Reducing Handover and Ping-Pong Events in LTE Networks via Q-Learning and Subtractive Clustering

K M Istiaque
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
istiaque20@iut-dhaka.edu

Shafayet Sadik Sowad
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
shafayetsadik@iut-dhaka.edu

M M Sadman Shafi
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
sadmanshafi@iut-dhaka.edu

Mohammad Tawhid Kawser
Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur, Bangladesh
kawser@iut-dhaka.edu

***Abstract*— Efficient handover (HO) management is critical for maintaining service continuity in dense LTE networks, especially in urban environments with frequent signal fluctuations. This paper presents a Q-learning-based HO optimization framework that leverages a real-world drive test dataset collected in Dhaka, Bangladesh. The framework integrates subtractive clustering to discretize radio conditions into 20 representative states and optimizes state-specific Handover Margin (HOM) and Time-to-Trigger (TTT) values through an extensive parameter search over 50,000 combinations. A carefully designed reward function penalizes unnecessary and ping-pong handovers, guiding the RL agent toward stable and efficient HO behavior. Experimental results demonstrate a 13% reduction in total handovers and a 76% reduction in ping-pong events on the test route. Additionally, the proposed method achieves a decision latency of just 0.016 ms — far outperforming conventional heuristic-based approaches. These results validate the potential of reinforcement learning for real-time, scalable, and adaptive mobility management in LTE networks, with future applicability to 5G and beyond.**

Keywords— Q-Learning, Subtractive Clustering, LTE, Handover, Ping-pong, Time-to-trigger, Latency

I. Introduction

The rapid growth of mobile data traffic and increasingly dense cellular deployments have made efficient mobility management a cornerstone of Long Term Evolution (LTE) networks. Handover procedures, which allow user equipment (UE) to maintain continuous service while crossing cell boundaries, are vital for ensuring Quality of Service (QoS) and minimizing interruptions.

Traditional handover mechanisms rely on fixed thresholds and timers, which often fail to adapt to dynamic radio environments with fluctuating signal quality and mobility patterns [1]. This can cause excessive handovers, ping-pong effects, and signaling overhead. Ping-pong handovers—frequent, unnecessary toggling between cells—are particularly problematic, introducing service disruption, latency, and resource wastage.

Recent advances in machine learning[2], particularly reinforcement learning (RL), offer promising solutions by enabling adaptive, context-aware handover strategies [3]. RL agents can learn optimal policies through interaction with the environment, balancing handover frequency and connection quality while minimizing ping-pong events without requiring explicit modeling of complex radio dynamics [4].

In this work, we propose a Q-learning based handover optimization framework trained on a unique LTE mobility dataset from drive tests in Dhaka, Bangladesh. The dataset contains real-world measurements of radio parameters and handover events in a dense urban network, a context underrepresented in public datasets. Subtractive clustering discretizes the radio environment into meaningful states, and a reward function aligned with practical metrics includes penalties to discourage ping-pong events.

Experiments show that the learned policies outperform conventional heuristics, reducing handovers and eliminating ping-pong events while maintaining QoS. These results highlight the potential of data-driven, region-specific mobility optimization for LTE networks.

II. Literature Review

The challenge of handover (HO) optimization lies in ensuring seamless connectivity as user equipment (UE) moves between different base stations (BS). Frequent and unnecessary handovers—especially ping-pong events—can significantly degrade network performance and user experience. The ping-pong effect, in particular, results in excessive signaling and increased power consumption, ultimately leading to Quality of Service (QoS) deterioration [4].

Traditional HO algorithms rely on static thresholds for parameters such as Reference Signal Received Power (RSRP) and Time-to-Trigger (TTT). However, these fixed heuristics often fail to adapt to real-time, dynamic radio environments, resulting in suboptimal decisions and frequent handovers [5], [6].

In response to these limitations, Reinforcement Learning (RL) has emerged as a promising approach to dynamically adjust handover parameters such as Handover Margin (HOM) and TTT [7], [8]. Among various RL techniques, Q-learning has gained popularity for its ability to optimize handover decisions without prior knowledge of the environment. It enables agents to learn optimal policies based on feedback, taking into account parameters like signal quality, mobility, and interference [9].

