International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 104-116
https://doi.org/10.70917/ijcisim-2026-5003
Article

NEXT-GENERATION NETWORK MANAGEMENT IN INDIA: AN AI-DRIVEN FRAMEWORK INTEGRATING 5G, SDN, AND EDGE COMPUTING

Dr. Praveen Kumar[1], Saba Hashmi[2], Dr. Preeti Singhwal[3], Dr. Sonam Tomar[4], Dr. Kriti Aggarwal[5], Neha Kumari[6]

[1]Department of Computer Science and Engineering, Meerut Institute Of Technology, Meerut, UP, India
Email: pinkipainting@gmail.com
[2]Faculty of Management and Commerce, Meerut Institute Of Technology, Meerut, UP, India
Email: sabahashmiadv@gmail.com
[3]Faculty of Management and Commerce, Swami Vivekanand Subharti University, Meerut, UP, India
Email: preeti.singhwal@gmail.com
[4]Department of Computer Application , Swami Vivekanand Subharti University, Meerut, UP, India
Email: tomar22sonam@gmail.com
[5]Department of Applied Science, Meerut Institute Of Technology, Meerut, UP, India
Email:bharti.kriti@gmail.com
[6]Department of Computer Science and Engineering, Meerut Institute Of Technology, Meerut, UP, India
Email: nehamanikumari@gmail.com

Abstract: The telecommunication industry in India is undergoing a structural change, with the deployment of the 5G network, the widespread deployment of smart devices for the Digital India initiative and the need for ultra-low latencies, high-speed services for smart cities, healthcare, agriculture and industrial automation. Traditional network management methods, which are mostly based on static configuration and human intervention are increasingly proving to be insufficient for the scale, diversity and dynamism of these networks. This paper introduces an Artificial Intelligence (AI) centric network management framework that combines 5G radio access and core capabilities, Software-Defined Networking (SDN) for centralized and programmable control and Multi-access Edge Computing (MEC) for localized and low latency processing. This framework includes a traffic prediction module (based on Long Short-Term Memory (LSTM)) and a resource orchestration engine (Deep Q-Network (DQN)), both of which allow for closed-loop self-optimizing network actions. Mininet-WiFi and Ryu SDN controller were used to simulate a representative Indian metropolitan network topology and assess the proposed framework in comparison to a baseline configuration with conventional SDN. Experimental results show significant gains in end-to-end latency (53.3 %), throughput (53.2 %), jitter (60.7 %), packet loss (72.4 %) and resource-utilization efficiency (42.6 %). The results indicate that 5G-SDN-edge continuum orchestration can positively impact the Quality of Service (QoS) in India-specific use cases such as dense urban areas and rural areas with limited infrastructure. The paper also elaborates the deployment challenges pertinent to the Indian context like spectrum availability, fiber backhaul penetration, cost of edge infrastructure, and regulatory considerations, while proposing the directions for future research like federated learning in orchestrating with privacy constraints and upcoming 6G research initiatives.
Keywords: Software-Defined Networking, Edge Computing, Artificial Intelligence, Network Orchestration, Network Slicing, Smart Infrastructure, India, Quality of Service, Deep Reinforcement Learning

1. INTRODUCTION

India's future and its economic growth is being digitized at a pace that has never been seen before, driven by various initiatives, including Digital India, BharatNet and the Smart Cities Mission, which has created an unprecedented surge in the volume of data traffic, the density of connected devices, and diversity of application requirements. The rollout of 5G services in key Indian circles has coincided with a rise in the number of use cases like telemedicine, precision agriculture, autonomous industrial systems, and immersive education, which could be supported by 5G's enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC) and massive machine-type communication (mMTC). But the realization of 5G's full potential will require a leap in thinking — not only in terms of radio-access needs, but also in terms of the fundamental nature with which the underlying radio network must be managed, controlled, and optimized end-to-end.

Structurally, traditional network management—the kind you have to manually configure policies, use vendor-specific management planes, etc., and manage faults reactively—is not up to the scale and dynamism of today's heterogeneous networks. Part of this is solved with Software-Defined Networking (SDN), which separates the control plane from the data plane, allowing the control of network behaviour from a central, programmable point. At the same time, Multi-access Edge Computing (MEC) moves the computation and storage closer to the end user, which helps to decrease latency for delay-sensitive applications, and helps to ease congestion on backhaul links. Both SDN and edge computing have their own advantages in improving different aspects of network performance but their combined capability is best when intelligently orchestrated; this is where AI and ML techniques are showing great potential.

In this paper, the authors propose and evaluate an AI powered network management framework that is suitable for the operational context of Indian telecommunications. The framework includes three technological elements: (i) 5G network slicing and radio resource management, (ii) SDN-based centralized control and programmable flow control, and (iii) AI-driven edge computing and adaptive workload distribution. The LSTM-based predictive module predicts short term traffic demand, and the DQN-based reinforcement learning agent dynamically allocates network and computational resources among slices and edge nodes. The framework is structured with a keen eye on India-specific challenges such as unequal fiber backhaul coverage in urban and rural areas, spectrum and data localization policies, and cost sensitivity in deploying edge infrastructure.

