

LSTM-Based Packet Loss Prediction in Communication Networks: A Comprehensive Review

Neelam Rani, Tejinder Kaur, Harmeet Kaur, Hitakshi, Manpreet Singh

^{1, 2, 3, 4, 5} Assistant Professor, Department of Computer Science Engineering, Sant baba bhag Singh university, Khiala, Punjab

Abstract— Packet loss is a critical factor affecting the performance, reliability, and Quality of Service (QoS) of modern communication networks. Accurate packet loss prediction enables proactive congestion management, efficient resource allocation, and improved network reliability. With the rapid growth of 5G/6G networks, Internet of Things (IoT), cloud computing, and Software-Defined Networking (SDN), traditional statistical and machine learning techniques have become insufficient for modelling the complex temporal characteristics of network traffic. Long Short-Term Memory (LSTM), a deep learning architecture designed for sequential data, has emerged as an effective solution due to its ability to capture long-term dependencies and nonlinear traffic patterns. This review presents a comprehensive analysis of LSTM-based packet loss prediction techniques by examining conventional prediction methods, recent deep learning models, comparative performance, application domains, research challenges, and future research directions. The review highlights the effectiveness of LSTM while identifying current limitations and emerging opportunities for developing intelligent, adaptive, and efficient packet loss prediction frameworks for next-generation communication networks.

Keywords— Packet Loss Prediction, Long Short-Term Memory (LSTM), Deep Learning, Network Traffic Prediction, Congestion Control, Time Series Forecasting, Quality of Service (QoS), Software-Defined Networking (SDN), Internet of Things (IoT), 5G/6G Networks.

1. Introduction

The rapid growth of modern communication networks, including the Internet of Things (IoT), cloud computing, Software-Defined Networking (SDN), and fifth- and sixth-generation (5G/6G) wireless networks, has significantly increased the demand for reliable and efficient data transmission [19]. As network traffic continues to grow in complexity and volume, maintaining consistent Quality of Service (QoS) and Quality of Experience (QoE) has become a major challenge [1], [7]. Among the various network performance issues, packet loss remains one of the most critical factors affecting communication reliability [57]. Packet loss occurs when one or more data packets fail to reach their intended destination during transmission, resulting in degraded network performance, increased latency, reduced throughput, and poor user experience. [22]. Applications such as video streaming, online gaming, Voice over Internet Protocol (VoIP), telemedicine, and autonomous systems are particularly sensitive to packet loss, making its prediction and mitigation essential [18], [61].

Accurate packet loss prediction enables network administrators to anticipate congestion and implement proactive traffic management strategies before network degradation occurs [57]. Traditional approaches generally detect packet loss after it has already happened, limiting their effectiveness in preventing performance deterioration [22]. Predictive models, on the other hand, estimate future packet loss based on historical network behaviour, allowing intelligent routing, congestion avoidance, dynamic resource allocation, and adaptive transmission control [19]. As modern networks generate continuous streams of temporal data, predicting packet loss has become a time-series forecasting problem that requires advanced analytical techniques capable of modelling complex sequential dependencies [2], [5].

In recent years, deep learning has emerged as a powerful solution for network traffic analysis due to its ability to automatically learn complex patterns from large-scale data without extensive manual feature engineering [3], [10]. Unlike conventional machine learning algorithms, deep learning models can effectively capture nonlinear relationships and temporal dependencies present in network traffic [19]. Among various deep learning architectures, Long Short-

Term Memory (LSTM) networks have gained considerable attention because of their ability to retain long-term information and overcome the vanishing gradient problem commonly encountered in traditional Recurrent Neural Networks (RNNs)[63]. These characteristics make LSTM particularly suitable for packet loss prediction, where historical traffic conditions significantly influence future network behaviour [22].

The primary objective of this review is to provide a comprehensive analysis of LSTM-based packet loss prediction techniques reported in the recent literature [57]. This paper reviews conventional statistical and machine learning approaches before examining the evolution of deep learning methods, particularly LSTM architectures [22]. Furthermore, it compares existing methodologies, datasets, evaluation metrics, and application domains while identifying current research challenges, limitations, and future research directions [36]. By presenting a structured overview of the existing body of knowledge, this review aims to serve as a valuable reference for researchers and practitioners working in intelligent network traffic prediction.

