combinations for nineteen T20 bowlers, identified using the proposed Markov Decision Process (MDP)-based Q-learning model. Each state is represented by a combination of match phase (Powerplay, Middleover, and Deathover), delivery length, and bowling line. At the same time, the associated action denotes the bowling variation that yields the highest expected reward (Q-value). Higher Q-values indicate more strategically effective bowling decisions under specific match conditions.

The results demonstrate that optimal bowling strategies vary considerably across bowlers and are strongly influenced by match context. A dominant trend is observed in the Deathover phase, where fast bowlers consistently achieve their highest Q-values through defensive and wicket-taking deliveries such as yorkers, slower balls, bouncers, and seam variations. For example, Arshdeep Singh recorded his highest Q-value of 43.34 with a slower ball delivered at full length wide outside off during death overs, while Jofra Archer achieved a Q-value of 46.10 with a yorker targeting outside off stump in the same phase. Similarly, Shaheen Shah Afridi produced one of the highest values overall (54.37) using a yorker aimed at middle stump during death overs, highlighting the effectiveness of yorkers in restricting scoring and increasing wicket-taking probability under pressure. Among spin bowlers, distinct strategic preferences were observed in the Middleover and Deathover phases. Rashid Khan achieved the highest Q-value in the dataset (64.77) through a googly bowled on a good length outside off stump during death overs, emphasizing the strategic value of deceptive spin deliveries in high-pressure situations. Likewise, Gudakesh Motie, Ravi Bishnoi, and Adil Rashid frequently showed high Q-values for spin-based variations such as orthodox spin, legbreak, googly, and flatter deliveries, particularly when targeting outside-off and wide lines.

Several bowlers demonstrated strong specialization in the Powerplay phase. Nathan Ellis, Josh Hazlewood, and Akeal Hosein generated their highest rewards in early overs through disciplined line-length combinations combined with swing or slower-ball variations. For instance, Josh Hazlewood's highest Q-value (33.76) was associated with an inswinger delivered on a good length targeting middle stump, reflecting his effectiveness in exploiting new-ball movement. A recurring pattern across multiple bowlers is the frequent occurrence of wide outside-off lines, full-length deliveries, and variation-based actions such as slower balls, yorkers, and cutters. These patterns suggest that the reinforcement learning model identifies bowling strategies that maximize reward by balancing wicket-taking opportunities with run containment across different innings phases. Overall, the findings confirm that the proposed Q-learning framework effectively captures bowler-specific tactical intelligence by learning optimal actions under varying match states. The results provide practical insights for captains, analysts, and team selectors by identifying not only statistically successful bowlers but also those with superior context-aware decision-making capabilities.

Table # 6: Summary Statistics and Ranking of Bowlers Using Q-Value Metrics and State-Action Space in the Markov Decision Process Framework

Rank	Bowler	Number of State Action Pairs	Mean-Q	Std-Q	Max-Q	Min-Q	Median-Q
1	Arshdeep Singh	95	20.883	8.459	43.344	2.953	18.429
2	Rashid Khan	73	22.565	13.452	64.769	-7.029	22.959
3	Gudakesh Motie	40	15.952	7.019	34.527	6.092	14.054
4	Mustafizur Rahman	86	16.802	8.309	35.942	-5.292	16.877
5	Axar Patel	57	16.553	9.976	36.414	-4.940	16.344
6	Adil Rashid	47	20.415	11.570	36.958	-5.897	22.485
7	Jofra Archer	42	14.707	8.331	46.095	2.792	12.946
8	Nathan Ellis	76	13.608	9.743	39.907	-3.293	14.932
9	Lockie Ferguson	73	11.668	6.869	30.189	-6.595	11.400
10	Maheesh Theekshana	55	13.614	11.876	37.387	-14.992	16.239
11	Shaheen Shah Afridi	67	10.854	9.095	54.373	-7.332	10.359
12	Ravi Bishnoi	74	10.253	5.052	20.932	-3.524	10.483
13	Adam Zampa	79	12.192	13.698	45.665	-23.652	13.094
14	Tim Southee	71	8.736	8.754	31.641	-8.269	7.543
15	Akeal Hosein	87	10.072	8.499	31.952	-6.243	8.401
16	Anrich Nortje	70	7.321	9.262	22.292	-13.885	8.822
17	Fazalhaq Farooqi	56	8.038	10.126	33.530	-10.245	9.350
18	Josh Hazlewood	60	7.616	12.072	33.758	-12.328	4.686
19	Mitchell Santner	83	5.534	5.626	20.332	-12.220	6.650

