International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Comparative Study of Traditional QoS Mechanism with Recent AI Based QoS Mechanism

Pawan Agrawal*, Dr. Kismat Chhillar**, Prof. Saurabh Shrivastava***,
Dr. Deepak Tomar****

*(PhD Research Scholar, Department of Mathematical Sciences and Computer Applications, Bundelkhand University, Jhansi, Uttar Pradesh, India, Email: pawanagrawal@bujhansi.ac.in)

**(Assistant Professor, Department of Mathematical Sciences and Computer Applications, Bundelkhand University, Jhansi, Uttar Pradesh, India, Email: dr.kismat@bujhansi.ac.in)

***(Professor, Department of Mathematical Sciences and Computer Applications, Bundelkhand University, Jhansi, Uttar Pradesh, India, dr.saurabh@bujhansi.ac.in)

****(System Analyst, Computer Center, Bundelkhand University, Jhansi, Uttar Pradesh, India)
Email: dr.deepak@bujhansi.ac.in)

----------**********----------

Abstract:

This paper provides a thorough comparison of traditional Quality of Service (QoS) methods and new AI-based architectures. It looks closely at the main principles, operational models, and performance features of Integrated Services (IntServ) and Differentiated Services (DiffServ). The paper points out the trade-off between reliable guarantees and network scalability. It also introduces the change brought by artificial intelligence and discusses key techniques like Deep Reinforcement Learning (DRL) and predictive analytics, which allow for proactive, dynamic, and automated network management. The discussion highlights the important role of enabling technologies such as Software-Defined Networking (SDN) and 5G in implementing AI-driven QoS. A detailed operational and architectural comparison follows, along with a deep dive into the major challenges and a plan for a future where AI smartly organizes and improves traditional QoS policies. The analysis concludes that while traditional methods offer a solid base, AI is crucial for meeting the needs of dynamic, complex, and intent-driven networks, providing unmatched efficiency, reliability, and flexibility.

Keywords — **QoS, Quality of Service, Computer Networks, Networking**

----------**********----------

I. INTRODUCTION

The modern digital landscape is shaped by a wide variety of applications. These range from low-latency, real-time services like Voice over IP (VoIP), video conferencing, and online gaming to bandwidth-heavy streaming and essential industrial control systems. Each application has its own requirements for network performance, including bit rate, delay, jitter, and packet loss. The default "Best Effort" service model is commonly used on the public Internet and treats all traffic the same. However, it does not meet the diverse and often strict requirements because it offers no guarantees on delay, throughput, or packet loss. This situation has led to the creation of Quality of Service (QoS) mechanisms, designed to prioritize different applications, users, or data flows to ensure a specific level of performance. Traditional QoS mechanisms have been crucial for managing network performance for many years. However, they are mostly static and rule-based. They depend on manual setups and human oversight to monitor and manage network performance and security [1]. This reactive approach struggles to keep up with the complex, ever-changing nature of modern networks. Relying on manual oversight increases the risk of mistakes and limits the network's ability to quickly respond to changes in traffic and demands. This issue is especially noticeable in environments with rapidly shifting conditions, such as wireless connections in tactical networks or large-scale

Internet of Things (IoT) deployments. The main problem is that static, pre-defined rules cannot effectively manage networks that are in constant flux. This leads to performance decline and vulnerabilities that can reduce a business's ability to innovate and meet market needs.

The rise of artificial intelligence (AI), especially machine learning (ML) and deep reinforcement learning (DRL), presents a new approach to network management that overcomes the weaknesses of traditional QoS. AI-enhanced networks use advanced data analysis and machine learning algorithms to learn from network traffic and performance data [2]. This allows them to make smart, data-driven decisions. The shift involves moving from fixed rules to dynamic learning and from reactive management to proactive and automatic responses to network changes. AI-powered networks can predict traffic flows, detect problems, and respond to changes without human intervention, resulting in enhanced efficiency and reliability. This paper offers a detailed comparison of traditional and AI-based QoS mechanisms.

