AI/ML-Driven Predictive Mobility Management in 3GPP: Technical Insights and Standardization Trends

Wonho Lee[†], Heeju Chae[‡], Inkyu Bang[§] Eunkyung Kim[‡], Taehoon Kim[†], and Gosan Noh[¶]
[†]Department of Computer Engineering, Hanbat National University, Daejeon 34158, South Korea
[‡]Department of Artificial Intelligence Software, Hanbat National University, Sejong 30139, South Korea
[§]Department of Intelligence Media Engineering, Hanbat National University, Daejeon, 34158, South Korea
[¶]Department of Electronic Engineering, Hanbat National University, Daejeon, 34158, South Korea
Email: {wonholee, 30242860}@edu.hanbat.ac.kr, {ikbang, ekim, thkim, gsnoh}@hanbat.ac.kr

***Abstract*—This paper provides an overview of recent standardization efforts in 3GPP to integrate AI/ML-based predictive mobility management into 5G-Advanced networks. It highlights key motivations for transitioning from reactive handover mechanisms to predictive models, summarizes the progression from Release 17 through Release 19, and outlines technical challenges and future directions for intelligent mobility solutions.**

***Index Terms*—Mobility management, AI/ML, 3GPP, RAN2, handover prediction, standardization survey**

I. Introduction

In the context of 5G-Advanced and beyond, mobility management remains a fundamental component for ensuring seamless connectivity and maintaining service continuity during user mobility. Conventional mobility management in 3GPP networks, as defined in TS 38.331 [4], is predominantly event-driven—relying on triggers such as Event A2 (serving cell quality falls below threshold) and Event A3 (neighbor cell becomes better than serving cell). Although widely deployed, these reactive mechanisms often struggle to cope with the rapid signal variations encountered in high-mobility and millimeter-wave (FR2) scenarios, leading to suboptimal handovers, increased control-plane overhead, and degraded quality of experience (QoE).

Recognizing these limitations, 3GPP has gradually incorporated AI/ML into its standardization roadmap to enable predictive mobility management. Release 17 initiated this transition by defining a systematic data collection framework in TR 37.817 [5] to support ML model training. Release 18 introduced the concept of AI/ML model lifecycle management (LCM) in TR 38.843 [6], enabling model deployment, versioning, and fallback strategies across the RAN. Building on this, Release 19 (TR 38.744 [7]) marked a major milestone by specifying a unified framework for predictive mobility, including direct and indirect prediction models, measurement prediction strategies, evaluation metrics, and enhancements to RRC signaling through the introduction of the `predictionReport` Information Element (IE).

State-of-the-art prediction models based on deep learning architectures, such as Long Short-Term Memory (LSTM) networks and Transformers, have shown promising improvements in simulation studies, particularly for anticipatory handover and link-quality estimation. Nonetheless, real-world deployment challenges persist, such as ensuring model generalization across diverse topologies, minimizing the computational load on user equipment (UE), and ensuring model interpretability and reliability in operational networks [12]. Future enhancements are expected to include tight integration with reconfigurable intelligent surfaces (RIS), sub-terahertz (sub-THz) bands, and enriched mobility context information, which can support more accurate and timely predictions of user trajectories and channel conditions.

In this paper, we present a comprehensive overview of the standardization progress and technical directions of AI/ML-driven predictive mobility management in 3GPP. We first analyze the limitations of conventional event-triggered mobility mechanisms and highlight the motivations behind the transition toward predictive models. We then review the developments from Release 17 to Release 20, focusing on data collection frameworks, AI/ML model lifecycle management, measurement prediction schemes, and recent advancements in handover event forecasting. In addition, we examine the practical deployment challenges associated with model generalization, computational constraints, and explainability. Finally, we discuss ongoing and future standardization efforts, including the emergence of AI-native mobility architectures, and identify key research challenges that must be addressed to enable scalable and interoperable predictive mobility solutions.

II. RRM Measurement Prediction

3GPP Release 19 introduces a standardized framework for AI/ML-based Radio Resource Management (RRM) measurement prediction, building on the foundational studies conducted in Release 17 (TR 37.817 [5]) and model lifecycle principles introduced in Release 18 (TR 38.843 [6]). The objective of this framework is to enhance the efficiency of mobility-related measurements by enabling the network to forecast signal quality indicators—such as Reference Signal Received Power (RSRP)—across beams, frequencies, and cells, thereby reducing measurement overhead and improving the timeliness and accuracy of mobility decisions.