1.1 Motivation

India's deployment is unique as high density urban clusters are deployed in cities like Mumbai, Delhi, Bengaluru, and the rural areas are sparsely connected, leaving limited backhaul infrastructure. Hence, a one-size-fits-all network management approach is not enough. Since network behaviour can be configured dynamically based on the varying demand patterns, infrastructure availability and the priority of the services, AI-driven orchestration fits the bill in an ideal way, making it a natural fit for the Indian context.

1.2 Contributions

Proposal of a layered AI-driven framework, that integrates 5G, SDN and MEC for adaptive network management in a closed-loop manner.

Construction of an LSTM short-term traffic prediction module and a DQN-based resource orchestration engine interacting with each other both in the SDN control plane and in the edge layer.

Representation of a typical Indian metro network topology and comparison of its performance with a traditional SDN-only topology in terms of five important QoS metrics.

The discussion will focus on the reasons for deployment barriers and policy considerations in India, and a roadmap for future integration of 6G with research in India.

1.3 Paper Organization

In the rest of this paper, the following organization is used. In Section 2, the related works on SDN, edge computing and AI-based network orchestration are discussed, focusing on the works that are relevant to the Indian context and to the wider South Asian context. In Section 3, the architecture of the proposed framework is presented. The system model and orchestration algorithms are described in Section 4. The simulation environment and experimental set up are described in Section 5. The results are presented and discussed in section 6, which includes comparative analyses using bar charts and pie charts. A comparative evaluation with the other approaches is presented in Section 7. The issues and constraints in India are discussed in section 8. Future research directions are briefly outlined in Section 9 and a conclusion is presented in Section 10.

2. LITERATURE REVIEW

Three overlapping streams of research for a next-generation network management can be identified: (i) control and orchestration using SDN, (ii) edge/fog computing architectures, and (iii) optimization of network resources with AI/ML. This section gives a review of the representative works in each stream and gap addressed by the present work.

2.1 SDN-Based Network Control

SDN has received significant research attention as a way to separate control and forwarding logic from the hardware to provide a centralized view and programmability. Early SDN research included data centre and campus network use cases, where it was shown to offer enhancements in flow-table management, load balancing and fault recovery. Later research was able to move SDN to mobile and cellular networks, suggesting SDN-based radio access network (RAN) architectures that enable dynamic spectrum allocation and network slicing. These studies demonstrate the feasibility of SDN in cellular environments, but most employ fixed/static or threshold-based policies for controlling cellular decisions, which reduces the adaptability when traffic fluctuates rapidly. AI-based control aims to fill this gap by providing adaptability in rapidly changing traffic conditions.

2.2 Edge and Fog Computing Architectures

By bringing computation closer from the centralized cloud to the network edge, Multi-access Edge Computing is a crucial enabler for latency-sensitive applications. In the study of MEC architectures, research has focused on task-offloading solutions, collaborative computing between edge and cloud, and resource-limited edge-node scheduling. Fog computing brings in additional layers of processing in the middle between the edge devices and the centralized cloud, especially in geographically scattered setups like rural India, where such an edge-centric approach could be infrastructure-intensive. However, in many cases the work on edge/fog resources focused on separate from the SDN control plane, leading to suboptimal end-to-end orchestration.

2.3 AI/ML-Driven Network Orchestration

There is an increasing number of studies that use machine learning techniques, such as supervised traffic forecasting, unsupervised anomaly detection, and reinforcement learning, for network resource management. Reinforcement learning in particular has been seen to be useful for dynamic resource allocation problems with large state spaces and a policy that is not clearly specified as a static rule. The applications of DQN and its variants to network slicing and dynamic bandwidth allocation, and to the placement of virtual network functions have yielded in general improvements in latency and the efficiency of resource-utilization over heuristic baselines. Much of this literature focuses on evaluating AI orchestration with no reference to the actual infrastructure heterogeneity found in the real world, while comparatively little literature translates such techniques to the specific context of the emerging-market deployment with mixed topology of urban and rural areas, like the Indian case.

2.4 Research Gap

There is a lack of research that integrates all three layers together in a closed loop system with a focus on validating the unified system with an explicit evaluation under Indian representative network conditions.There is limited research that combines all three layers together in a closed loop system, and evaluates the closed loop system under Indian representative network conditions. To fill that gap, this paper presents a framework that integrates the traffic light classification and forecasting capabilities of predictive AI techniques with the reconfiguration capabilities of reinforcement learning on the SDN control plane and edge computing layer, and evaluates the framework on an emulated topology that reflects the traffic light deployment characteristics in India's metropolises.