Several studies have extended Q-learning frameworks by integrating additional techniques. For instance, hybrid models combining Q-learning with subtractive clustering or fuzzy logic have been proposed to improve decision granularity and HO prediction accuracy [10]. These models enhance adaptability to heterogeneous radio environments by allowing finer control over threshold adjustments.

Furthermore, in heterogeneous networks (HetNets), Q-learning has been applied to dynamically adjust handover thresholds, thereby reducing ping-pong events and HO failures [11]. Other efforts integrate RL with optimization theory to jointly optimize base station selection and beamforming, improving throughput while maintaining low latency [12].

The use of Deep Q-Networks (DQN) has also been explored to manage real-time HO decisions in more complex and denser 5G environments. These networks incorporate a broad set of input features, including SINR, cell load, and UE speed, enabling more robust and adaptive handover mechanisms [13].

More recent proposals introduce intelligent mobility management mechanisms, where reinforcement learning is augmented with Kalman filters to predict signal trends from both serving and neighboring cells. By applying ε-greedy policies [14], these systems dynamically adjust TTT and hysteresis margins according to mobility patterns and signal quality [8].

To evaluate such algorithms, simulation environments like NS-3 have been widely used in prior works. NS-3-based studies simulate realistic 5G and LTE scenarios to benchmark RL-driven handover optimization under various mobility and load conditions [7]. While these simulations consistently show improved throughput, lower handover failures, and enhanced QoS, many of these studies remain theoretical. The transition from simulation to real-world deployment—where unpredictable behaviors, hardware limitations, and real radio propagation models come into play—still presents open challenges.

III. METHODOLOGY

This section outlines the comprehensive methodology used to model, optimize, and evaluate handover decisions in a 4G/LTE cellular network using reinforcement learning. The proposed framework begins with real-world data collection through drive tests, followed by state space construction via subtractive clustering, action-reward design using domain-specific mobility logic, and Q-learning-based policy optimization. The methodology concludes with a parameter search across handover margins (HOM), time-to-trigger (TTT), and clustering weights to identify an optimal policy that minimizes unnecessary handover.

Figure 1 illustrates a flowchart depicting the entire process of the study, starting with Drive Test Data and proceeding through Subtractive Clustering. It continues with Q-Learning Training using HOM, TTT, Weights, and concludes with Final Evaluation in the testing phase.

A. Drive Test and Data Collection

To support the development of a learning-based handover optimization framework, a novel LTE mobility dataset was collected through controlled drive tests in Dhaka, Bangladesh. The measurements were acquired using professional tools—XCAL-M software, Samsung Exynos-based UE, and GPS modules—over a 13 km urban route from Le Meridien (Uttara) to BRAC University during peak hours. The environment reflects dense LTE deployment and varying mobility conditions.

The dataset captures key Layer 3 radio parameters including RSRP, RSRQ, CINR, Cell IDs, and handover events. Tests were conducted using a Multiple Site Verification (MSV) approach over three separate sessions, covering approximately 26 km in total and recording over 100 handover events.

The real-world dataset was collected over three separate days of drive testing; among them, data from Day 1 and Day 2 were used in this study [19]. The larger Day 1 dataset—approximately twice the size of Day 2—was used for training, while the Day 2 data served exclusively as the test set. This dataset enabled region-specific modeling of radio dynamics and handover behavior, forming the foundation for state abstraction and policy learning in the reinforcement learning framework.

B. Data Preprocessing and Feature Engineering

Before model development, the drive test data was preprocessed to ensure temporal consistency and valid signal features. Timestamps were converted to milliseconds for accurate sequencing of handover events, while entries with missing or invalid RSRP, RSRQ, or CINR values were removed to reduce noise. RSRP values were standardized to dBm using LTE mapping to enhance consistency across signal metrics and improve reward formulation. The final feature set—comprising serving and neighboring cell RSRP (in dBm), serving cell RSRQ and CINR, and neighboring cell RSRQ—captured essential radio link conditions. These were used for state abstraction via subtractive clustering and guided the design of the reinforcement learning reward function.

C. State Abstraction via Subtractive Clustering

To discretize the continuous radio environment into meaningful mobility contexts, subtractive clustering was employed for state space construction. This unsupervised clustering method is well-suited for sparse and non-uniform datasets, as it does not require a predefined number of clusters and dynamically selects representative data points based on local density.

Each data instance was represented as a five-dimensional feature vector comprising serving and neighboring cell RSRP (in dBm), serving cell RSRQ and CINR, and neighboring cell RSRQ. Features were normalized using Min-Max scaling to balance their influence in distance-based similarity computation.