The remainder of this paper is organised as follows. Section 2 presents the fundamental concepts of packet loss, including its causes, performance impact, Packet Loss Rate (PLR), and its relationship with QoS and QoE[1], [7], [18]. Section 3 reviews conventional packet loss prediction techniques based on statistical and machine learning methods [22], [57]. Section 4 discusses the fundamentals of Long Short-Term Memory networks and their application to packet loss prediction [63]. The subsequent sections analyse recent literature, compare existing approaches, identify research gaps, discuss future research opportunities, and conclude the review.

2. Fundamentals of Packet Loss

Packet loss refers to the failure of one or more data packets to successfully travel from the source to the destination during network communication [19]. In packet-switched networks, data are divided into smaller packets that traverse different paths before being reassembled at the receiving end [22]. When packets are dropped due to transmission errors or network limitations, the receiving device either requests retransmission or experiences missing information, depending on the communication protocol [57]. Excessive packet loss reduces communication efficiency and negatively affects both network performance and user satisfaction [1], [7].

Several factors contribute to packet loss in modern communication networks. Network congestion remains the most common cause, occurring when routers and switches receive more packets than they can process, resulting in buffer overflow and packet dropping [18], [61]. Hardware failures, faulty cables, damaged network interfaces, and malfunctioning routing devices may also lead to packet loss [19]. Wireless communication introduces additional challenges such as signal interference, fading, weak signal strength, and environmental obstacles that increase transmission errors [47]. Furthermore, software configuration errors, routing protocol failures, overloaded servers, cyberattacks such as Distributed Denial-of-Service (DDoS), and insufficient bandwidth can significantly increase packet loss in large-scale networks [57].

The impact of packet loss extends beyond simple data retransmission. High packet loss increases communication delay, decreases throughput, wastes bandwidth, and introduces jitter that affects real-time applications [18]. Interactive services such as video conferencing, online gaming, and Voice over Internet Protocol (VoIP) are particularly vulnerable, as packet loss can produce audio distortion, video freezing, interrupted communication, and degraded user experience [1], [7]. In industrial automation, healthcare systems, and autonomous transportation, excessive packet loss may compromise operational reliability and system safety [61].

Packet Loss Rate (PLR) is a widely used metric for evaluating network reliability [57]. It represents the proportion of lost packets relative to the total number of transmitted packets over a specified period [22]. PLR is mathematically expressed as the ratio of lost packets to transmitted packets, usually presented as a percentage [57]. A lower PLR indicates a more reliable communication channel, whereas higher values suggest network congestion, instability, or communication failures [19]. PLR is commonly employed as both a performance indicator and a prediction target in intelligent network management systems [57].

Packet loss directly influences both Quality of Service (QoS) and Quality of Experience (QoE) [1], [7]. QoS focuses on objective network performance parameters such as throughput, delay, jitter, bandwidth utilisation, and packet delivery reliability [18]. In contrast, QoE represents the subjective perception of end users regarding service quality and application performance [1]. Maintaining low packet loss is therefore essential for ensuring high QoS while simultaneously improving users' overall experience across multimedia streaming, cloud services, IoT applications, and next-generation communication networks [7], [61].

3. Conventional Packet Loss Prediction Techniques

Before the emergence of deep learning, packet loss prediction primarily relied on statistical modelling and conventional machine learning techniques [57]. These approaches aimed to estimate future network conditions based on historical observations and predefined mathematical relationships [22]. Although these methods achieved satisfactory performance in relatively stable network environments, their ability to model highly dynamic and nonlinear traffic patterns remained limited [19].

Among statistical approaches, the Autoregressive Integrated Moving Average (ARIMA) model has been widely used for forecasting network traffic and packet loss [57]. ARIMA analyses historical time-series data to identify temporal trends and predict future values using autoregressive and moving-average components [22]. While ARIMA performs well for linear and stationary datasets, its prediction accuracy decreases significantly when network traffic exhibits sudden fluctuations or nonlinear behaviour [19].

Markov Models have also been extensively applied to packet loss prediction by representing network conditions as a sequence of probabilistic state transitions [57]. These models estimate future packet loss based on the current network state and predefined transition probabilities [22]. Although Markov-based methods are computationally efficient and mathematically interpretable, they often assume memoryless transitions, limiting their ability to capture long-term dependencies present in modern network traffic [19].