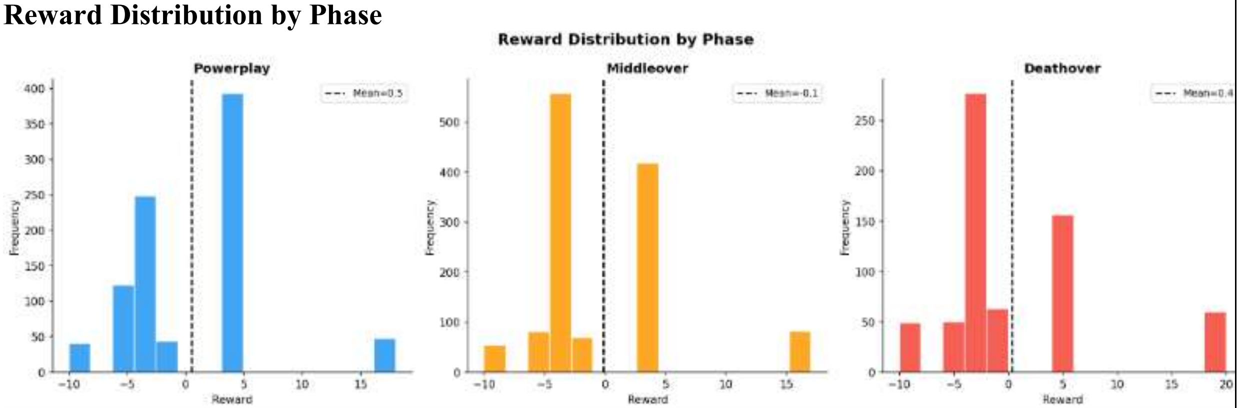
Table # 6 presents the overall performance ranking of T20I bowlers using Q-value metrics generated through the Markov Decision Process (MDP) and Q-learning framework. Higher Mean Q-values indicate more effective bowling strategies and better long-term rewards across different match situations. Rashid Khan recorded the highest Mean Q-value (22.5648), indicating exceptional strategic effectiveness, while Arshdeep Singh ranked first overall due to his high Mean Q-value (20.8831), strong Median Q-value (18.4291), and the largest number of state-action pairs (95), demonstrating consistency and adaptability across multiple bowling situations. Adil Rashid also showed a strong performance with a Mean Q-value of 20.4145. Bowlers such as Mustafizur Rahman, Axar Patel, and Gudakesh Motie achieved strong positive Mean Q-values, reflecting effective tactical bowling and reward-maximizing strategies. In contrast, Mitchell Santner (5.5337), Anrich Nortje (7.3212), and Josh Hazlewood (7.6162) recorded lower Mean Q-values, indicating comparatively lower effectiveness. The Standard Deviation values show performance variability across different states. Rashid Khan and Adam Zampa displayed higher variability, suggesting both highly rewarding and risky bowling outcomes, while Ravi Bishnoi and Mitchell Santner showed more stable but lower-impact performances. Rashid Khan also achieved the highest Max_Q value (64.7692), highlighting his ability to generate highly effective bowling outcomes in specific match situations.