It includes a side-by-side analysis of both approaches, discussing their technical foundations, operational philosophies, performance metrics, and scalability. It points out the relationships between network complexity and the need for intelligent systems. The paper shows that the main issue with traditional QoS is the significant trade-off between guaranteed performance and network scalability. It argues that AI provides a way to move beyond this fundamental compromise by enabling a learning-based approach to network control. Lastly, the report suggests a future research agenda focusing on hybrid solutions, where AI intelligently manages and improves traditional QoS policies in real-time.

II. RELATED WORK

A. *Overview of Foundational Concepts*

Quality of Service (QoS) is defined as the ability to provide different priorities to different applications or data flows to guarantee a certain level of performance [3]. The primary goal of QoS is to ensure that mission-critical applications have the necessary resources and priority to achieve high performance, even when network capacity is limited. This is achieved by managing and controlling key performance metrics that directly impact the user experience [4]. Network performance is evaluated using several critical metrics including latency, jitter, packet loss, and throughput. Latency or delay refers to the time taken for a data packet to travel from its source to its destination, and excessive latency can severely affect real-time applications such as VoIP and online gaming by making them practically unusable. Jitter denotes the variation in packet arrival delays, where high jitter can distort audio or video streams, leading to choppy or incoherent playback.

Packet loss occurs when packets fail to reach their destination due to congestion or errors within the network, necessitating retransmission that further increases delay and may cause network collapse under severe conditions. Throughput or bandwidth represents the rate at which data is successfully transmitted through a communication channel, with Quality of Service (QoS) mechanisms ensuring that applications needing specific bit rates, like video streaming, receive the required dedicated bandwidth for optimal performance. Most networks operate under a default Best Effort model, characterized by the absence of QoS controls, where the network simply attempts to deliver data using the maximum available performance and bandwidth at any given time without guaranteeing delay, throughput, or packet loss constraints. This limitation of the Best Effort model underscores the need for advanced QoS architectures that provide predictable and efficient service delivery for critical applications.

B. *Integrated Services (IntServ) Model*

The Integrated Services (IntServ) model is a deterministic, per-flow QoS architecture designed to provide explicit, end-to-end performance guarantees. Its design philosophy is rooted in the idea of providing a predictable and quantifiable service level for a specific data stream [5]. The IntServ model operates on the principle of explicit resource reservation. Before an application begins transmitting data, it uses a signaling protocol, typically the Resource Reservation Protocol

(RSVP), to request and reserve network resources (e.g., bandwidth, delay) along the entire path from the source to the destination.

Every routing device in the end-to-end path must negotiate and support these RSVP reservations. The network performs a deterministic Admission Control (AC) check to verify if the requested resources are available. If all intermediary devices can reserve the required resources, the application is cleared to send traffic within its requested profile, and the network guarantees to maintain the specified level of service. The primary advantage of IntServ is its ability to provide quantifiable and predictable performance guarantees. This is ideal for small, highly controlled environments, such as a military or industrial network, where mission-critical applications require an absolute assurance of service quality that cannot be compromised.

However, the model's major weakness is its poor scalability, which is a direct consequence of its per-flow state design. To provide end-to-end guarantees, every network node must create, support, and maintain state information for every single flow traversing it, which includes the source and destination IP addresses and ports [6]. This creates significant overhead in terms of CPU processing and memory consumption, making it impractical for large-scale networks like the public Internet, which may have thousands or even millions of flows. The continuous signaling required by RSVP also adds to this overhead.

C. Differentiated Services (DiffServ) Model

The Differentiated Services (DiffServ) model was developed to address the scalability shortcomings of the IntServ model [7]. It is a class-based, soft-QoS model designed for high scalability across large networks. The DiffServ model eliminates per-flow state by classifying packets into a limited number of traffic classes and marking them with a Differentiated Services Code Point (DSCP) in the IP header. This marking, which can also use values like MPLS EXP or CoS, allows for traffic classification at the network edge [8]. Network nodes then apply a defined.Per-Hop Behavior (PHB) to each class, which dictates how the packet should be forwarded, queued, or shaped. The PHB is a standardized method for handling traffic at each hop in a DiffServ-enabled network and refers to the way scheduling, queueing, policing, and shaping are configured on a router.