Hidden Markov Models (HMMs) extend traditional Markov Models by introducing hidden states that cannot be directly observed but are inferred from observable network measurements [57]. HMMs have demonstrated improved capability in modelling uncertain and stochastic network behaviour [22]. However, their performance depends heavily on accurate parameter estimation and predefined probability distributions, making them less suitable for highly dynamic communication environments [19].

With the advancement of artificial intelligence, machine learning techniques became increasingly popular for packet loss prediction [57]. Decision Tree algorithms classify network conditions by constructing hierarchical decision rules based on traffic features [22]. They provide fast prediction and high interpretability but are prone to overfitting when dealing with complex datasets [19]. Support Vector Machines (SVMs) identify optimal decision boundaries for packet loss classification and regression problems and often perform well with limited datasets [57]. However, SVM performance depends on kernel selection and parameter tuning, making large-scale implementation computationally expensive [22].

Ensemble learning algorithms such as Random Forest improve prediction accuracy by combining multiple decision trees, thereby reducing variance and improving generalisation [57]. Random Forest demonstrates strong robustness against noisy network data but requires increased computational resources as the number of trees grows [22]. More recently, Extreme Gradient Boosting (XGBoost) has gained attention for its ability to achieve high predictive accuracy through gradient boosting optimisation [57]. XGBoost effectively captures nonlinear relationships and handles missing data efficiently; however, extensive hyperparameter tuning and increased model complexity remain practical challenges [19].

Despite their contributions, conventional statistical and machine learning methods exhibit several limitations [57]. Most algorithms require manual feature engineering, struggle to capture long-term temporal dependencies, and experience reduced performance when network traffic becomes highly dynamic and nonlinear [22]. These limitations have encouraged researchers to explore deep learning techniques capable of automatically learning hierarchical representations from sequential network traffic data [19].

4. LSTM for Packet Loss Prediction

Recurrent Neural Networks (RNNs) were developed to process sequential data by maintaining information from previous time steps through recurrent connections [19]. Unlike traditional feed forward neural networks, RNNs utilise internal memory to model temporal relationships within time-series data, making them suitable for applications such as speech recognition, natural language processing, financial forecasting, and network traffic prediction [22]. However, conventional RNNs suffer from the vanishing and exploding gradient problems during training, preventing them from effectively learning long-term dependencies in lengthy sequences [63].

Long Short-Term Memory (LSTM) networks were introduced to overcome these limitations by incorporating specialised memory cells and gating mechanisms that regulate information flow [63]. An LSTM cell consists of three primary gates: the forget gate, input gate, and output gate [57]. The forget gate determines which historical

information should be discarded, the input gate controls the storage of new information, and the output gate decides which information is propagated to the next hidden state [22]. This architecture enables LSTM networks to preserve important information over long sequences while selectively forgetting irrelevant data, thereby significantly improving prediction performance [63].

The working principle of LSTM involves sequentially processing network traffic observations while continuously updating the internal memory state [57]. At each time step, the model receives current network features, combines them with historical information stored in the memory cell, and generates predictions for future packet loss [22]. Through iterative training using backpropagation through time (BPTT), LSTM learns complex temporal relationships that conventional statistical and machine learning models often fail to capture [63]. Consequently, LSTM effectively models dynamic network behaviour characterised by nonlinear traffic fluctuations, congestion events, and varying communication conditions [57].

LSTM has become particularly suitable for packet loss prediction because network traffic inherently exhibits temporal dependencies, where previous network conditions strongly influence future performance [19]. Parameters such as packet arrival rate, queue occupancy, bandwidth utilisation, transmission delay, jitter, and congestion levels evolve continuously over time rather than independently [22]. LSTM's ability to retain long-term contextual information enables accurate modelling of these evolving patterns, resulting in improved packet loss prediction under diverse networking scenarios, including IoT systems, cloud computing, SDN, wireless communication, and emerging 5G and 6G networks [57].

Compared with conventional statistical models and traditional machine learning algorithms, LSTM offers several significant advantages [63]. It automatically learns complex feature representations directly from raw sequential data, reducing the need for manual feature engineering [57]. Its capability to capture nonlinear relationships and long-term dependencies improves prediction accuracy under highly dynamic network conditions [19]. Furthermore, LSTM demonstrates strong adaptability across diverse communication environments and large-scale datasets [22]. These advantages have established LSTM as one of the most widely adopted deep learning architectures for intelligent packet loss prediction and proactive network traffic management [63].