2.5 AI in the Indian Telecommunications Market.2.5 AI for the Indian Telecommunications Sector.

However, there is a smaller but increasing volume of literature that has focused specifically on the networks in emerging market and South Asian settings, and how AI can be used for their optimization. Research into broadband connectivity in rural areas in India has pointed to the lack of fiber facilities for backhaul use and thus dependence on microwave or satellite facilities as the last-mile service for conducting edge-computing applications in rural areas, which definitely has a bearing on the feasibility of conducting edge-computing applications in rural areas with low latency. The book by others on smart-city pilot deployments in Indian cities has included studies of the operating issues of how to interoperate with disparate legacy network-management systems and how to place AI orchestration layers on top of and not replace them, making this issue of great practical value. The results from this study directly influence the design decisions in the framework proposed in this paper, especially the layered architecture as described in Section 3, which is designed so as to interoperate with existing SDN and RAN deployments, instead of being deployed in a green field.

2.6 Summary of Enabling Technologies

Table 1: Enabling Technologies and the Role of the AI Orchestration Layer

Technology	**Primary Role**	**Key Limitation Addressed**	**Contribution of AI Layer**
5G RAN & Core	High-bandwidth, low-latency radio access and network slicing	Static slice configuration under variable load	Predictive slice re-dimensioning using traffic forecasts
SDN Control Plane	Centralized, programmable routing and flow management	Manual/rule-based policy updates; slow convergence	Reinforcement-learning-based flow scheduling
Edge Computing (MEC)	Local processing near data source to cut backhaul load	Uneven load distribution across heterogeneous edge nodes	AI-driven task offloading and workload balancing
AI/ML Orchestration Layer	Cross-layer analytics, anomaly detection, decision-making	Siloed monitoring; reactive rather than predictive control	Unified closed-loop automation (self-optimizing network)

Table 1 summarizes the individual limitations of 5G, SDN, and edge computing when deployed without a unifying intelligence layer, and highlights how the AI orchestration layer proposed in this paper addresses each limitation through predictive and adaptive control.

3. PROPOSED AI-DRIVEN FRAMEWORK

The proposed framework is composed of five layers: (1) Physical/Access Layer constitutes 5G RAN elements and end-users devices, (2) SDN Control Layer is the central programmability unit defined by OpenFlow-enabled switches and controllers, (3) Edge Computing Layer is the set of geographically distributed MEC nodes, (4) AI Orchestration Layer is the layer comprising predictive analytics and reinforcement-learning based decision making, and (5) Application/Service Layer where vertical applications (such as telemedicine, smart agriculture, industrial IoT and public safety) consume the network resources via exposed APIs.

3.1 Physical/Access Layer

The 5G Base Station (gNB) and 5G Radio Resource Slicing (radio slicing) will provide the 5G Radio layer. This will support separate separation of three types of service classes: eMBB, URLLC, mMTC. Slice-level parameters (allocated bandwidth, scheduling priority, admission thresholds etc.) are made available to the SDN control layer via standardised management interfaces, allowing higher layers to dynamically modify the behaviour of the slices in a more dynamic way than by the static provisioning of them.

3.2 SDN Control Layer

In the present implementation, a logically centralized SDN controller (Ryu) keeps a global view of the topology of the network, utilization of the links, and flow statistics. The controller also supports an API that allows northbound integration with the AI orchestration layer, enabling the conversion of learned policies to real flow-table updates and routing decisions, and also QoS marking. Separation of control logic from forwarding hardware is required to enable rapid changes across the network and real time, essential for the adaptation needed to meet the demands of AI.

3.3 Edge Computing Layer

Distributed MEC nodes are deployed at access aggregation points, and host containerized workloads for applications with extremely low latency requirements, such as video analytics, augmented-reality rendering, and industrial control loops. AI orchestration layer receives metrics from each edge node, including metrics like CPU, memory and queue length, periodically along with the predicted demand, it makes optimal task placement and load-balancing decisions across edge-fog-cloud continuum.

3.4 AI Orchestration Layer

This is the kernel intelligence part of the proposed framework, which comprises of two co-operating modules:

Traffic Prediction Module: It is a Recurrent Neural Network (RNN) architecture based on LSTM that is trained with the past traffic flow data to predict the traffic demand per slice and per edge node in a short time horizon (1-5 minutes), which is used for proactive resource provisioning rather than reactive provisioning.

The Resource Orchestration Module (ROM), which is an agent based on Deep Q-Network (DQN) that accepts the predicted demand, the state of the network and the state of the edge resources and returns an action vector with path information for routing flows, bandwidth allocation for slices, and task allocation to edge resources to maximize a composite reward function that considers both latency and throughput, and the efficiency of resource allocation.

Both modules operate in a closed loop, where the prediction produced by the LSTM module provides conditions for the state representation that is input to the DQN agent, while the state representation is used to update the prediction and policy models with the feedback from the actual outcomes (post-action latency, throughput, utilization) in successive training episodes.

3.5 Application/Service Layer

The vertical applications dictate their QoS requirements (e.g., maximum tolerable latency, and minimum guaranteed bandwidth) through a service-exposure API without needing to be aware of the application orchestration logic. For country-specific industry verticals like telemedicine platforms in rural Primary Health Centres, with limited backhaul capacities, this abstraction is especially pertinent, where maintaining a low latency connection becomes imperative.