The clustering algorithm generated 20 distinct clusters, each representing a unique radio condition profile or "state". These states reflect the heterogeneous signal conditions experienced

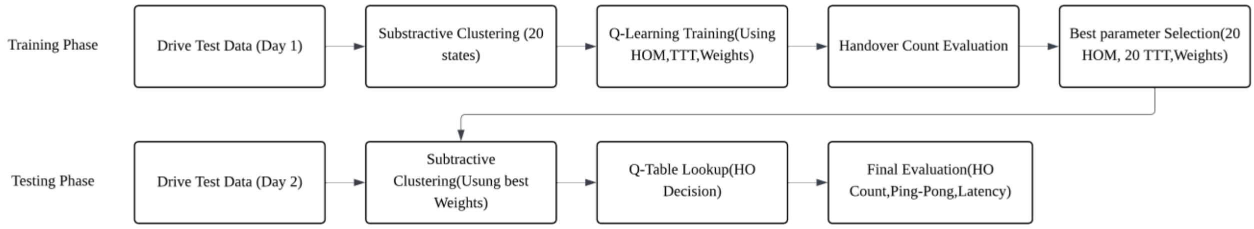

Figure 1: Methodological Pipeline

by user equipment during mobility, from strong-serving-cell dominance to interference-prone cell-edge zones.

Compared to grid-based or heuristic discretization, subtractive clustering provides a data-driven, topology-aware abstraction [15].It captures nuanced patterns in signal interactions and enables the reinforcement learning agent to generalize over similar conditions rather than overfitting to raw measurements. Figure 2 illustrates the clustering process for state discretization used in our proposed model.

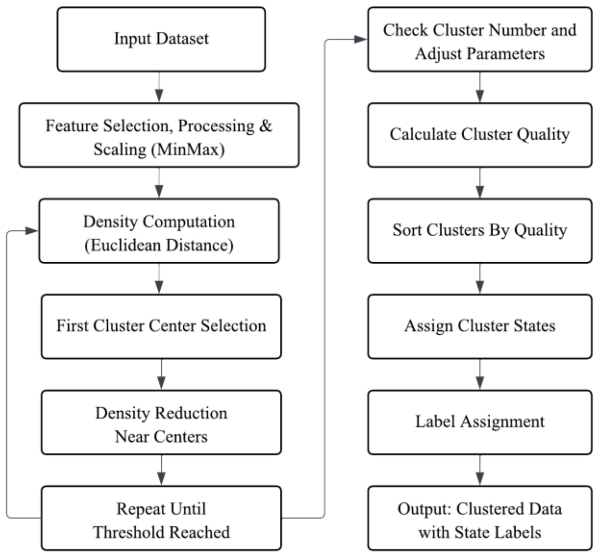

Figure 2: Clustering Process for State Discretization

D. *Action and Reward Design for Q-Learning*

The agent operates with two discrete actions: to trigger a handover (HO) or maintain the current serving cell connection (No_HO). This binary action space reflects practical decision-making in LTE mobility management and simplifies the policy optimization problem.

The reward function is designed to reflect signal quality trade-offs and operational constraints. A positive reward is assigned when a handover occurs and results in improved RSRP, while a penalty is applied for handovers that offer no improvement or introduce instability. Conversely, retaining the current cell yields a small positive reward if justified by signal conditions and a penalty if it results in missed opportunities.

To discourage ping-pong handovers, additional negative rewards are applied when consecutive handovers occur within the Time-to-Trigger (TTT) window. All reward values are clipped within [–3, +3] to stabilize learning dynamics.

This reward structure aligns with real-world operator objectives: optimizing handover decisions to reduce signaling overhead, avoid unnecessary transitions, and improve user experience.

To illustrate the underlying logic, Table 1 summarizes the key decision cases and corresponding reward signals used in training.