Table # 7: Performance Evaluation of T20I Bowlers Across Match Phases Using Mean Q-Values

Rank	Bowler	Powerplay Mean-Q	Middle over Mean-Q	Death over Mean-Q
1	Arshdeep Singh	16.8655	20.4039	24.6719
2	Rashid Khan	1.7970	23.7827	30.3303
3	Gudakesh Motie	0.0000	13.9054	30.2768
4	Mustafizur Rahman	14.2913	20.3074	15.9292
5	Axar Patel	9.8605	23.2675	6.0698
6	Adil Rashid	16.0095	26.0344	1.1415
7	Jofra Archer	13.0063	12.0431	21.5931
8	Nathan Ellis	18.8572	3.8571	17.4149
9	Lockie Ferguson	7.9016	11.7177	15.7164
10	Maheesh Theekshana	15.8439	19.1729	-6.6618
11	Shaheen Shah Afridi	13.2401	2.8172	12.0804
12	Ravi Bishnoi	8.7513	10.9605	10.5858
13	Adam Zampa	5.3642	15.1035	7.1966
14	Tim Southee	5.3797	8.3951	13.0485
15	Akeal Hosein	15.4201	7.1050	0.4926
16	Anrich Nortje	-5.2258	13.2472	8.4481
17	Fazalhaq Farooqi	13.2935	-5.0201	12.0696
18	Josh Hazlewood	18.0135	-3.9806	4.1281
19	Mitchell Santner	0.6873	7.4643	4.1959

Table # 7 presents the phase-wise performance of T20I bowlers using Mean Q-values generated through the Q-learning framework. Higher Q-values indicate more effective bowling strategies that produced better long-term rewards such as wickets, dot balls, and run restriction, while negative Q-values represent ineffective bowling performance. In the Powerplay phase, Nathan Ellis (18.8572), Josh Hazlewood (18.0135), and Arshdeep Singh (16.8655) achieved the highest Mean Q-values, indicating strong effectiveness with the new ball and during field restrictions. In contrast, Anrich Nortje (-5.2258) showed weaker performance in this phase. During the Middle overs, spin bowlers dominated the rankings. Adil Rashid (26.0344), Rashid Khan (23.7827), and Axar Patel (23.2675) recorded the highest Mean Q-values, highlighting the importance of spin variations and controlled bowling in restricting scoring opportunities during this phase. However, Fazalhaq Farooqi (-5.0201) and Josh Hazlewood (-3.9806) were comparatively less effective. In the Death overs, Rashid Khan (30.3303), Gudakesh Motie (30.2768), Arshdeep Singh (24.6719), and Jofra Archer (21.5931) demonstrated exceptional effectiveness under high-pressure situations. Their high Q-values indicate successful execution of yorkers, slower balls, and variation-based strategies. Conversely, Maheesh Theekshana (-6.6618) struggled in the death overs.

Reward Distribution by Phase



The figure # 5 shows the reward distributions of the reinforcement learning model across the Powerplay, Middleover, and Deathover phases, with dashed lines representing the mean reward of each phase.In the Powerplay, rewards are mostly concentrated around small positive values, indicating relatively stable and effective bowling outcomes, with only a few negative rewards. The Middleover shows the widest distribution, reflecting greater tactical variation and a balance between positive and negative outcomes due to experimentation with different bowling strategies. In contrast, the Deathover displays a more polarized distribution, with many negative rewards and some high positive rewards, highlighting the high-risk, high-reward nature of bowling under scoring pressure.