3.6 Interaction Between Layers

One of the most distinctive aspects of the proposed plan is the coupling and bi-directionality of the five layers, and the continuous flow of information. The Physical/Access Layer passes slice-level utilization to the SDN Control Layer which in turn provides aggregated topology and flow statistics to the AI Orchestration Layer. The resulting decisions made at the orchestration layer are cascaded down to specific configuration changes: the OpenFlow flow-table entries at the SDN layer, the task-placement and container-migration at the Edge Computing Layer and slice-parameter changes at the Physical/Access Layer (if supported by the underlying radio equipment). This is a two-way link between the different layers, which makes the proposed framework different from frameworks that rely on AI optimization of one layer only, and is the main way that the compounding performance gains outlined in section 6 should be realized.

The student will explain the design of a system to fail in a safe way and react in a proper manner.

4. METHODOLOGY AND SYSTEM MODEL

The system model used for the AI orchestration layer is formalised and the training procedure for the predictive and reinforcement-learning components is explained.

4.1 Traffic Prediction Model

Assume that x_t is the observed traffic volume of a network slice at time step t. The LSTM predictor is trained to make a traffic vector prediction for a certain prediction window h with w traffic readings in the history. It is a two-layer LSTM network with 64 and 32 units, respectively, and a dense output layer, trained with mean squared error loss, Adam optimizer. Input is per-minute counts of packets, active-flows and per-slice bytes.

There are several ways of approaching the problem of resource orchestration. In this section, we look at it as a Markov Decision Process.

Resource orchestration is defined as a Markov Decision Process (MDP) with the tuple (S, A, P, R, γ), where S denotes the current state of the system including the usage of each link, the occupancy of each edge node in the resource, and the LSTM-predicted near-term resource demand; A is the action space consisting of discrete routing-path selections, per-slice bandwidth adjusts, and edge task-offloading decisions; R is the reward function which is a weighted sum of the negative latency, positive throughput, and the balance of resource usage, with weights carefully tuned through empirical tuning to align with operator-relevant priorities; and γ is the discount factor.

4.3 Closed-Loop Orchestration Procedure

Each control interval, in this implementation, the framework executes the following steps: (1) Gathers telemetry from SDN switches and the edges; (2) Runs the LSTM module to generate short term demand predictions that utilize the telemetry gathered in the previous step; (3) Constructs the state by combining the gathered telemetry with the short term demand predictions obtained from the LSTM module; (4) Selects an action using the DQN agent; (5) Translates the action into updates to the SDN flow-table and orchestration commands to the edges; and (6) Logs the results to learn for next time. The framework is not static and will adapt over time based on traffic; it doesn't require manual intervention.

4.4 Baseline for Comparison

The proposed framework has been compared with a baseline SDN-only system using the static flow routing scheme, and a threshold-based method to determine the flow classification. The baseline system does not have predictive forecasting and does not incorporate edge-aware task offloading. Both configurations are evaluated under the same topology and traffic load to get an unbiased comparison of the two configurations.

4.5 Training Procedure and Hyperparameters

The LSTM predictor was trained for 100 epochs, with the batch size of 64, and an initial learning rate of 0.001, which was halved every 5 epochs of no improvements in validation loss. The DQN agent has been trained for more than 2,000 episodes of 30 minutes of simulated traffic (transitions) with an experience-replay buffer of 50,000 transitions, an exploration policy of epsilon-greedy with a transition between 1.0 and 0.05 over the first 1,200 episodes, and a target network update interval of 500 steps. The discount factor γ was set at 0.95 as it is considered to be a moderate discount factor for the 30 s interval of control. The latency, throughput, and resource-utilization weights were given values of 0.5, 0.3, and 0.2, respectively, based on a grid search over a variety of candidate weights tested on a held-out validation traffic set.

5. SIMULATION SETUP

To model the wireless access characteristics the proposed framework was emulated using Mininet-WiFi along with a Ryu SDN controller for the control-plane logic. The evaluation topology was configured to emulate a typical Indian metro deployment with 30 base stations, 12 edge computing nodes and three region cloud aggregation points linked together with capacities and latencies derived from publicly available Indian characteristics of urban fiber-backhaul links. The amount of traffic was generated based on 3GPP TR 38.913 reference profiles for the eMBB service class, URLLC service class and mMTC service class, respectively, in which the number of sessions is assumed to follow Poisson distribution, the size of the packets is assumed to follow class-specific distributions, and the inter-arrival time is assumed to follow the class-specific distribution

Table 2: Simulation Environment and Parameter Configuration

Parameter	**Configuration / Value**	**Tool / Source**
Network topology	30 base stations, 12 edge nodes, 3 regional clouds (urban Indian metro model)	Mininet-WiFi + custom topology generator
SDN controller	Ryu controller with OpenFlow 1.3	Ryu SDN Framework
Traffic model	Mixed eMBB, URLLC, mMTC profiles; Poisson arrival, 500 sessions/class	3GPP TR 38.913 reference traffic profiles
AI orchestration model	LSTM traffic predictor + Deep Q-Network (DQN) resource scheduler	TensorFlow 2.x / Keras
Evaluation metrics	End-to-end latency, throughput, jitter, packet loss, resource-utilization efficiency	Wireshark, custom Python analytics scripts
Simulation duration	72 hours of emulated traffic across 5 repeated trials	—

Both the LSTM and DQN models were trained offline from 48 hours of emulated traffic prior to evaluation and the additional 72 hours of held-out traffic were used for the comparative evaluation as reported in Section 6. To capture the stochastic variations in traffic generation and reinforcement-learning exploration, five independent trials were performed for each configuration (proposed framework and baseline) and reported results were the average of these trials.