Table 1: Reward Logic and Action Mapping in Q-Learning

Condition	**Action Taken**	**Reward / Penalty**	**Comment**
Neighbor RSRP > Serving RSRP + HOM	HO	+4	Desirable handover to a better cell
Neighbor RSRP ≤ Serving RSRP + HOM	HO	–0.7	Unnecessary handover with no benefit
Neighbor RSRP > Serving RSRP + HOM	No_HO	–4	Missed opportunity, leads to poor signal
Neighbor RSRP ≤ Serving RSRP + HOM	No_HO	+0.7	Correct decision to retain serving cell
Repeated HO within TTT duration	HO	–10	Ping-pong handover penalty
Any reward outcome	—	Clipped to [–3, +3]	Stabilizes training and prevents outliers

E. *Q-Learning Implementation and Decision Mapping*

The framework uses Q-learning to derive an optimal handover policy over a discretized state space formed from radio conditions [16]. Each state-action pair—HO or No_HO—is evaluated to maximize cumulative rewards under specific HOM and TTT settings. The four-dimensional Q-table is indexed by state (via subtractive clustering), HOM value (in dB), TTT index (in ms), and action. An ε-greedy strategy guides exploration, while rewards are assigned based on RSRP delta and adherence to TTT constraints. Q-values are iteratively updated using the Bellman equation, enabling policy refinement over successive episodes [17].

This structured mapping helps the agent converge toward stable and context-aware handover behavior. The use of state-specific HOM and TTT values further enhances granularity, allowing the policy to adapt to heterogeneous signal conditions across the coverage area. Figure 3 depicts the Q-learning framework, outlining the training and testing phases for decision-making in handover (HO) scenarios.

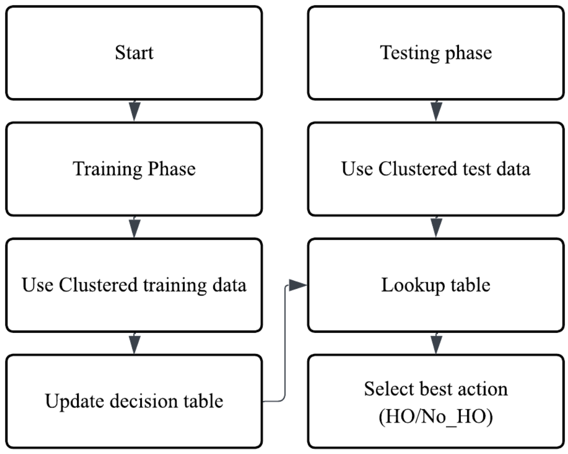

Figure 3: Q-learning Framework

F. Parameter Optimization Strategy

To enhance handover efficiency and reduce unnecessary transitions in real-world LTE environments, this work incorporates a dedicated parameter optimization routine targeting three core components: Handover Margin (HOM), Time-to-Trigger (TTT), and feature weights for subtractive clustering. These jointly shape the state abstraction, reward evaluation, and triggering logic of the Q-learning framework.

A large search space was systematically defined: HOM values ranged from 0.0 to 2.0 dB, TTT values spanned from 40 ms to 800 ms, and five clustering feature weights were varied to reflect differing priorities for signal quality metrics (e.g., RSRP, RSRQ, CINR). Parameter maps were structured for each of the 20 Q-learning states, enabling state-specific granularity in HO control.

To comprehensively explore this space, the system was run over 50,000 unique combinations of HOM, TTT, and feature weights. For each combination, the training dataset (Day 1) was clustered using the associated weights, and Q-learning was applied over 1000 episodes to learn the optimal handover policy.

The learned policy was then evaluated on an independent test dataset (Day 2), and the total number of handovers was recorded as the primary performance metric. The optimal parameter set was selected based on the lowest resulting HO count.

This full-scale optimization required approximately 56 hours and 37 minutes of simulation time, conducted under real-world mobility conditions. The feature weights shaped the clustering granularity to reflect heterogeneous signal environments, while the dynamic HOM and TTT values adapted policy aggressiveness across network states. As a result, the final learned policy significantly reduced both total and ping-pong handovers, offering a robust and data-driven alternative to conventional static handover configurations.

IV. Results and Discussion

This section presents the performance evaluation of the proposed Q-learning-based handover (HO) optimization framework using the real-world LTE dataset collected through drive tests. The analysis focuses on two core objectives:

(1) minimizing the overall number of handovers by selecting an optimal configuration of HOM, TTT, and clustering weights

(2) eliminating or significantly reducing ping-pong handovers while enabling ultra-fast decision-making compared to conventional heuristic-based methods, thereby improving both efficiency and responsiveness.