This approach reflects a shift from a reactive, measurement-triggered handover mechanism toward a predictive paradigm where AI models proactively estimate future radio conditions based on historical patterns. The resulting predictions can be used to support smarter measurement gap configuration, early beam switching, and inter-frequency handover preparation.

The measurement prediction framework introduced in Release 19(TR 38.744 [7]) defines key architectural components including prediction input granularity (e.g., sliding/non-sliding windows), data source integration, model update cycles, and signaling enhancements for conveying predicted metrics. In parallel, Release 20 is expected to provide further enhancements by expanding the scope toward multi-RAT (Radio Access Technology) prediction and AI-native control loop integration. For example, prediction mechanisms are being studied to anticipate the degradation of LTE radio conditions, enabling earlier measurement triggering on NR cells, which improves handover reliability across RAT boundaries.

Additionally, Release 20 explores enhanced RRM prediction capabilities that tightly integrate with the AI-native mobility architecture (RP-240004 [9]), where predicted measurements are not just input to mobility logic but actively guide scheduling and radio resource allocation. This enables the gNB to dynamically reconfigure measurement configurations, such as measurement gaps and reporting frequency, based on anticipated signal transitions.

Together, these developments support the long-term goal of reducing UE-side measurement burden, increasing RAN intelligence, and enabling fully predictive mobility behaviors that scale across frequency layers, RATs, and vendor implementations.

A. Prediction Approaches

1) Temporal Prediction: Temporal prediction focuses on estimating future signal quality (e.g., Layer 3 Reference Signal Received Power, or L3-RSRP) based on historical measurements within the same frequency band. Two representative prediction strategies are studied: non-sliding (static) and sliding (dynamic) window methods.

The static (non-sliding) method uses only past actual measurements without reusing previous predictions. This limits error accumulation but may underperform in situations where continuous estimation is necessary. In contrast, the sliding method simulates real-world measurement skipping by using both past predictions and actual measurements as input for ongoing prediction. This approach reflects practical deployments where reducing reporting frequency can lower signaling load but increases the risk of compounding prediction errors.

Fig.1 visualizes the core difference between the two approaches. In the non-sliding case, the prediction model is fed only with a fixed-size sequence of past measurements and is updated periodically without overlap. In contrast, the sliding method employs overlapping windows where predictions from previous windows can feed into subsequent ones, better capturing temporal dynamics at the cost of potential error propagation.

These methods are formally studied in 3GPP Release 19 under the AI/ML for Mobility study item (TR 38.744 [7]) as part of the RRM measurement prediction use case. The sliding window strategy is particularly relevant to high-mobility scenarios such as vehicles or trains, where the signal quality changes rapidly and requires continual forecasting. In such cases, maintaining a low reporting burden while ensuring prediction robustness is critical.

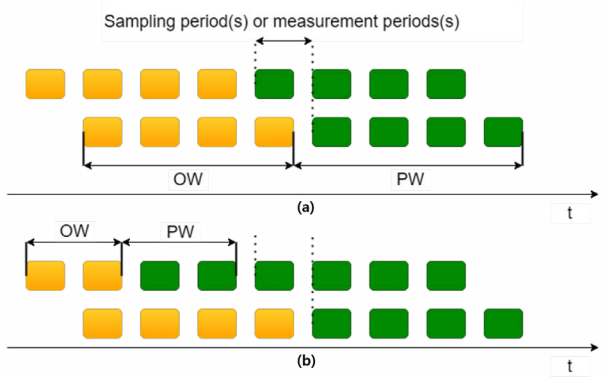

Fig. 1: Temporal prediction methods: (a) non-sliding prediction using disjoint windows, (b) sliding prediction with overlapping windows and recursive input.

To improve accuracy under sliding scenarios, filtering mechanisms have been investigated. These include reusing Layer 1 (L1) filtered data or combining previous predictions with new observations. The effectiveness of these filtering methods depends on user mobility speed, prediction window size, and variability of radio conditions. 3GPP also discusses the potential for adaptive window lengths, where the Observation Window (OW) and Prediction Window (PW) are dynamically adjusted based on environment, speed, or service type, to balance latency and reliability.