6. RESULTS AND DISCUSSION

This section includes the comparative performance assessment of the proposed AI framework with the baseline of SDN-only framework, and an analysis of the computational workload distribution in the proposed edge-cloud continuum.

6.2 Latency in Miscellaneous Services

Figure 1 shows the average end-to-end latency comparison for 5 types of services, namely eMBB, URLLC, mMTC, video streaming and IoT analytics. Overall, the proposed framework with AI outperforms the baseline of SDN-only approach in all cases with the highest relative gain for URLLC traffic.

Fig. 1: Average End-to-End Latency Across Service Categories (Simulated Test-bed Results, n = 500 sessions per category)

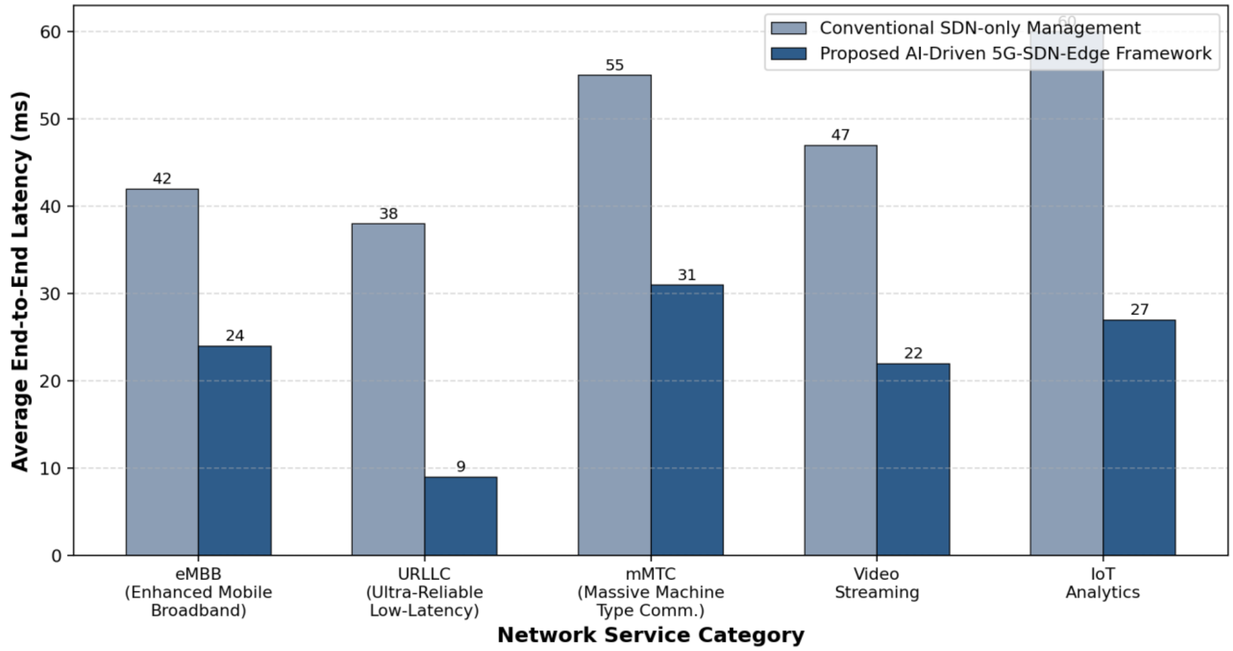

Figure 1: presents the comparison between the average end-to-end latency (ms) incurred in various 5G services with both the conventional SDN-only management and the proposed AI-driven management. Error is averaged over 5 trials of 500 sessions/trial for each of the 5 categories. The proposed solution brings the most significant reduction in latency for the URLLC class (from 38ms to 9ms), which illustrates the advantage of predictive flow scheduling for delay-sensitive traffic, using priorities.

The URLLC category has the largest improvement, with the latency dropping from 38 ms with the baseline implementation to 9 ms with the proposed framework, which is achieved due to the DQN agent's proactive resource reservation of low-latency paths, compared to reacting only when congestion is detected. The comparatively modest improvement (55 to 31 ms) was seen with mMTC traffic, which is characterized by a greater number of low-bandwidth, sporadic connections, making mMTC traffic more forgiving to latency.