5. Literature Review

[1] Yao et al. developed a data-driven framework for TCP that achieved flexible quality of service (QoS) control in mobile data networks. Their framework established foundational principles for adaptive network management in heterogeneous mobile environments. The authors demonstrated through extensive experimentation that their approach effectively balanced multiple QoS objectives while maintaining TCP compatibility.

[2] Das et al. introduced a decoder-only foundation model for time-series forecasting. Their model demonstrated the potential of transformer architectures in capturing temporal dependencies across network traffic patterns. The authors showed that their foundation model achieved superior performance compared to traditional forecasting approaches through extensive empirical validation.

[3] Zhou et al. proposed a machine learning-based framework for dynamic selection of congestion control algorithms. Their framework significantly advanced the field by enabling context-aware algorithm switching based on real-time network conditions. The authors validated their approach through comprehensive experiments showing improved network performance across diverse scenarios.

[4] Xia et al. presented a multi-objective reinforcement learning perspective on internet congestion control. Their work addressed the complex trade-offs inherent in network optimization by considering multiple performance objectives simultaneously. The authors demonstrated that their reinforcement learning approach effectively balanced competing objectives in dynamic network environments.

[5] Nie et al. contributed to the forecasting domain by demonstrating that time series data could be effectively represented using limited vocabulary sizes in transformer models. Their work achieved long-term prediction capabilities with remarkable efficiency. The authors showed that their approach maintained high accuracy while significantly reducing computational requirements.

[6] Du et al. introduced a unified congestion control framework that accommodated diverse application preferences and network conditions. Their framework provided a comprehensive solution for varying quality-of-service requirements across different applications. The authors validated their framework through extensive experiments showing adaptability to diverse network scenarios.

[7] Yao et al. extended their previous work by incorporating deep reinforcement learning (DRL) to optimize QoS over mobile networks. Their unified framework demonstrated significant improvements in network performance through adaptive learning mechanisms. The authors showed that their DRL-based approach achieved superior performance compared to traditional methods in mobile environments.

[8] Tian et al. addressed the critical challenge of fairness generalizability in learning-based congestion control through their Jury framework. Their work ensured equitable resource allocation across heterogeneous network environments. The authors demonstrated through rigorous experimentation that their framework achieved fair resource distribution without compromising overall efficiency.

[9] Hong et al. advanced long-term multi-energy load forecasting by developing Patchformer, a patch and transformer-based approach that captured both local and global patterns effectively. Their model achieved state-of-the-art performance in energy load prediction tasks. The authors showed that their patch-based approach effectively captured multi-scale temporal patterns in energy data.

[10] Rahmatinia and Hosseini-Seno implemented an attention-based deep learning model for multi-horizon prediction of load, price, and wind power generation. Their model demonstrated the versatility of transformer architectures across diverse forecasting applications. The authors validated their approach through comprehensive experiments showing accurate predictions across multiple prediction horizons.

[11] Han et al. introduced Anole, which pragmatically blended classic and learning-based algorithms in congestion control. Their hybrid approach achieved superior performance through complementary strengths of traditional and modern techniques. The authors demonstrated that their blended approach effectively leveraged the advantages of both algorithmic paradigms.

[12] Zhou et al. developed Antelope, a framework for dynamic selection of congestion control algorithms that operated at the network protocol level. Their framework enabled real-time adaptation to changing network conditions. The authors showed through extensive experimentation that their framework achieved significant performance improvements across diverse network conditions.

[13] Lu et al. refined multivariate forecasting with adaptive temporal-contextual learning through their ARM framework. Their framework effectively captured both temporal and contextual dependencies in multivariate time series data. The authors demonstrated that their adaptive approach achieved superior performance compared to static forecasting methods.

[14] Liao et al. presented Astraea, which focused on achieving fairness and efficiency in learning-based congestion control through innovative architectural designs. Their framework addressed the persistent challenge of maintaining fairness while optimizing for efficiency. The authors validated their approach through comprehensive experiments showing balanced performance across fairness and efficiency metrics.

[15] Niu et al. explored attention mechanisms as robust representations for time series forecasting. Their work established attention-based representations as effective features for forecasting tasks. The authors demonstrated that attention mechanisms provided robust and interpretable representations for time series data.