2) Spatial Prediction for Beam Reduction: In beam-based wireless systems, UEs often need to measure signal strength across multiple spatial beams, which creates a heavy measurement burden. Spatial prediction techniques aim to reduce this overhead by enabling the network to infer the signal quality of unmeasured beams based on a subset of observed beams.

According to 3GPP TR 38.744 [7], both cell-level and beam-level measurement prediction are within the scope of RRM measurement prediction. In particular, spatial-domain beam prediction, referred to as Beam Measurement Case 1 (BM-Case 1), allows the network to estimate the signal quality of non-measured beams using a subset of directly measured beams. This approach is supported by AI/ML models trained to learn spatial correlations among beams, often based on physical proximity, angular separation, or historical signal patterns.

By selectively measuring only a fraction of the available beams and predicting the rest, the UE can significantly reduce energy consumption and signaling overhead. However, the effectiveness of this strategy depends on the beam selection policy—such as deterministic selection or learned skipping—and the ability of the AI model to generalize under partial observations. 3GPP further discusses mapping relationships between wide beams and narrow beams, and how to exploit such structure to improve prediction accuracy.

Fig. 2 illustrates this concept. A subset of beams is directly measured by the UE (shown in yellow), while the remaining beams (shown in green) are inferred by the network using

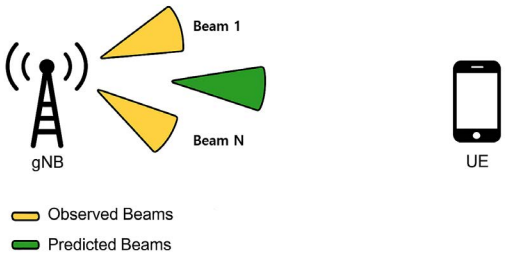

Fig. 2: Spatial beam prediction in a gNB-UE system.

spatial prediction. This mechanism supports overhead-efficient beam management and enables scalable mobility enhancements for multi-beam systems.

3) Generalization Across Frequencies and Cells: A significant challenge in AI/ML model deployment is ensuring that trained models can generalize across different frequency layers or cell configurations. For inter-frequency prediction, model accuracy varies depending on the degree of correlation between frequency bands (e.g., predicting 2 GHz performance based on 4 GHz data). Therefore, correlation analysis between bands is critical before reusing models in new spectrum environments.

Similarly, inter-cell generalization—using a model trained on one cell in another—requires that the spatial propagation characteristics of the cells are sufficiently similar. Techniques such as cluster-based training, where models are trained using data from multiple neighboring cells, have shown more consistent performance. These methods help mitigate the limitations of deploying single-cell trained models in complex and diverse network topologies.

B. Standardization Considerations

To support these AI-based prediction mechanisms, 3GPP has proposed protocol extensions that ensure interoperability and integration with existing RRC frameworks. For instance, the `predictionReport` Information Element (IE), introduced in Release 19, is designed to convey predicted measurement values and related metrics to the base station. This facilitates proactive handover and measurement skipping operations while maintaining control plane integrity [4].

Overall, the RRM measurement prediction framework introduced in Release 19 [7] represents a shift toward intelligent, context-aware mobility management. It offers a modular architecture that accommodates both temporal and spatial prediction needs, while also emphasizing the importance of cross-domain generalization for scalable deployment.

III. Measurement Event Predictions

A. AI/ML for Predictive Mobility Management

Conventional mobility management in cellular networks relies heavily on reactive measurement event reporting mechanisms. In this traditional approach, user equipment (UE)

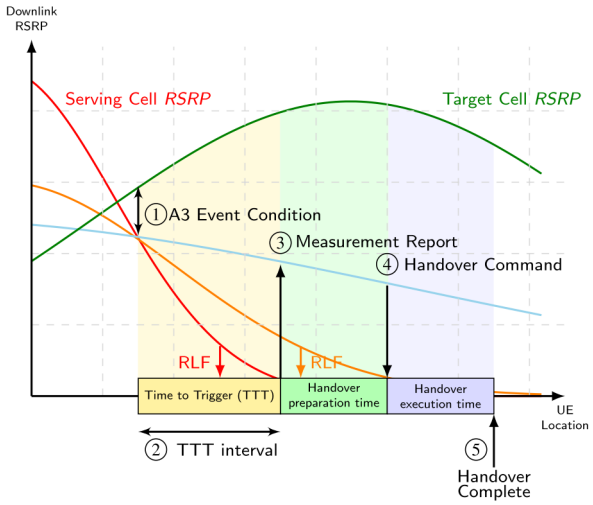

Fig. 3: RSRP-based Event A3 triggering and handover sequence.