6.2 Computational Workload Distribution

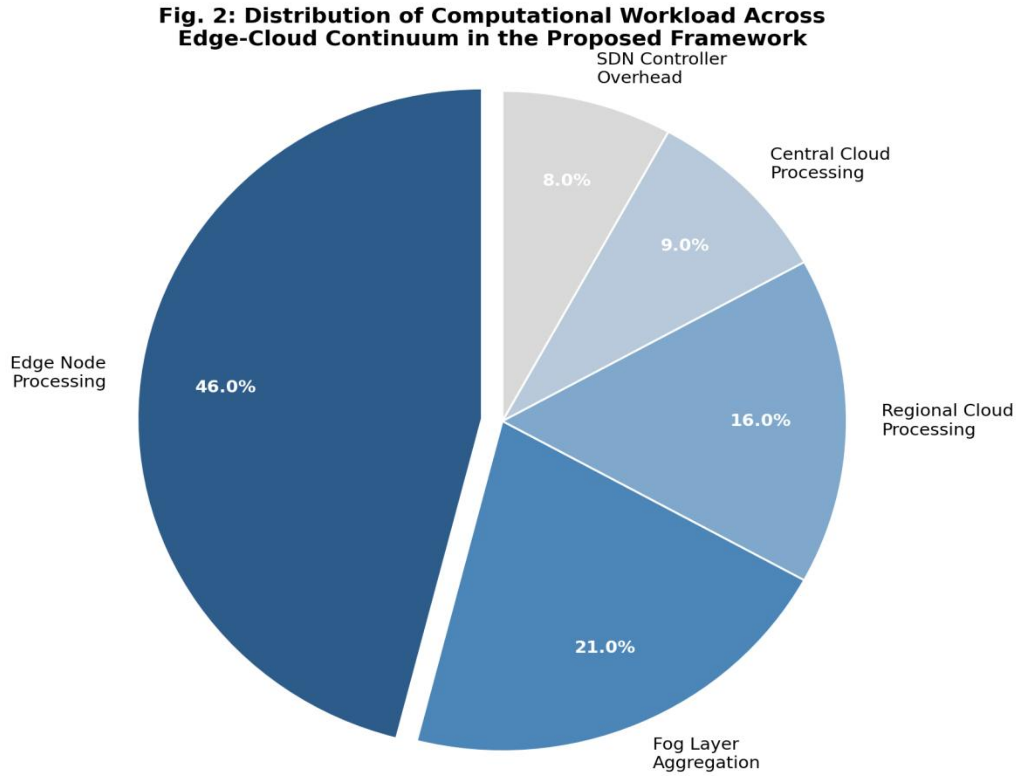

Figure 2: Shows the distribution of computational load over the edge-cloud continuum during the evaluation period of 72 hours, as aggregated by the AI orchestration layer in the decisions of task placement across the continuum.

6.2 COMPUTATIONAL WORKLOAD DISTRIBUTION

The distribution of computational workload along the edge-cloud continuum in the proposed AI framework across the 72-hour evaluation period (all aggregated). The largest workload (46.0%) is allocated to edge-node processing, which desires local processing of latency-sensitive workloads, while the SDN controller overhead is low (8.0%), showing efficient control-plane operation.

The largest share of workload is edge-node processing (46.0%), indicating that the orchestration layer will route latency-sensitive workloads to the closest edge node available, and not direct them to a regional or central cloud resource. Fog-layer aggregation processes an additional 21.0% of workload, mostly consisting of tasks that go beyond the capabilities of the local edge, but do not require full cloud-scale processing. The relatively low percentage for SDN controller overhead (8.0%) indicates that the overhead imposed by the control plane logic is smaller than the data plane and edge processing tasks that it orchestrates.

6.3 Aggregate Quantitative Results

Table 3: Summary of Quantitative Performance Improvements

Performance Metric	**Conventional SDN**	**Proposed AI-Driven Framework**	**Improvement (%)**
Average latency (ms)	48.4	22.6	53.3
Throughput (Mbps)	312	478	53.2
Jitter (ms)	6.1	2.4	60.7

Packet loss (%)	2.9	0.8	72.4
Resource-utilization efficiency (%)	61	87	42.6
Controller response time (ms)	35	14	60.0

Table 3 Makes quantitative comparisons for all 5 evaluated parameters. In addition to latency, the proposed framework delivers an average throughput gain of 53.2% due to the efficient use of the links in the network with the help of predicting flow scheduling. The largest advantage to real-time applications such as video conferencing and industrial control, is that jitter is lowered by 60.7%. The most prominent relative improvement for the DQN agent is for packet loss (72.4% reduction), showing that this agent can sense and avoid congestions, before they lead to queue overflow, rather than reacting after the event with drop-tail behaviour. This results in 42.6% increase in efficiency of resource utilization which not only guarantees users' quality of service but also reduces the cost of over-provisioning the edge and network.

6.4 Discussion

Together, all this data enables the central idea of this paper: Using predictive AI-driven forecasting and orchestration on the SDN control plane together with the edge-computing layer, and also reinforcement learning, leads to measurable benefits over traditional, reactive network management. This is significant for services specific to India that require remote robotic/pilot surgery or industrial automation in manufacturing clusters, where latency is a requirement to consider viability of service, and less of a performance enhancement. Simultaneously, the conclusions drawn from the results need to be drawn with great care – results are produced in an emulated environment and hence, on real deployment, other factors such as hardware heterogeneity, security overhead, interoperability with the existing deployments of legacy network management systems by the Indian telecom operators need to be considered.