The Differentiated Services (DiffServ) architecture defines several key Per-Hop Behaviors (PHBs) that determine how packets are treated across network nodes to achieve varying levels of Quality of Service. The Expedited Forwarding (EF) PHB offers a low-loss, low-latency, and low-jitter service by assigning the highest priority to a particular traffic class, effectively functioning as a virtual leased line ideal for delay-sensitive applications like telephony and Voice over IP (VoIP). The Assured Forwarding (AF) PHB provides four distinct forwarding classes, labeled AF1 through AF4, each incorporating three drop precedence levels—low, medium, and high—to enable granular control over traffic handling and ensure bandwidth assurance for different aggregated traffic types based on priority. In contrast, the Best Effort (BE) PHB represents the default traffic class that operates without any guarantee of QoS, handling general data transmissions that do not require preferential treatment [9]. Together, these PHBs enable DiffServ networks to balance efficiency, reliability, and prioritization according to application requirements and service-level agreements.

DiffServ's core strength is its high scalability and low overhead. Since it doesn't require complex signaling protocols or the maintenance of per-flow state, DiffServ is a much more efficient model for large deployments. It is also much easier to configure and maintain than IntServ, making it the dominant model for large enterprise and service provider networks [10]. The main drawback, however, is the lack of hard, end-to-end performance guarantees. DiffServ is based on statistical preferences per traffic class, and the QoS provided is a local, per-hop affair based on locally defined policies. While this approach is effective under normal conditions, it can still lead to service degradation during periods of high congestion because there is no explicit resource reservation.

C. The IntServ over DiffServ Hybrid Approach

The IntServ and DiffServ models are not mutually exclusive but can be combined to leverage their respective strengths. The IntServ over DiffServ hybrid model is a practical attempt to reconcile the fundamental trade-off between deterministic guarantees and scalability. The approach is to use the IntServ model at the network edge, where the number of connections is limited and per-flow state is manageable. This provides end-to-end guarantees for a limited number of applications. The core of the network, which handles a much larger volume of aggregated traffic, then uses DiffServ's scalable, class-based approach. This model seeks to provide the predictable performance of IntServ while leveraging the efficiency and scalability of DiffServ in the network core. The IntServ reservations at the edge are mapped to DiffServ traffic classes, which are then carried over the core network, thereby offering a more practical solution for large deployments without requiring per-flow state on every router.

The fundamental trade-off between deterministic performance guarantees and network scalability represents a core philosophical choice in network design. IntServ's mechanism of per-flow state and signaling directly provides the foundation for hard guarantees but inherently limits its scalability due to the computational and memory overhead on every router. Conversely, DiffServ's class-based aggregation and marking enable high scalability by abstracting away individual flows, but in doing so, it loses the granular control needed for hard guarantees, providing only statistical assurances. The "IntServ over DiffServ" hybrid is a direct, rule-based attempt to manage this trade-off, but it remains a static compromise. The system relies on a human administrator to pre-configure a set of rules (PHBs, RSVP signaling) that remain largely unchanged unless manually updated. This means that traditional QoS mechanisms are inherently reactive, only working to enforce the pre-configured rules once a traffic event occurs. They are not designed to learn or adapt to novel, unforeseen network conditions, which is the precise gap that AI aims to fill.

III. THE PARADIGM OF AI ENHANCED QoS MECHANISMS

A. The Shift to Intelligent Networking

AI-enhanced networks represent a significant leap forward in terms of automation and intelligence, moving beyond the static, rule-based operations of traditional networking. This new paradigm is characterized by a shift from a reactive to a proactive management philosophy, where networks are capable of learning from data, adapting to changes, and optimizing performance autonomously. The goal is to address the limitations of manual configurations in dynamic and complex networks by leveraging machine learning and data analytics to continuously monitor, analyze, and optimize network performance in real-time.