monitors radio conditions and transmits measurement reports to the base station when specific thresholds are met. According to 3GPP TS 38.331 [4], Event A3 is triggered when a neighboring cell's signal becomes better than that of the serving cell, while Event A2 is initiated when the serving cell's signal quality drops below a certain threshold. While these mechanisms are standardized and widely implemented, they are inherently reactive. The latency can be particularly detrimental in high-mobility scenarios such as vehicular environments or in high-frequency FR2 deployments, where radio conditions change rapidly and signal degradation can occur before a handover can be effectively initiated. Fig. 3 visualizes this procedure by showing the sequential timing of key handover events—starting from the A3 event trigger, through the Time-to-Trigger (TTT), to the measurement report and handover execution steps.

To address these shortcomings, AI/ML-based predictive mobility management is emerging as a transformative solution. The core idea is to shift from reactive to proactive mobility decisions by forecasting the future state of radio conditions and handover needs. This enables the network to initiate handover preparation and execution before adverse conditions are encountered, thereby minimizing handover failures and ensuring seamless connectivity.

Multiple AI/ML modeling approaches are under consideration. One popular approach treats the problem as a time-series forecasting task, where models such as LSTM networks or Transformer-based models predict future values of signal quality indicators like RSRP and RSRQ [7]. These predictions can be used to preemptively trigger handover actions.

Alternatively, classification models can be trained to predict whether a handover will be required in the near future based on input features such as historical signal metrics, UE location, speed, mobility pattern, and surrounding network topology. Regression models may also estimate the exact time at which

an event (e.g., A2 or A3) is likely to occur, providing fine-grained control over handover timing [12].

Federated Learning (FL) has been identified as a key enabling technology for training predictive mobility models in distributed environments. In FL, model training is performed locally at UEs or edge nodes using their own data, and only the model updates are shared with the central server [8]. This approach preserves data privacy, reduces backhaul communication overhead, and enables collaborative training across diverse network segments. This is particularly important for real-world deployments, where centralized training may be impractical or undesirable.

These predictive methods offer several benefits over traditional mechanisms. They enable proactive handover preparation, reduce the risk of radio link failure, and improve quality of service for end users. Moreover, by predicting mobility trajectories and dwell times in cells, they can help avoid unnecessary handovers and reduce signaling overhead, contributing to more efficient radio resource utilization.

Recent developments in 3GPP Release 20 further expand the role of AI/ML in predictive mobility by exploring its applicability to radio link failure (RLF) forecasting and adaptive Conditional Handover (CHO) management [9]. RLF prediction aims to proactively detect and mitigate impending link breakdowns using real-time degradation patterns and historical handover performance. Meanwhile, CHO procedures are being enhanced by allowing AI-based models to dynamically configure trigger conditions—such as offset thresholds and Time-to-Trigger (TTT)—based on predicted signal quality trends. These directions reflect a broader transition toward AI-native mobility architectures, where model-driven inference loops are embedded directly into control-plane logic to support autonomous, context-aware handover strategies.

B. Standardization Activities and Technical Frameworks

3GPP has formally acknowledged the applicability of AI/ML in mobility management through a series of study and work items. Release 17 marked the beginning of standardized efforts to support AI/ML in mobile networks, focusing on establishing robust data collection frameworks. In Technical Report 37.817 [5], 3GPP outlined mechanisms to collect relevant radio and context data at both the UE and network side, which are essential for training and validating machine learning models. This includes logging of measurement events, mobility-related KPIs, and metadata like location and UE capability information.

Building on this foundation, Release 18 introduced the AI/ML Model Lifecycle Management (LCM) framework, documented in TR 38.843 [6]. This framework defines the operational processes required to support AI/ML models within the RAN, including model training, deployment, validation, version control, and fallback strategies. It also considers interfaces and signaling mechanisms needed for model inference and feedback within the gNB, and ensures coordination between RAN layers and AI agents.