Table # 8: Hybrid Ranking Model for T20 Bowlers Using Classical Metrics and Reinforcement Learnin

Rank	Name	Weighted wicket	Weighted economy rate	Weighted strike rate	Weighted bowling average	Weighted dot-ball %	Weighted boundary	Sum of all weighted classical metrics	Normaliz Q-learnin score
1	Arshdeep Singh	0.28	0.0852	0.1213	0.0806	0.0854	0.0158	0.6683	1
2	Rashid Khan	0.2178	0.1261	0.1181	0.0892	0.0803	0.0354	0.6669	0.995
3	Gudakesh Motie	0.1556	0.1011	0.1358	0.0909	0.0741	0.0288	0.5861	0.8928
4	Mustafizur Rahman	0.2333	0.0641	0.1104	0.0694	0.0729	0.0151	0.5651	0.7378
5	Axar Patel	0.14	0.0947	0.1014	0.0735	0.0847	0.0263	0.5205	0.733
6	Adil Rashid	0.1244	0.063	0.1106	0.0692	0.0836	0.0142	0.465	0.7805

7	Jofra Archer	0.1244	0.0947	0.1114	0.0782	0.0866	0.0201	0.5153	0.6117
8	Nathan Ellis	0.1556	0.0966	0.0974	0.0722	0.0866	0.0205	0.5289	0.5767
9	Lockie Ferguson	0.1556	0.1073	0.1185	0.0847	0.0829	0.0315	0.5805	0.4659
10	Maheesh Theekshana	0.14	0.1012	0.0948	0.0724	0.0762	0.0276	0.5122	0.4895
11	Shaheen Shah Afridi	0.1867	0.0462	0.1238	0.0729	0.0822	0.0037	0.5155	0.4115
12	Ravi Bishnoi	0.1556	0.0878	0.0947	0.0681	0.0845	0.0199	0.5106	0.3372
13	Adam Zampa	0.1867	0	0.1167	0.0567	0.0507	0	0.4108	0.4158
14	Tim Southee	0.1556	0.0475	0.104	0.0609	0.0639	0.0121	0.444	0.3187
15	Akeal Hosein	0.1244	0.0551	0.0701	0.0431	0.0736	0.0094	0.3756	0.3598
16	Anrich Nortje	0.1556	0.0852	0.0914	0.0656	0.0975	0.0153	0.5105	0.0793
17	Fazalhaq Farooqi	0.0778	0.069	0.0397	0.0319	0.0822	0.0123	0.3129	0.1629
18	Josh Hazlewood	0.0933	0.0845	0.0134	0.0263	0.1	0.0148	0.3324	0.0989
19	Mitchell Santner	0.0778	0.0525	0	0	0.063	0.0183	0.2115	0

Table 8 presents the hybrid ranking model, which integrates classical bowling metrics (wickets, economy rate, strike rate, bowling average, dot ball percentage, and boundary percentage) with reinforcement-learning-derived Q-scores using a 65:35 weighting scheme. The results show that Arshdeep Singh ranks first overall with the highest classical sum (0.6683) and maximum normalized Q-score (1.000), yielding a final hybrid score of 0.7844, closely followed by Rashid Khan (0.7817), who demonstrates nearly identical classical performance (0.6669) and an exceptionally high Q-score (0.9950). Gudakesh Motie ranks third with a final score of 0.6935, though a noticeable gap exists between the top two and the third position. A key insight from the hybrid model is the divergence between classical and RL-based evaluations. Adil Rashid (ranked 6th with a final score of 0.5754) has a relatively lower classical sum (0.4650) but a strong Q-score (0.7805), indicating that his strategic decision-making quality exceeds what traditional metrics alone capture, causing him to benefit significantly from RL inclusion. In contrast, Lockie Ferguson (ranked 9th with a final score of 0.5404) shows the opposite pattern: a strong classical sum (0.5805, the 4th highest among all bowlers) but a substantially lower Q-score (0.4659), suggesting that while his raw outcomes are excellent, his context-aware decision-making may be less optimal. The most striking divergence occurs with Anrich Nortje, who possesses a respectable classical sum (0.5105, comparable to mid-ranked bowlers) but an extremely low Q-score (0.0793), causing his hybrid rank to drop to 16th with a final score of 0.3596—substantially lower than his classical performance would suggest. Similarly, Josh Hazlewood (0.2507) and Mitchell Santner (0.1375) occupy the bottom two positions, with Santner recording a zero normalized Q-score, indicating the weakest performance across both evaluation dimensions. Overall, the hybrid model demonstrates that classical metrics alone are insufficient for comprehensive bowler evaluation, as the RL-derived Q-score captures tactical intelligence and context-aware decision-making that traditional statistics miss, providing a more complete and nuanced assessment of bowler quality for team selection and strategic planning.