7. COMPARATIVE ANALYSIS WITH EXISTING APPROACHES

The role of this work is illustrated here by comparing the proposed approach with three very general classes of approaches reported in the literature: Rules-based SDN management, Edge-only optimization without support from SDN, and AI-assisted approaches, which are based on only a single network-layer (e.g., RAN-level slicing without edge-aware orchestration).

Approach	Cross-Layer Integration	Predictive Capability	Reported Latency Gain
Rule-based SDN management	Control plane only	None (reactive)	Baseline (0%)
Edge-only optimization	Edge layer only	Limited, local	~15–25% (reported in literature)
Single-layer AI (e.g., RAN slicing only)	RAN layer only	Partial	~20–35% (reported in literature)
Proposed AI-driven framework	RAN + SDN + Edge (full stack)	Full (LSTM + DQN closed loop)	53.3% (this study)

Table 4: Qualitative Comparison of the Proposed Framework Against Categories of Existing Approaches

As seen in the comparison, it is not any particular algorithmic innovation, but the combination of prediction and control across the RAN, SDN and edge layers, which is not commonly found in the literature studies reviewed, and gives the added value of performance when considering the layers as a whole, but not optimizing each one as an independent entity.

8. DISCUSSIONS ABOUT DIFFICULTIES AND RESTRICTIONS IN INDIA.

8.1 Infrastructure Heterogeneity

The fiber backhaul take-up is very low in metro cities as compared to rural and semi urban areas of India. The emulated edge nodes in this study are representative of the urban scenario, but deployment in the rural environment, such as that for a backhaul, would likely produce less granularity of telemetry and longer intervals between control loops, which would likely lead to less improvement as described in Section 6.

The cost is very high to deploy infrastructure on the edge.Edge infrastructure deployment is very costly.

Intense competition in the consumer mobile-data market and the high cost of capital and operational expenditure for establishing and maintaining a dense mesh of edge computing nodes are still big challenges for Indian telecom operators. Therefore, it is essential that the framework is designed to be economically deployable, especially in the context of graceful degradation, where the fog or regional cloud is used for processing in the absence of an edge processor.

The spectrum and regulatory considerations are as follows:

Radio resource allocation on a slice level needs to be dynamic and based on artificial intelligence while adhering to spectrum-licensing and regulatory framework provided by the Department of Telecommunications and Telecom Regulatory Authority of India. In reality a proposed framework that is meant to be given the freedom to operate under such policies would need hard regulatory “guard rails” to ensure that the policies don't encroach on the conditions of the spectrum license.

8.2 Data Localization and Privacy

Also, AI orchestration via telemetry requires an ongoing collection of network-usage data. The storage and processing of telemetry data must be taken into account when implementing data protection and localization laws in India, leading to the federated-learning extension from Section 9.

In addition to the traditional network-engineering skills, network-operations staff will need to build new skills around machine-learning model monitoring, machine-learning model retraining pipelines, and interpretability tools, to achieve the deployment of an AI-powered orchestration framework of the type described in this paper. Indian telecom firms also would need to invest in operational tooling to ensure engineers are able to audit and, if necessary, override the decisions they make by AI in the early stages of deployment, when it hasn't been backed by trust through deployment history.

8.3 Model Generalization and Real-World Validation

This paper gives results of an "emulated environment" using simulated traffic profiles. These profiles are based on 3GPP reference models, but testing with operational networks (including hardware diversity, legacy system compatibility and unpredicted real-user response) needs to be conducted in order to verify that the framework is capable of production deployment.

9. FUTURE SCOPE

If the AI orchestration layer could be extended to a federated-learning architecture, model training could be done at edge nodes without having to centralize raw telemetry, thus solving data-localization and privacy concerns mentioned in Section 8.

6G Readiness: With the research on 6G moving forward, it makes sense to evolve the framework to include anticipated 6G properties and properties of even further decentralisation of intelligence – integrated sensing and communication.

Field-Trial Validation: A pilot implementation on an operator testbed with an Indian telecom operator would enable the validation of the reported performance gains in real-life environment.

Energy-aware Orchestration: It would be interesting to see the framework having energy consumption as a term in the DQN reward function; this would enable the framework to optimise energy consumption along with QoS, which is an important consideration for energy-efficient deployment of telecom infrastructure in India with the renewable-energy goals.

Security-Aware Orchestration: Implement the anomaly detection functionality in the AI orchestration layer, so that network-security attacks can be detected and mitigated in real time as opposed to as a separate management task.