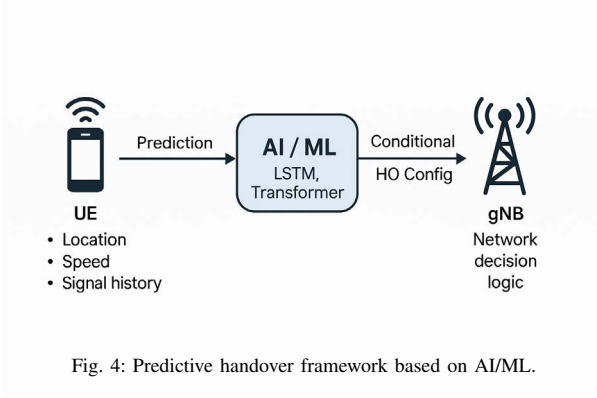

Fig. 4: Predictive handover framework based on AI/ML.

Release 19 represents a significant step forward with TR 38.744 [7], which introduces concrete predictive mobility management schemes under RAN2. Two primary approaches are standardized:

- **Indirect Prediction:** Models predict future signal quality metrics, such as RSRP and RSRQ. These predicted values are compared against conventional thresholds (e.g., A2, A3) to determine if handover preparation should begin.
- **Direct Prediction:** AI/ML models estimate the probability or timing of future measurement events directly, bypassing the need for intermediate metric predictions. This allows for faster and potentially more accurate decision-making.

These predictive models are intended to be integrated seamlessly with the existing RRC architecture. To facilitate this, the concept of a new RRC signaling message, `predictionReport`, has been proposed for standardization in TS 38.331 [4]. This IE enables the UE to report predicted future events or measurements, which the gNB can use to trigger conditional or proactive handover procedures. This aligns with ongoing enhancements to CHO functionality and is designed to support proactive mobility strategies.

Figure 4 illustrates the overall predictive handover framework enabled by AI/ML. In this architecture, the UE provides context-aware inputs such as signal history, location, and speed to a trained prediction model. The model forecasts mobility events in advance, allowing the gNB to prepare conditional handover configurations ahead of time.

Overall, these standardization activities aim to build a modular, extensible framework that accommodates a wide variety of predictive mobility models while ensuring interoperability across vendors and deployment scenarios. The long-term vision includes the integration of additional enabling technologies, such as RIS and sub-THz bands, to further enhance the accuracy and responsiveness of AI-driven mobility solutions [7].

In this context, Release 20 marks a significant evolution by launching the “AI-Native Mobility Architecture” study item (RP-240004 [9]), which extends predictive mobility from a functional enhancement to a foundational control framework within the RAN. This work item envisions embedding model-

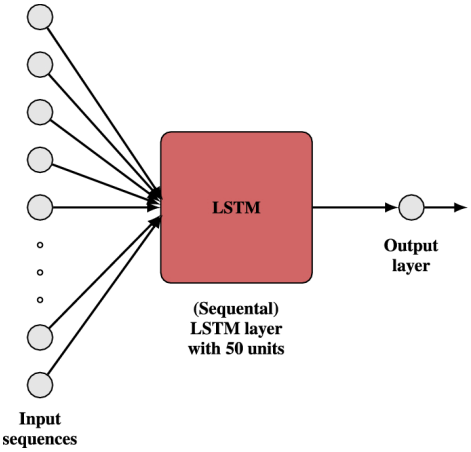

Fig. 5: LSTM-based model for predicting mobility events using time-sequential features such as signal quality, location history, and speed.

driven inference loops directly into RRC procedures and handover decision logic, allowing for real-time, closed-loop control of mobility state transitions. Furthermore, it explores cross-layer coordination across RRC, NGAP, and OAM signaling interfaces to support dynamic AI capability negotiation, model selection, and policy-driven mobility orchestration. These efforts signal a shift from modular AI enhancements to system-wide AI-native integration in future 5G-Advanced and 6G architectures.

C. Technical Considerations

The predictive mobility architecture requires not only accurate modeling but also careful consideration of practical deployment constraints in real-world mobile networks. While AI/ML algorithms can demonstrate high accuracy in controlled environments, their integration into commercial cellular systems must account for computational feasibility, energy efficiency, and real-time responsiveness.

One major challenge arises from the limited computational resources and power budgets of UE. Complex neural network models, such as LSTMs and Transformers, can be computationally intensive and may not be suitable for real-time inference on battery-powered devices without significant optimization. Techniques such as model pruning, quantization, and edge-optimized neural networks are therefore being considered to reduce latency and energy consumption during on-device inference [6].