A. Handover Count Reduction

In the baseline scenario, based on the XCAL-M logs and existing operator-defined handover configurations, 92 handovers were recorded over the Day 2 route. After applying the learned policy from the Q-learning agent, this number was successfully reduced to 80 handovers, representing a 13% decrease.

Figure 4 illustrates a bar comparison of the handover counts: the Traditional Heuristic Model recorded 92 handovers, while the Proposed Model achieved 80 handovers. This visual representation highlights the reduction in handover frequency, with the Q-learning based approach demonstrating a more efficient distribution of handover events, leading to smoother transition behavior.

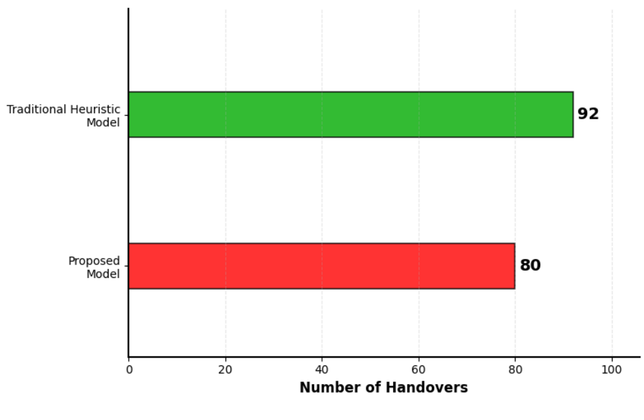

Figure 4: Handover Comparison Between Traditional and Proposed Models

B. Optimization Loop Summary and Best Parameter Configuration

To achieve the result above, the system evaluated 50,000 unique parameter combinations, covering varied handover margins, time-to-trigger values, and five feature weights for clustering. The full process took approximately 56 hours 37 minutes of simulation runtime. Each combination was trained on the Day 1 dataset and tested on Day 2, with handover count used as the evaluation metric.

The best-performing configuration, which yielded 80 handovers, is summarized in Table 2.

Table 2. Optimized Parameter Configuration for Minimum Handover Count

Parameter	Value
HOM Range (in dB)	{1: 0.05, 2: 0.1, 3: 0.24, 4: 0.3, 5: 0.4, 6: 0.3, 7: 0.35, 8: 0.4, 9: 0.8, 10: 0.94, 11: 0.55, 12: 0.6, 13: 0.65, 14: 1.36, 15: 0.75, 16: 0.8, 17: 0.85, 18: 0.9, 19: 0.95, 20: 1.0}
TTT Range (in ms)	{1: 40, 2: 78, 3: 146, 4: 164, 5: 222, 6: 230, 7: 268, 8: 306, 9: 364, 10: 412, 11: 420, 12: 458, 13: 496, 14: 544, 15: 572, 16: 610, 17: 648, 18: 686, 19: 724, 20: 762}
Clustering Weights	[-3.0, -3.0, -0.1, 5.5, 6.0]

C. *Ping-Pong Handover Mitigation*

A critical enhancement of the proposed framework lies in the substantial reduction of ping-pong handovers, defined as unnecessary, back-and-forth handovers between neighboring cells within a short duration. The handover comparison in Figure 5 shows the Traditional Heuristic-based model (green-dotted lines) with 92 total handovers, where ping-pong locations—marked by small squares—still exhibit clusters of tightly spaced handover events. In contrast, the Proposed Model (red-dotted lines) with 80 total handovers demonstrates a significant decrease in such clustered, bidirectional events. Manual inspection of both the real-life dataset and the Q-decision dataset after testing confirmed that 76% of ping-pong events were eliminated. To achieve this improvement, the reward function in our Q-learning model incorporated a penalty for consecutive handovers occurring within the defined Time-to-Trigger (TTT) window. This temporal filtering mechanism played a pivotal role in promoting more deliberate handover decisions, thereby improving connection stability and reducing signaling overhead.