Conclusion

This study developed an MDP-based Q-learning framework for optimizing bowling strategies in T20 cricket using real ball-by-ball data, successfully modeling bowling as a sequential decision-making problem, learning optimal policies directly from historical match data, evaluating these policies across different match phases, and integrating traditional bowling metrics with RL-derived Q-scores into a hybrid bowler ranking framework. The proposed MDP framework effectively captured the sequential nature of T20 bowling by defining states based on match phase, delivery length, and bowling line, with actions representing bowling variations such as yorkers, slower balls, bouncers, and spin deliveries, and the Q-learning algorithm successfully learned optimal policies without requiring pre-specified transition probabilities. The evaluation of learned policies across match phases revealed distinct optimal strategies: in the Powerplay (overs 1-6), bowlers such as Nathan Ellis (Mean-Q = 18.86), Josh Hazlewood (18.01), and Arshdeep Singh (16.87) achieved the highest Q-values, indicating that disciplined line-length combinations and swing-based variations are most effective during field restrictions; during the Middle overs (7-15), spin bowlers dominated with Adil Rashid (26.03), Rashid Khan (23.78), and Axar Patel (23.27) recording the highest Mean Q-values, highlighting the importance of spin variations in restricting scoring; and in the Death overs (16-20), Rashid Khan (30.33), Gudakesh Motie (30.28), Arshdeep Singh (24.67), and Jofra Archer (21.59) demonstrated exceptional effectiveness through yorkers, slower balls, and variation-based strategies, though the polarized reward distribution confirmed the high-risk, high-reward nature of death-over bowling. The analysis of optimal state-action combinations revealed that fast bowlers achieved their highest Q-values through yorkers and slower balls in death overs, spin bowlers achieved optimal rewards through googlies, legbreaks, and orthodox deliveries targeting outside-off lines, and powerplay specialists generated highest rewards through inswingers and disciplined line-length combinations with the new ball. The hybrid ranking model integrated classical metrics (wickets, economy rate, strike rate, bowling average, dot ball percentage, boundary percentage) with normalized Q-scores using a 65:35 weighting scheme, with Arshdeep Singh ranking first overall (0.7844) with the highest classical sum (0.6683) and maximum Q-score (1.000), Rashid Khan ranking second (0.7817) with nearly identical classical performance (0.6669) and exceptional Q-score (0.9950), Adil Rashid benefiting most from RL inclusion by rising from 10th in classical sum to 6th in hybrid rank due to a strong Q-score (0.7805), and Anrich Nortje showing the largest divergence with a respectable classical sum (0.5105) but a very low Q-score (0.0793), dropping to 16th in hybrid rank. The study makes several academic contributions, including the first comprehensive MDP formulation of T20 bowling that includes match phase, delivery length, bowling line, and batter context as state variables; the first application of Q-learning to learn optimal delivery-type policies directly from real ball-by-ball T20 match data rather than simulated environments; phase-specific evaluation of learned bowling policies across powerplay, middle, and death overs; and a novel hybrid ranking framework demonstrating that strategic decision-making quality provides information not captured by traditional statistics. Practically, coaches and analysts can use the identified optimal state-action combinations to design phase-specific bowling strategies, team selectors can utilize the hybrid ranking model to evaluate bowlers based on both traditional outcomes and strategic decision-making quality, captains can make data-informed decisions about bowling changes based on Q-values, and player development programs can use the framework to identify weaknesses in context-aware decision-making. However, limitations include the Markov property assumption which may not fully capture long-term dependencies such as batter learning across multiple deliveries, the reward function which may not perfectly reflect all strategic objectives, and the focus on T20 International matches from a single year which may limit generalizability. Future work should extend the framework to deep reinforcement learning for handling larger state spaces, multi-agent reinforcement learning to model both bowler and batter as learning agents, real-time strategy recommendation systems for captains, cross-format comparison across ODI and Test cricket, incorporation of additional context such as batter weakness profiles and field placements, and longitudinal validation studies to examine whether bowlers with higher Q-scores demonstrate greater improvement or longer careers. In conclusion, this study demonstrates that reinforcement learning, specifically MDP-based Q-learning, provides a powerful and practical framework for optimizing bowling strategies in T20 cricket, and the findings confirm that bowlers with high strategic intelligence measured through Q-scores may be systematically undervalued by traditional metrics alone, suggesting that RL-based evaluation should complement conventional statistical analysis in future cricketing decision-making.