Figure 5 illustrates the architecture of an LSTM-based model designed to predict mobility events. The model takes time-sequential input features—such as signal quality, location history, and user speed—and processes them through a sequential LSTM layer with 50 units. By learning temporal patterns in the input sequences, the model produces an output that can be used to anticipate future mobility events (e.g., A3 trigger), enabling proactive handover preparation.

In predictive mobility, LSTMs have been successfully applied to forecast mobility event timing and to model signal behavior under measurement skipping scenarios [7]. Compared to shallow models, they offer greater robustness in high-mobility and mmWave environments, where signal variation is both rapid and unpredictable.

Beyond the hardware limitations, ensuring the generalization of AI/ML models across diverse deployment scenarios is another critical concern. Networks are deployed in heterogeneous environments—urban, rural, indoor, high-speed vehicular, and more. A model trained in one scenario may fail to perform accurately in another due to domain shift. This underscores the need for extensive cross-scenario validation and possibly the use of transfer learning or meta-learning approaches to enhance model adaptability [6].

To mitigate the risks associated with model drift and changing network conditions, the AI/ML Model Lifecycle Management (LCM) framework introduced in TR 38.843 [6] recommends robust validation pipelines, periodic retraining procedures, and fallback mechanisms. These mechanisms ensure that outdated or poorly performing models can be quickly identified and replaced or bypassed, thus maintaining the reliability and safety of mobility decisions made by the AI system.

Despite the simulation-based performance improvements reported in TR 38.744 [7], the primary objective of standardization remains to develop a generic, interoperable framework that is vendor-neutral and adaptable to real-world networks. The intent is not only to achieve performance gains but also to create trustable and explainable models that mobile operators can deploy and manage effectively.

In conclusion, the success of predictive mobility management hinges not only on algorithmic accuracy but also on practical deployability, cross-environment generalization, lifecycle robustness, and compatibility with evolving network architectures. These considerations remain central to the ongoing work in 3GPP and are likely to shape future releases as AI/ML becomes an integral part of mobile network intelligence.

IV. Open Challenges and Future Work

To provide a structured perspective on future directions related to Release 20, this section categorizes the challenges and opportunities into three domains: technical, standardization, and broader implementation considerations.

A. Technical Challenges and Enhancements

Model Generalization and Computation: AI/ML models for predictive handover are often trained on deployment-specific data, which limits generalization across diverse environments. Lightweight and energy-efficient models (e.g., quantized or pruned LSTMs) are necessary to support real-time decision-making under tight latency constraints on user equipment (UE) with limited compute capabilities. Release 20 emphasizes the importance of supporting model negotiation and inference offloading strategies [9].

Real-Time Inference: Mobility decisions in high-mobility scenarios require inference within tens of milliseconds. Release 20 addresses this by enabling inference capability signaling during RRC connection setup and supporting offline fallback models when real-time prediction is infeasible.

Emerging Technologies: Reconfigurable Intelligent Surfaces (RIS) and sub-terahertz (sub-THz) communications are expected to enhance spatial and temporal prediction fidelity. These technologies enable more controllable radio environments and richer contextual inputs for AI models [9].

B. Standardization Aspects

Lifecycle Management and Model Interoperability: Release 20 introduces model-ID-based signaling and lifecycle management (LCM) to improve interoperability in multi-vendor networks. However, challenges remain regarding model update synchronization, fallback handling, and cross-layer alignment.

Cross-Layer Signaling: The AI-Native Mobility Architecture study in Release 20 proposes embedding AI-driven decision-making into RRC, Next Generation Application Protocol (NGAP), and Operations, Administration, and Maintenance (OAM) signaling. Harmonizing these layers to support distributed inference and lifecycle coordination is critical for scalable deployments.

Explainability and Trust: 3GPP SA5 and SA3 are defining standards for AI trust, robustness, and traceability [10]. Release 20 incorporates the need for explainable AI systems that provide confidence levels, trigger explanations, and audit mechanisms to support reliable operations.

C. Other Implementation Considerations

Data Availability and Training: Collecting high-quality mobility traces and signal measurements remains difficult due to regulatory and operational constraints. OAM-based data collection and federated learning frameworks are being explored to address these limitations [11].