B. Core AI Techniques for QoS Optimization

Deep Reinforcement Learning (DRL) is an advanced artificial intelligence approach in which an agent autonomously learns an optimal policy through trial-and-error interactions with an environment, guided by a reward signal. Applied in the context of Quality of Service (QoS), DRL effectively addresses the challenges of dynamic routing and resource allocation in complex network environments. In this framework, the agent embedded within the network's control plane makes intelligent routing and resource management decisions based on real-time network states, such as topology, traffic load, and link conditions. The environment comprises the network infrastructure itself, while the reward function plays a pivotal role by incentivizing actions that enhance QoS metrics, including reduced latency, increased throughput, and minimized packet loss.

The learning mechanism involves the agent continuously observing the network's state, selecting routing actions such as determining the optimal next hop, and receiving feedback in the form of rewards that reflect the impact of its decisions. Over time, using advanced algorithms like Deep Deterministic Policy Gradient (DDPG), the DRL agent refines its policy to maximize cumulative rewards, thereby evolving an optimal and adaptive routing strategy that maintains superior QoS even under dynamic traffic and

topology variations. This data-driven model provides a scalable and intelligent solution for complex, non-linear problems in network management. Figure 1 shows deep reinforcement learning cycle for QoS enhanced Mechanisms.

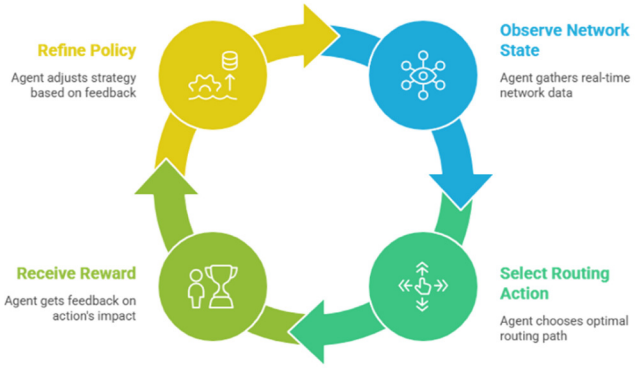

Figure 1. Deep Reinforcement Learning Cycle for QoS Enhanced Mechanisms

AI-enhanced networking empowers systems to operate proactively through predictive analytics and anomaly detection, shifting from traditional reactive approaches to intelligent foresight-driven management. By leveraging machine learning models such as Graph Convolutional Networks (GCNs) and Gated Recurrent Units (GRUs), vast historical and real-time network data can be analyzed to uncover patterns and forecast future conditions, predicting potential traffic congestions and performance bottlenecks before they impact service quality. This predictive capability marks a transformative shift from the manual and reactive nature of conventional QoS mechanisms. With proactive management, AI systems can anticipate service-level agreement (SLA) violations or network degradations and initiate corrective measures automatically such as reallocating bandwidth or rerouting traffic along less congested paths before disruptions occur, thereby enhancing both efficiency and reliability.

Expanding on this intelligence, hybrid AI models integrate the strengths of multiple learning strategies to achieve higher adaptability and precision. For instance, models combining GCNs and GRUs can predict evolving network topologies and performance metrics, while a Deep Reinforcement Learning (DRL) agent utilizes these predictions to dynamically optimize QoS parameters in real time. Such hybrid architectures are particularly effective in complex and dynamic environments like Mobile Ad Hoc Networks (MANETs), where topology and traffic conditions frequently change. Additionally, other hybrid frameworks incorporate deep learning methods such as Multilayer Perceptrons (MLP) and Deep Belief Networks (DBN) to model and predict Quality of Experience (QoE) by mapping QoS indicators to user satisfaction outcomes, thereby enabling comprehensive, intelligent, and adaptive network optimization.