[16] Wu et al. developed Autoformer, which leveraged decomposition transformers with auto-correlation for long-term series forecasting. Their approach achieved state-of-the-art performance through innovative architectural design.

The authors showed that their decomposition-based approach effectively captured both trend and seasonal components in time series data.

[17] Wu et al. introduced AutoPlex for inter-session multiplexing congestion control in large-scale live video services. Their work addressed the unique challenges of real-time multimedia delivery across multiple sessions. The authors demonstrated that their multiplexing approach effectively managed multiple concurrent video streams.

[18] Xu et al. proposed congestion control as a service framework for low-latency mobile uploading. Their approach provided on-demand congestion control capabilities for latency-sensitive applications. The authors validated their framework through experiments showing significant latency reductions in mobile upload scenarios.

[19] Wei et al. provided a comprehensive survey on the renaissance of congestion control with machine learning. Their review documented the evolution from classical to learning-based approaches in congestion control. The authors synthesized the state-of-the-art and identified key research directions for future investigation.

[20] Jiang et al. performed analysis and improvement of the delay-based Copa congestion control algorithm. Their work contributed significantly to the theoretical understanding of delay-sensitive protocols. The authors proposed enhancements that improved Copa's performance in diverse network conditions.

[21] Zhang and Yan developed Crossformer, which utilized cross-dimension dependency for multivariate time series forecasting. Their model effectively captured dependencies across different dimensions of multivariate data. The authors demonstrated through extensive experiments that their cross-dimensional approach achieved superior prediction accuracy.

[22] Mishra et al. proposed a delay-driven machine learning scheme for TCP data flow performance enhancement. Their scheme demonstrated significant improvements in data flow efficiency through delay-aware optimization. The authors validated their approach through experiments showing enhanced TCP performance in variable network conditions.

[23] Emara et al. introduced Eagle, which refined congestion control by learning from expert algorithms. Their approach demonstrated how learning from established algorithms could improve congestion control performance. The authors showed that their expert-guided learning approach achieved superior performance compared to purely learning-based methods.

[24] Fan et al. presented EchoCC, which refined learning-based congestion control with a WordBook mechanism. Their innovative approach enhanced the adaptability of learning-based congestion control systems. The authors demonstrated through experiments that their WordBook mechanism effectively captured and utilized historical network patterns.

[25] Tian et al. developed Spine, an efficient DRL-based congestion control system with ultra-low overhead. Their system addressed the computational constraints of real-time network operations while maintaining high performance. The authors validated their approach through extensive experiments showing minimal overhead with superior performance.

[26] Xing et al. applied patch channel-mixing transformer models to electricity load forecasting. Their approach effectively captured both temporal and channel-wise dependencies in electricity data. The authors demonstrated that their patch-based approach achieved accurate load predictions across multiple time horizons.

[27] Rahmatinia and Seno enhanced NTMA with simultaneous multi-QoS parameter prediction using transformer-based deep learning. Their approach enabled comprehensive QoS prediction across multiple parameters simultaneously. The authors showed that their multi-task learning approach effectively predicted various QoS metrics with high accuracy.

[28] Zhou et al. developed FEDformer, which incorporated frequency-enhanced decomposed transformers for long-term series forecasting. Their model achieved superior performance through frequency domain analysis. The authors demonstrated that their frequency-enhanced approach effectively captured periodic patterns in time series data.

[29] Li et al. proposed FPN-fusion, an enhanced linear complexity time series forecasting model. Their model achieved competitive performance with reduced computational complexity. The authors showed that their approach maintained high accuracy while significantly reducing training and inference times.

[30] Lan and Oermann introduced Gateformer, advancing multivariate time series forecasting through temporal and variate-wise attention with gated representations. Their model effectively balanced attention across temporal and variate dimensions. The authors demonstrated through experiments that their gated attention mechanism achieved superior performance across multiple forecasting benchmarks.