References

[1] Robel, M., Khan, M. A. R., Ahammad, I., Alam, M. M., & Hasan, K. (2024). Cricket players selection for national team and franchise league using machine learning algorithms. Cloud Computing and Data Science, 108-139.

[2] Premkumar, P., Chakrabarty, J. B., & Chowdhury, S. (2020). Key performance indicators for factor score based ranking in One Day International cricket. IIMB Management Review, 32(1), 85-95.

[3] Vohra, M., & Gordon, G. S. (2021). Markov Cricket: Using Forward and Inverse Reinforcement Learning to Model, Predict And Optimize Batting Performance in One-Day International Cricket. arXiv preprint arXiv:2103.04349.

[4] Ameen, M. R., Islam, A., Miah, A. S. M., Rafat, M. S., & Shin, J. (2025). Teaching Machine Learning Through Cricket: A Practical Engineering Education Approach. arXiv preprint arXiv:2510.22392.

[5] Williams, R. J. (2007). Reinforcement learning and Markov decision processes. CSG220, Spring.

[6] Khan, Q. W. (2024). Exploring Markov decision processes: a comprehensive survey of optimization applications and techniques. Igmin Research, 2(7), 508-517

[7] Clifton, J., & Laber, E. (2020). Q-learning: Theory and applications. Annual Review of Statistics and Its Application, 7(1), 279-301.

[8] Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. IEEE access, 7, 133653-133667.

[9] Bozkus, T., & Mitra, U. (2024). Multi-Timescale Ensemble Q -Learning for Markov Decision Process Policy Optimization. IEEE Transactions on Signal Processing, 72, 1427-1442.

[10] Chen, Y., Zhao, Q., & Swami, A. (2008). Distributed spectrum sensing and access in cognitive radio networks with energy constraint. IEEE transactions on signal processing, 57(2), 783-797.

[11] Zhang, F., & Lau, V. K. (2014). Closed-form delay-optimal power control for energy harvesting wireless system with finite energy storage. IEEE Transactions on Signal Processing, 62(21), 5706-5715.

[12] Chai, S., & Lau, V. K. (2020). Online trajectory and radio resource optimization of cache-enabled UAV wireless networks with content and energy recharging. IEEE Transactions on Signal Processing, 68, 1286-1299.

[13] Bertsekas, D. (2019). Reinforcement learning and optimal control (Vol. 1). Athena Scientific..

[14] Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1, pp. 9-11). Cambridge: MIT press.

[15] Hussain, A., Arshad, S., & Hassan, A. (2024). RunsGuard framework: Context aware cricket game strategy for field placement and score containment. Applied Sciences, 14(6), 2500.

[16] Bhavsar, J. P., Shah, S., & Khara, S. (2025, November). Advancing Cricket Strategy With Artificial Intelligence: A Review Of Predictive And Strategic Models For Performance Optimization. In 2025 2nd Global AI Summit-International Conference on Artificial Intelligence and Emerging Technology (AI Summit) (pp. 517-521). IEEE.