C. Enabling Architectures: SDN and 5G

The integration of AI for Quality of Service (QoS) optimization is deeply intertwined with modern networking architectures such as Software-Defined Networking (SDN) and 5G, which collectively provide the structural and operational foundation for AI's capabilities. SDN serves as a crucial enabler by decoupling the control plane from the data plane, thereby granting a centralized controller the ability to manage network behavior programmatically. This architecture creates a direct interface for AI analytics engines to issue dynamic configuration commands such as rerouting traffic, reallocating bandwidth, and adjusting priorities which the SDN controller enforces across the network in real time. While SDN delivers the mechanism for adaptive control, AI injects the intelligence necessary to interpret network states and make optimal, context-aware decisions. At the same time, 5G and intent-based networking extend AI's role by supporting ultra-low latency communication and flexible network slicing. Figure 2 shows AI driven QoS optimization.

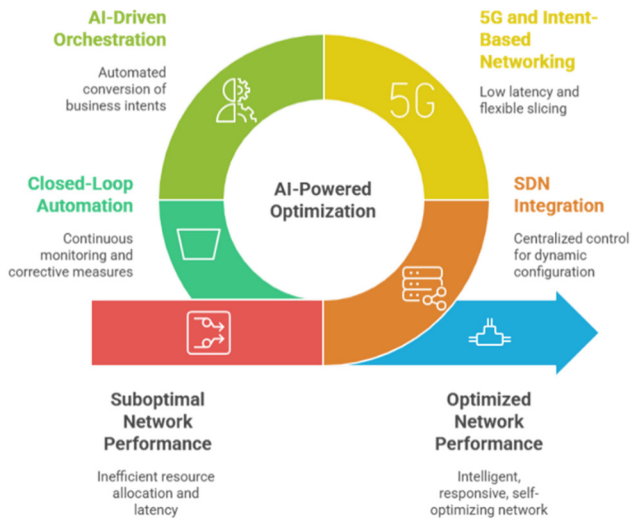

Figure 2. AI-Driven QoS Optimization

In this paradigm, high-level business intents, expressed in natural language (for instance, "ensure ultra-low latency video for remote surgery"), are automatically converted into granular, technical QoS configurations through AI-driven orchestration. AI agents manage these network slices to meet precise performance goals, using closed-loop automation systems to continuously monitor metrics such as latency, jitter, and packet loss. If deviations from performance targets are detected or predicted, the AI system autonomously triggers corrective measures to restore service levels. Together, SDN and 5G architectures amplify the potential of AI to deliver intelligent, responsive, and self-optimizing network environments that align with evolving user and application demands.

The fundamental difference between AI-enhanced and traditional QoS is the shift from rule-based to learning-based control. A traditional DiffServ policy will react to congestion by dropping lower-priority packets when a queue is full. An AI-enhanced system, however, might have already predicted the congestion using models like GCNs and GRUs, and proactively rerouted traffic to an uncongested path or dynamically adjusted queue sizes before any packets are dropped. This is not a reactive enforcement of a static rule but a dynamic, proactive optimization based on a learned policy. This learning process, guided by a reward function, allows the AI agent to explore a vast solution space to find optimal paths that a human administrator would be unable to configure statically.

The power of this approach is unlocked by the symbiotic relationship between AI and enabling architectures. Without a programmable control plane like SDN, an AI's optimal routing policy would be a theoretical conclusion, unable to be implemented in a distributed, device-by-device fashion. The synergy is profound: SDN provides the mechanism for dynamic control, and AI provides the intelligence to drive that control. 5G network slicing is the perfect application driver for this synergy, as it demands the real-time, dynamic orchestration that only an intelligent, automated system can provide.