Policy-Aware Prediction: Future systems must support policy-driven, service-differentiated mobility—e.g., Ultra-Reliable Low-Latency Communication (URLLC) flows receive higher mobility prediction priority than enhanced Mobile Broadband (eMBB) or massive Machine-Type Communication (mMTC). This introduces a new layer of complexity that spans AI, session management, and policy control.

Deployment-Ready Architectures: Beyond model accuracy, future architectures must be robust to adversarial conditions, auditable, and seamlessly integrated with existing infrastructure. Predictive mobility must co-exist with conventional fallback procedures and support real-world traffic variability.

In summary, Release 20 lays the groundwork for AI-native mobility through enhanced inference architecture, signaling support, and integration of emerging technologies. Future work must continue to address cross-domain interoperability, lightweight modeling, and trustworthy AI frameworks to realize predictive mobility at scale.

V. Conclusion

This paper has provided a comprehensive survey of AI/ML-driven predictive mobility management within the 3GPP standardization framework, with a particular emphasis on the technical developments from Release 17 through Release 20. We reviewed the transition from reactive to predictive mobility, discussed the design and application of measurement and event prediction models, and examined challenges in model deployment, generalization, and real-time execution. We also explored the emerging AI-native mobility architecture introduced in Release 20, highlighting the incorporation of closed-loop inference, multi-layer signaling support, and integration of RIS and sub-THz technologies.

As wireless networks evolve toward 6G, predictive mobility will be a cornerstone for achieving seamless connectivity and service continuity under diverse mobility conditions. Continued research is needed to enhance model robustness, optimize signaling overhead, and ensure multi-vendor interoperability. The findings and frameworks summarized in this paper serve as a foundation for future innovations in intelligent mobility management systems.

Acknowledgment

This research was supported by the ICT Talent Cultivation Project (ICT Innovation Human Resources 4.0 Program) and the University ICT Research Center (ITRC) Program funded by the Ministry of Science and ICT (MSIT), Republic of Korea. This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) under Grant No. IITP-2025-RS-2022-00156212 (50%) and IITP-2025-RS-2024-00437886 (50%).

References

[1] Z. Ali, M. Miozzo, L. Giupponi, et al., "Recurrent Neural Networks for Handover Management in Next-Generation Self-Organized Networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Dublin, Ireland, Jun. 2020.

[2] M. A. Habib, P. E. Iturria-Rivera, Y. Ozcan, et al., "Transformer-Based Wireless Traffic Prediction and Network Optimization in O-RAN," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Denver, CO, USA, Jun. 2024.

[3] D. Wang, A. Qiu, S. Partani, Q. Zhou, H. D. Schotten, "Mitigating Unnecessary Handovers in Ultra-Dense Networks through Machine Learning-based Mobility Prediction," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC2023-Spring)*, Florence, Italy, Jun. 2023, pp. 1–7, doi: 10.1109/VTC2023-Spring57618.2023.10200542.

[4] 3GPP TS 38.331 V17.0.0, "NR; Radio Resource Control (RRC); Protocol specification," Dec. 2022.

[5] 3GPP TR 37.817 V17.0.0, "Study on data collection for NR," Sept. 2022.

[6] 3GPP TR 38.843 V18.1.0, "Study on AI/ML for NR," Dec. 2023.

[7] 3GPP TR 38.744 V19.0.0, "Study on AI/ML for mobility in NR," Mar. 2024.

[8] 3GPP TDoc R2-2005127, "Federated Learning for Predictive Mobility," 3GPP RAN2 Meeting #112, 2020.

[9] 3GPP RP-240004, "Study on AI-Native Mobility Architecture," 3GPP RAN Plenary, Release 20, 2024.

[10] 3GPP TR 33.898 V0.4.0, "Study on Security Aspects of AI/ML in 5G," 3GPP SA3, Dec. 2023.

[11] 3GPP TS 28.105 V17.3.0, "Management and orchestration; Artificial Intelligence (AI)/Machine Learning (ML) management framework," 3GPP SA5, Dec. 2022.

[12] M. Akrout, A. Feriani, F. Bellili, A. Mezghani, and E. Hossain, "Domain Generalization in Machine Learning Models for Wireless Communications: Concepts, State-of-the-Art, and Open Issues," *IEEE Wireless Communications*, early access, Mar. 2023.