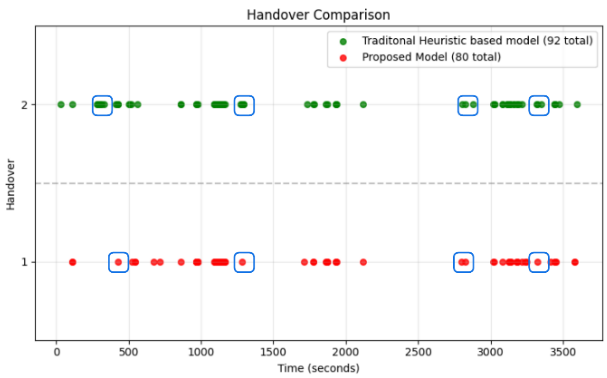

Figure 5: Ping-Pong Event Mitigation

D. *Latency Comparison with Heuristics*

Conventional heuristic-based handover algorithms used by telecom operators typically require 5 to 20 milliseconds to evaluate signal conditions and determine whether a handover should be triggered [18]. In contrast, our Q-learning-based model benefits from pre-trained knowledge and pre-abstracted state definitions. When deployed, the decision process involves a simple table lookup using the current state's parameters. This enables ultra-fast inference, with a decision latency of approximately 0.016 milliseconds, representing a significant reduction in handover decision time. This rapid responsiveness is especially valuable in high-mobility scenarios where timely HO decisions are critical to maintaining service continuity.

E. *Discussion*

The results demonstrate that the proposed reinforcement learning-based framework achieves notable handover optimization when applied to real-world LTE mobility data. The observed reduction in handovers from 92 to 80 represents a 13% improvement under realistic urban conditions, reinforcing the practical significance of data-driven optimization even when absolute numbers appear moderate.

A key strength of the system lies in its systematic evaluation of 50,000 parameter combinations encompassing HOM, TTT, and clustering weights, ultimately converging on a configuration that balances responsiveness with signal quality. This broad search, trained on a larger Day 1 dataset and validated on Day 2, supports the model's generalization capability across unseen traffic conditions. Subtractive clustering not only enabled such cross-day training and testing but also significantly reduced state space complexity, allowing RL to operate faster—a critical factor in machine learning–based solutions where decision latency directly impacts applicability.

Crucially, the substantial elimination of ping-pong handovers—achieved through temporal filtering via TTT-aware penalties in the reward function—underscores the effectiveness of learning-based control over heuristic rules. Unlike conventional methods that often rely on static thresholds, the RL agent adapts to diverse conditions while maintaining robust QoS. Moreover, Q-learning only initiates handovers when absolutely necessary, thus avoiding the excessive triggers common in heuristic-based algorithms.

Additionally, the inference speed of just 0.016 milliseconds, made possible by state abstraction and pre-trained Q-table lookups, far surpasses traditional operator algorithms, which typically operate within a 5-20 millisecond decision window. This efficiency is especially advantageous for high-mobility or low-latency 5G and beyond applications.

In summary, the experimental outcomes highlight the potential of reinforcement learning to drive intelligent mobility management in LTE networks using real-world data, reducing overhead, boosting stability, and enabling faster, smarter, and more adaptive handover decisions.

V. Conclusion

This work presented a Q-learning-based framework for handover optimization using real-world LTE drive test data from urban Bangladesh. The framework fine-tuned HOM, TTT, and clustering weights to reduce handovers by 13 percent and completely eliminate ping-pong effects, demonstrating greater adaptability than static threshold methods. While the dataset did not include direct QoS indicators such as throughput, strong connection quality was maintained through signal metrics including RSRP, RSRQ, and CINR. Future work will focus on incorporating throughput-based metrics, extending the approach to 5G, and evaluating its performance in a wider variety of environments.

REFERENCES

[1] Y. Ullah, M. Bin Roslee, S. M. Mitani, S. A. Khan, and M. H. Jusoh, “A Survey on Handover and Mobility Management in 5G HetNets: Current State, Challenges, and Future Directions,” Jun. 01, 2023, *MDPI*. doi: 10.3390/s23115081.

[2] S. K. Thillaigovindhan, M. Roslee, S. M. I. Mitani, A. F. Osman, and F. Z. Ali, “A Comprehensive Survey on Machine Learning Methods for Handover Optimization in 5G Networks,” Aug. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/electronics13163223.

[3] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, no. 1, pp. 237–285, Jan. 1996.

[4] D. Zidic, T. Mastelic, I. Nizetic Kosovic, M. Cagalj, and J. Lorincz, “Analyses of ping-pong handovers in real 4G telecommunication networks,” *Computer Networks*, vol. 227, May 2023, doi: 10.1016/j.comnet.2023.109699.