[17] Van Roy, M., Robberechts, P., Yang, W. C., De Raedt, L., & Davis, J. (2023). A Markov framework for learning and reasoning about strategies in professional soccer. Journal of Artificial Intelligence Research, 77, 517-562.

[18] Ding, N., Takeda, K., & Fujii, K. (2022). Deep reinforcement learning in a racket sport for player evaluation with technical and tactical contexts. IEEE Access, 10, 54764-54772.

[19] Bowala, S. M., Manage, A. B., & Scariano, S. M. (2021). Modeling T20I cricket bowling effectiveness: A quantile regression approach with a Bayesian extension. Journal of Sports Analytics, 7(3), 197-221.

[20] Jamil, M., Kerruish, S., Beato, M., & McErlain-Naylor, S. A. (2022). The effects of bowling lines and lengths on the spatial distribution of successful power-hitting strokes in international men's one-day and T20 cricket. Journal of sports sciences, 40(19), 2208-2216.

[21] Van Roy, M., Robberechts, P., Yang, W. C., De Raedt, L., & Davis, J. (2023). A Markov framework for learning and reasoning about strategies in professional soccer. Journal of Artificial Intelligence Research, 77, 517-562.

[22] Sandholtz, N., & Bornn, L. (2020). Markov decision processes with dynamic transition probabilities: An analysis of shooting strategies in basketball.

[23] Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. IEEE access, 7, 133653-133667.

[24] Samsuden, M. A., Diah, N. M., & Rahman, N. A. (2019, October). A review paper on implementing reinforcement learning technique in optimising games performance. In 2019 IEEE 9th international conference on system engineering and technology (ICSET) (pp. 258-263). IEEE.

[25] Kandasamy, C. A., Harishjayaraj, S., Geetha, R., Kanimozhi, J. K., Loganathan, E., & Prakash, N. (2025, May). Optimizing Sports Strategies: Achieving Convergence using SARSA and Q-Learning for Dynamic Decision-Making. In 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN) (pp. 1622-1626). IEEE.

[26] Hoffmeister, S. (2019). Sport-Strategy Optimization with Markov Decision Processes (Doctoral dissertation).

[27] Hoffmeister, S., & Rambau, J. (2020, December). Strategy Optimization in Sports via Markov Decision Problems. In Modeling, Simulation and Optimization of Complex Processes HPSC 2018: Proceedings of the 7th International Conference on High Performance Scientific Computing, Hanoi, Vietnam, March 19-23, 2018 (pp. 277-322). Cham: Springer International Publishing.

[28] Manju, M. K., & Philip, A. O. (2023). Novel method for ranking batsmen in Indian Premier League. Data Science and Management, 6(3), 158-173.

[29] Prakash, C. D., & Verma, S. (2022). A new in-form and role-based deep player performance index for player evaluation in T20 cricket. Decision Analytics Journal, 2, 100025.

[30] Vobhilineni, R. C., Anthireddy, S. R., Maddula, S., Pandi, V. R., & Sathyan, S. (2025, April). A Stacking Approach for T20 Cricket Performance Forecasting Using Machine Learning. In 2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA) (pp. 1-6). IEEE.

[31] Gour, P. N., & Khan, M. F. (2024). Utilizing machine learning for comprehensive analysis and predictive modelling of ipl-t20 cricket matches. Indian Journal of Science and Technology, 17(7), 592-597.

[32] Huang, J., Zhang, Y., Xu, M., Lv, Y., Zhang, J., & Shafieezadeh, M. M. (2025). Hybrid FAHP-FTOPSIS methodology for objective football player selection and ranking. Scientific Reports, 15(1), 27913.

[33] Bellman, R. I. C. H. A. R. D. (1957). Dynamic programming, princeton univ. Press Princeton, New Jersey, 39.

[34] Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1, pp. 9-11). Cambridge: MIT press.

[35] Puterman, M. L. (1990). Markov decision processes. Handbooks in operations research and management science, 2, 331-434.

[36] Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine learning, 8(3), 279-292.