IV. COMPREHENSIVE COMPARATIVE AND PERFORMANCE EVALUATION

A. Operational and Architectural Comparison

The operational and architectural differences between traditional and AI-enhanced QoS are stark, representing a fundamental philosophical divergence in network management. A comparative table can serve as a powerful tool to synthesize these distinctions. Table 1 shows the comparison of traditional and Enhanced QoS approaches. As shown in the table, traditional QoS mechanisms are built on a static, human-centric paradigm. The configuration process is manual and time-intensive, whether it involves the per-flow signaling of IntServ or the class-based marking of DiffServ. This results in low adaptability, as the network is limited by the administrator's pre-defined rules and cannot respond autonomously to unforeseen network changes or traffic spikes. The performance optimization philosophy is inherently reactive; it acts on events that have already occurred, such as a full queue triggering a drop precedence policy.

TABLE I
TRADITIONAL Vs. ENHANCED QoS

Aspect	IntServ	DiffServ	AI-Enhanced QoS
Configuration	Manual, per-flow signaling	Manual, class-based	Automated, dynamic, intent-based
Adaptability	static resource reservation	Low; static, rule-based PHBs	High; real-time learning &

			adaptation
Control Mechanism	Centralized, per-flow signaling	Distributed, per-hop forwarding	Centralized, learning-based control plane
Performance Opt.	Reactive, explicit reservation	Reactive, class-based rules	Proactive, predictive optimization
Core Philosophy	Hard, deterministic guarantees	Soft, statistical preferences	Dynamic, predictive optimization

In contrast, AI-enhanced QoS operates on an automated, dynamic, and intent-based model. AI systems continuously learn from network data, enabling high adaptability and real-time responses to changing conditions. The control mechanism shifts from a distributed, per-hop model or a centralized, per-flow signaling model to a centralized, learning-based control plane, often facilitated by an SDN controller. The core philosophy is not about enforcing a static rule but about a proactive, predictive optimization of the network to achieve a high-level intent, preventing issues before they arise.

B. Performance and Scalability Comparison

The comparison of IntServ, DiffServ, and AI-enhanced networking reveals key distinctions across performance guarantees, scalability, and resource utilization, highlighting AI's transformative potential in modern QoS architectures. IntServ provides deterministic, hard performance guarantees for individual traffic flows by explicitly reserving network resources along the entire transmission path, ensuring predictable service quality but at the cost of high complexity. Figure 3 shows comparison of recent QoS approaches.

Figure 3. Comparison of Recent QoS Approaches

DiffServ, on the other hand, offers statistical or soft guarantees based on Per-Hop Behaviors (PHBs), providing a scalable yet less precise mechanism where performance can still degrade under heavy congestion. AI-enhanced QoS systems bridge this gap by proactively managing congestion, forecasting traffic variations, and dynamically reallocating resources, enabling performance that approaches the deterministic reliability of IntServ while retaining the flexibility of DiffServ.

Scalability further differentiates the three frameworks—IntServ suffers from high overhead due to per-flow state maintenance on every device, making it unsuitable for large-scale networks, whereas DiffServ achieves superior scalability by aggregating traffic into a small number of classes. In contrast, AI-based QoS models, particularly those integrated with Software-Defined Networking (SDN), are inherently scalable and capable of handling network complexity far beyond traditional manual configurations. These systems are limited only by the computational capacity of the AI engine, which can be elastically scaled using cloud or edge resources.

When it comes to resource utilization, IntServ often results in inefficiency due to static reservations that lock bandwidth even when flows are inactive, while DiffServ, despite being more efficient, can still waste resources due to rigid, pre-configured policies that fail to adapt to changing conditions. AI-driven systems overcome these inefficiencies through continuous monitoring, adaptive optimization, and predictive analytics, enabling dynamic allocation and reclamation of bandwidth in real time. This results in sustainably higher resource utilization, cost-effectiveness, and responsiveness, establishing AI as a pivotal enabler for intelligent QoS management in future network infrastructures.