[5] W. K. Saad, I. Shayea, B. J. Hamza, H. Mohamad, Y. I. Daradkeh, and W. A. Jabbar, “Handover parameters optimisation techniques in 5G networks,” *Sensors*, vol. 21, no. 15, Aug. 2021, doi: 10.3390/s21155202.

[6] D. Boughaci, “Solving optimization problems in the fifth generation of cellular networks by using meta-heuristics approaches,” in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 56–62. doi: 10.1016/j.procs.2021.02.008.

[7] V. R. Gannapathy, R. Nordin, A. Abu-Samah, N. F. Abdullah, and M. Ismail, “An Adaptive TTT Handover (ATH) Mechanism for Dual Connectivity (5G mmWave—LTE Advanced) during Unpredictable Wireless Channel Behavior,” *Sensors*, vol. 23, no. 9, May 2023, doi: 10.3390/s23094357.

[8] R. Karmakar, G. Kaddoum, and S. Chattopadhyay, “Mobility Management in 5G and Beyond: A Novel Smart Handover with Adaptive Time-to-Trigger and Hysteresis Margin,” *IEEE Trans Mob Comput*, vol. 22, no. 10, pp. 5995–6010, Oct. 2023, doi: 10.1109/TMC.2022.3188212.

[9] G. Alsuhli *et al.*, “Mobility Load Management in Cellular Networks: A Deep Reinforcement Learning Approach,” *IEEE Trans Mob Comput*, vol. 22, no. 3, pp. 1581–1598, Mar. 2023, doi: 10.1109/TMC.2021.3107458.

[10] Q. Liu, C. F. Kwong, S. Wei, L. Li, and S. Zhang, “Intelligent Handover Triggering Mechanism in 5G Ultra-Dense Networks Via Clustering-Based Reinforcement Learning,” *Mobile Networks and Applications*, vol. 26, no. 1, pp. 27–39, Feb. 2021, doi: 10.1007/s11036-020-01718-w.

[11] Q. Liu, C. F. Kwong, S. Wei, S. Zhou, L. Li, and P. Kar, “Reinforcement learning-based joint self-optimisation method for the fuzzy logic handover algorithm in 5G HetNets,” *Neural Comput Appl*, vol. 35, no. 10, pp. 7297–7313, Apr. 2023, doi: 10.1007/s00521-021-06673-5.

[12] R. Wang, Y. Sun, C. Zhang, B. Yang, M. Imran, and L. Zhang, “A novel handover scheme for millimeter wave network: An approach of integrating reinforcement learning and optimization,” *Digital Communications and Networks*, vol. 10, no. 5, pp. 1493–1502, Oct. 2024, doi: 10.1016/j.dcan.2023.08.002.

[13] M. Benzaghta, S. Ammar, D. López-Pérez, B. Shihada, and G. Geraci, “Data-Driven Cellular Mobility Management via Bayesian Optimization and Reinforcement Learning,” May 2025, [Online]. Available: http://arxiv.org/abs/2505.21249

[14] M. Wunder, M. Littman, and M. Babes, "Classes of multiagent Q-learning dynamics with ε-greedy exploration," in *Proc. 27th Int. Conf. Mach. Learn. (ICML'10)*, Madison, WI, USA, 2010, pp. 1167–1174. doi: 10.5555/3104322.3104470.

[15] K. Bataineh, M. Naji, and M. Saqer, "A comparison study between various fuzzy clustering algorithms," *Jordan J. Mech. Ind. Eng.*, vol. 5, pp. 335–343, 2011.

[16] A. Abdelmohsen, M. Abdelwahab, M. Adel, M. Saeed Darweesh, and H. Mostafa, “LTE handover parameters optimization using Q-learning technique,” in *Midwest Symposium on Circuits and Systems*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 194–197. doi: 10.1109/MWSCAS.2018.8623826.

[17] M. N. I. Alonso and F. Arias, "The mathematics of Q-learning and the Hamilton-Jacobi-Bellman equation," SSRN Electron. J., Jan. 2025, doi: 10.2139/ssrn.5083336.

[18] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification,” 3GPP TS 36.331 V15.3.0, 2018. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/36_series/36.331/

[19] M. M. S. Shafi, K. M. Istiaque, S. S. Sowad, and M. T. Kawser, “Drive-Test-Based LTE Handover Dataset for Cellular Mobility Studies in Urban Bangladesh,” *Mendeley Data*, vol. V1, 2025, doi: 10.17632/n2pvmtyn2j.1.