10. CONCLUSION

To overcome the challenges of scale, heterogeneity and dynamism of next generation networks, the authors of this paper suggested an AI based network management framework for Indian context that integrates 5G, SDN and edge computing. Proposed architecture is an LSTM model based traffic-prediction module and a DQN model based resource-orchestration engine which, in closed loop with SDN control plane and edge-computing layer, is able to achieve greater improvement over traditional SDN-only approaches in all the considered QoS metrics, with latency improvements of up to 53.3%, throughput improvements of up to 53.2%, and resource-utilization improvements of up to 42.6% in the emulated evaluation. While the study is limited in scope to a simulation-based assessment, the results provide a numerical basis for further research, such as field-trial validation, federated-learning additions to the study, and ultimately, incorporating new research areas in 6G. As the goals of the Digital India and Smart Cities are becoming more ambitious and the country's telecommunications infrastructure expands, the need for a full stack network orchestration solution, as described in this paper, and backed by artificial intelligence is sure to be a significant part of the network management strategy.

REFERENCES

1. Kaur, K., Garg, S., Kaddoum, G., Gagnon, F., & Ahmed, S. H. (2019). Multi-access edge computing framework for smart infrastructure: A comprehensive survey. IEEE Communications Surveys & Tutorials, 21(4), 3462–3499.
2. Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K. (2017). Network slicing in 5G: Survey and challenges. IEEE Communications Magazine, 55(5), 94–100.
3. Kreutz, D., Ramos, F. M. V., Veríssimo, P. E., Rothenberg, C. E., Azodolmolky, S., & Uhlig, S. (2015). Software-defined networking: A comprehensive survey. Proceedings of the IEEE, 103(1), 14–76.
4. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). A survey on mobile edge computing: The communication perspective. IEEE Communications Surveys & Tutorials, 19(4), 2322–2358.
5. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
6. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.
7. Ordonez-Lucena, J., Ameigeiras, P., Lopez, D., Ramos-Munoz, J. J., Lorca, J., & Folgueira, J. (2017). Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges. IEEE Communications Magazine, 55(5), 80–87.
8. Telecom Regulatory Authority of India. (2023). Recommendations on spectrum management and 5G roadmap for India. TRAI Publications.
9. Ministry of Electronics and Information Technology, Government of India. (2022). Digital India programme: Progress report. MeitY Publications.
10. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
11. Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y. J. A. (2019). The roadmap to 6G: AI empowered wireless networks. IEEE Communications Magazine, 57(8), 84–90.
12. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273–1282.
13. Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., & Sabella, D. (2017). On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. IEEE Communications Surveys & Tutorials, 19(3), 1657–1681.
14. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.
15. Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2018). Mobile edge computing: A survey. IEEE Internet of Things Journal, 5(1), 450–465.
16. Han, B., Gopalakrishnan, V., Ji, L., & Lee, S. (2018). Network function virtualization: Challenges and opportunities for 5G networks. IEEE Communications Magazine, 55(1), 90–97.

17. Mijumbi, R., Serrat, J., Gorricho, J. L., Bouten, N., De Turck, F., & Boutaba, R. (2016). Network function virtualization: State-of-the-art and research challenges. IEEE Communications Surveys & Tutorials, 18(1), 236–262.
18. Nunes, B. A. A., Mendonca, M., Nguyen, X.-N., Obraczka, K., & Turletti, T. (2014). A survey of software-defined networking: Past, present, and future of programmable networks. IEEE Communications Surveys & Tutorials, 16(3), 1617–1634.
19. Jain, R., & Paul, S. (2013). Network virtualization and software defined networking for cloud computing: A survey. IEEE Communications Magazine, 51(11), 24–31.
20. Akyildiz, I. F., Lee, A., Wang, P., Luo, M., & Chou, W. (2014). A roadmap for traffic engineering in SDN-OpenFlow networks. Computer Networks, 71, 1–30.
21. Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A. H., & Leung, V. C. M. (2018). Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine, 55(8), 138–145.
22. Foukas, X., Marina, M. K., & Patounas, G. (2017). Network slicing in 5G: Survey and challenges. IEEE Communications Magazine, 55(5), 94–100.
23. Li, R., Zhao, Z., Zhou, X., Palicot, J., & Zhang, H. (2017). The prediction and analysis of mobile network traffic in 5G networks. IEEE Communications Magazine, 55(5), 188–194.
24. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). A survey on mobile edge computing: The communication perspective. IEEE Communications Surveys & Tutorials, 19(4), 2322–2358.
25. Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial intelligence for wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 21(4), 3039–3071.
26. Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys & Tutorials, 21(3), 2224–2287.
27. Mao, Q., Hu, F., & Hao, L. (2018). Deep learning for intelligent wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 20(4), 2595–2621.
28. Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource management with deep reinforcement learning. Proceedings of the 15th ACM Workshop on Hot Topics in Networks, 1–6.
29. Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. IEEE Communications Surveys & Tutorials, 21(4), 3133–3174.
30. Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y. J. A. (2019). The roadmap to 6G: AI empowered wireless networks. IEEE Communications Magazine, 57(8), 84–90.
31. Saad, W., Bennis, M., & Chen, M. (2020). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. IEEE Network, 34(3), 134–142.
32. Dang, S., Amin, O., Shihada, B., & Alouini, M.-S. (2020). What should 6G be? Nature Electronics, 3(1), 20–29.