C. Applicability and Use Cases

The IntServ, DiffServ, and AI-enhanced QoS models each excel in distinct network environments based on their design principles and operational strengths. The IntServ model is most appropriate for small, controlled, and mission-critical networks where deterministic performance guarantees are

essential, such as in military command-and-control infrastructures, dedicated industrial automation systems, or compact campus networks with a manageable number of real-time flows. Its resource reservation mechanism ensures strict quality assurance but limits scalability, making it unsuitable for large-scale implementations. DiffServ, by contrast, serves as the preferred standard for large enterprise, campus, and service provider networks where scalability, efficiency, and simplicity are prioritized. It aggregates traffic into service classes to offer differentiated service levels without maintaining per-flow state, providing a practical and scalable solution for managing varied QoS requirements across numerous users. However, its statistical guarantees make it less suited for highly dynamic or unpredictable scenarios.

AI-enhanced QoS represents the next evolutionary stage, designed for modern, complex, and adaptive network infrastructures such as 5G, which demand ultra-low latency, high throughput, and intelligent orchestration to support use cases like enhanced mobile broadband and large-scale IoT deployments. It is particularly advantageous in environments like tactical Mobile Ad Hoc Networks (MANETs), cloud-native infrastructures, and mission-critical systems where predictive maintenance, real-time decision-making, and dynamic adaptability are necessary to sustain optimal performance and reliability. Through continuous learning, predictive analytics, and automation, AI-driven QoS systems provide the flexibility and intelligence required to meet the demands of next-generation networking.

V. CHALLENGES AND OPEN ISSUES

Despite the significant advantages of AI-enhanced QoS, several major challenges and open issues must be addressed before widespread adoption.

A. Technical and Algorithmic Challenges

AI-driven QoS optimization, while highly promising, faces several ongoing challenges that must be addressed to achieve efficient and sustainable deployment in real-world network environments. One major concern lies in the computational complexity of many deep learning architectures, which imposes significant processing and energy demands. For time-sensitive applications operating at the network edge, this necessitates the design of lightweight and resource-efficient AI models capable of delivering intelligent decision-making without introducing notable latency or computational overhead. Equally critical are data-related challenges, as the accuracy and generalization ability of machine learning models depend heavily on the quality, diversity, and quantity of training data. Building large, representative datasets remains difficult due to the complexities of data collection, potential biases, and the labor-intensive process of labeling network traffic flows according to specific QoS parameters.

Future research must therefore emphasize advanced data augmentation, transfer learning, and synthetic environment generation to support robust AI training. Additionally, maintaining model performance over time presents its own difficulty due to "model drift," a condition where learned policies degrade as underlying network patterns evolve. This problem underscores the need for continuous learning mechanisms, periodic retraining pipelines, and adaptive model architectures that can evolve alongside dynamic network conditions. Together, addressing these challenges will be key to realizing fully autonomous, resilient, and scalable AI-driven QoS systems capable of adapting seamlessly to future networking demands.

B. Architectural and Integration Challenges

The shift toward AI-enhanced networking introduces several significant implementation challenges that must be carefully managed to ensure successful adoption and operation. One of the foremost issues is integration with legacy infrastructure; the transition from traditional networking systems to AI-driven platforms is rarely straightforward, as existing network hardware and software—often proprietary and vendor-specific—may be incompatible or only partially compatible with new AI technologies. This can necessitate complex adjustments, middleware development, or outright replacement of components, all of which increase cost and implementation complexity.

In parallel, security and privacy concerns become more prominent in AI-powered networks due to their extensive reliance on real-time telemetry and other sensitive data feeds. Safeguarding this data from breaches and ensuring that AI models are resistant to adversarial manipulations are critical research priorities that must be addressed to maintain trust and operational integrity. Another vital consideration is the explainability and trustworthiness of AI-driven systems. Many advanced AI models operate as "black boxes," generating decisions that are difficult for network operators and administrators to interpret or justify. As organizations expect transparency and auditability—especially when lives, assets, or compliance are at stake—ongoing research into explainable AI (XAI) remains essential. Empowering administrators to understand, validate, and trust automated recommendations and actions will be crucial for the broader acceptance and integration of AI within production network environments.

VI. CONCLUSION

This paper highlights how traditional Quality of Service (QoS) mechanisms, especially Integrated Services (IntServ) and Differentiated Services (DiffServ), have played a crucial role in managing network performance. However, they face limitations due to their static, rule-based nature. This lack of adaptability and scalability has paved the way for the rise of AI-enhanced architectures. AI brings a revolutionary approach to QoS, allowing for real-time, dynamic, and predictive network management—something traditional methods simply can't achieve. The collaboration between AI and technologies like Software-Defined Networking (SDN) and 5G isn't just a coincidence; it's essential. These architectures offer the programmable control and application drivers that enable AI to truly shine. While there are still significant challenges ahead, the future of QoS is clearly heading towards a hybrid model. In this model, intelligent systems will serve as a dynamic orchestration layer, enhancing and optimizing traditional mechanisms to provide unmatched levels of performance, scalability, and reliability, thus meeting the needs of both current and future communication systems.

VII. FUTURE SCOPE

A practical future direction lies in a hybrid model where AI enhances rather than replaces traditional Quality of Service (QoS) mechanisms. Acting as an intelligent orchestration layer over a DiffServ framework, AI could predict congestion and dynamically adjust Per-Hop Behavior (PHB) weights or traffic shaping policies in real time. This merges the reliability of classical systems with AI's adaptability. Future research should develop hybrid AI models—such as Graph Convolutional Networks for prediction and Deep Reinforcement Learning for optimization—to advance network management. Similarly, in speech recognition under noisy conditions, the emphasis is on adaptive, privacy-aware systems using multimodal deep learning, self-supervised pre-training, edge computing, and context-aware integration for robust real-world performance.

REFERENCES

[1] M. Natkaniec, K. Kosek-Szott, S. Szott and G. Bianchi, "A Survey of Medium Access Mechanisms for Providing QoS in Ad-Hoc Networks," *IEEE communications surveys & tutorials,* vol. 15, no. 2, pp. 592-620, 2012.

[2] G. Sunkara, "The Role of AI and Machine Learning in Enhancing SD-WAN Performance," *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology,* vol. 14, no. 4, pp. 1-9, 7 December 2022.

[3] M. Karakus and A. Durresi, "Quality of service (QoS) in software defined networking (SDN): A survey," *Journal of Network and Computer Applications,* vol. 80, no. 1, pp. 200-218, 15 February 2017.

[4] K. Bouraqia, E. Sabir, M. Sadik and L. Ladid, "Quality of experience for streaming services: measurements, challenges and insights," *IEEE Access,* vol. 8, no. 1, pp. 13341-13361, 2020.

[5] Z. Mammeri, "Framework for parameter mapping to provide end-to-end QoS guarantees

in IntServ/DiffServ architectures," *Computer Communications,* vol. 28, no. 9, pp. 1074-1092, 2 June 2005.

[6] S. R. Lima, P. Carvalho and V. Freitas, "Admission control in multiservice IP networks: architectural issues and trends," *IEEE Communications Magazine,* vol. 45, no. 4, pp. 114-121, 16 April 2007.

[7] M. H. Mohamed, B. H. Ali, A. I. Taloba, A. O. Aseeri, M. A. Elaziz and S. El-Sappagah, "Toward an Accurate Liver Disease Prediction Based on Two-Level Ensemble Stacking Model," *IEEE Access,* vol. 12, no. 1, pp. 180210-180237, September 2024.

[8] A. Bahnasse, F. E. Louhab, H. A. Oulahyane, M. Talea and A. Bakali, "Novel SDN architecture for smart MPLS traffic engineering-DiffServ aware management," *Future Generation Computer Systems,* vol. 87, no. 1, pp. 115-126, 1 October 2018.

[9] M. Radivojević and M. Petar, "Quality of Service Implementation," *The Emerging WDM EPON,* vol. 1, no. 1, pp. 35-66, 13 May 2017.

[10] L. Han, Y. Qu, L. Dong and R. Li, "Flow-Level QoS Assurance via IPv6 In-Band Signalling," in *27th Wireless and Optical Communication Conference (WOCC 2018)*, Hualien, Taiwan, 2018.