Table 5.1 Tabular form Literature Review

Sr.No	Author Name	Year	Jounral/Conference	Key Findings
31	Yang et al.	2023	IEEE INFOCOM 2023	Proposed Gemini , a divide-and-conquer learning-based internet congestion control framework that decomposes congestion control into smaller sub-problems, achieving practical deployment and improved network performance.
32	Zhao et al.	2025	International Joint Conference on Neural Networks (IJCNN)	Developed GLoSyformer , a global-local synergistic transformer that enhances multivariate time-series forecasting by capturing both global and local dependencies.
33	Li et al.	2025	International Seminar on Computer Technology, Mechanical and Electrical Engineering (ISCME)	Proposed Hadaformer , which employs Hadamard attention to reduce computational complexity while maintaining high forecasting accuracy.
34	Khanal et al.	2025	IEEE International Conference on Data Mining Workshops (ICDMW)	Introduced InfoST , an information-theoretic framework for evaluating spatio-temporal prediction models with rigorous performance metrics.
35	Cao et al.	2023	AAAI Conference on Artificial Intelligence	Developed InParformer , an evolutionary decomposition transformer with interactive parallel attention that improves long-term time-series forecasting performance.

36	Abbasloo	2023	arXiv	Presented a comprehensive benchmark of Internet congestion control algorithms, providing standardized evaluation methodologies for future research.
37	Liu et al.	2023	arXiv	Proposed iTransformer , an inverted transformer architecture that achieves competitive time-series forecasting performance with lower computational complexity.
38	Xu et al.	2024	IEEE International Conference on Communications (ICC)	Developed KerDqn , a kernel-level deep reinforcement learning congestion control algorithm that reduces latency and improves network efficiency.
39	Du et al.	2025	IEEE/ACM Transactions on Networking	Proposed Libra , an adaptive congestion control framework capable of handling diverse application requirements and varying network conditions.
40	Huang et al.	2025	International Conference on Machine Learning and Computing Applications (ICMLCA)	Introduced LWaveNet , combining learnable wavelet transforms and differential attention to improve multi-scale time-series forecasting accuracy.
41	Dong et al.	2024	IEEE Smart World Congress (SWC)	Developed MAI-Unet , a multi-scale attention interactive network that enhances long-term multivariate time-series forecasting performance.
42	Jia et al.	2024	ACM Web Conference	Proposed a hybrid congestion control algorithm that effectively mitigates RTT jitter and improves network stability under dynamic conditions.
43	Bai et al.	2025	IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	Introduced MetaCon , a meta-reinforcement learning-based congestion control framework capable of rapidly adapting to changing network

				environments.
44	Klotergens et al.	2025	Semantic Scholar (Preprint)	Investigated mixer networks for irregular multivariate time-series forecasting, demonstrating competitive prediction performance on irregular datasets.
45	Luo and Wang	2024	Semantic Scholar (Preprint)	Proposed ModernTCN , a pure convolution-based architecture that achieves forecasting performance comparable to transformer models with lower complexity.
46	Liu et al.	2021	IEEE Global Communications Conference (GLOBECOM)	Developed a constrained reinforcement learning approach for multi-objective network congestion control, effectively balancing multiple performance objectives.
47	Zhong et al.	2022	IEEE INFOCOM	Proposed Muses , a lightweight learning-based congestion control framework designed for resource-constrained mobile devices with improved efficiency.
48	Liu et al.	2022	Advances in Neural Information Processing Systems (NeurIPS)	Introduced Non-stationary Transformers , addressing non-stationary time-series forecasting through specialized transformer architectures.
49	Yue et al.	2025	arXiv	Proposed OLinear , a lightweight linear forecasting model based on orthogonal transformations, achieving competitive accuracy with low computational cost.
50	Zhou et al.	2023	Advances in Neural Information Processing Systems (NeurIPS)	Introduced One Fits All , demonstrating that pretrained language models can achieve state-of-the-art performance across a wide range of time-series analysis tasks.

51	Xiao et al.	2022	IEEE International Conference on Big Data	Proposed an optimal rate control mechanism for latency-constrained, high-throughput big data applications, improving throughput while satisfying latency requirements.
52	Sacco et al.	2023	IEEE Transactions on Network and Service Management	Introduced a reinforcement learning-based partially oblivious congestion control framework that improves network efficiency while adapting to dynamic Internet conditions.
53	Wei et al.	2022	China Communications	Developed Plume , a lightweight deep reinforcement learning-based congestion control framework that provides efficient and generalized congestion management.
54	Zhou et al.	2023	arXiv	Proposed a pretrained language model (PLM) approach for time-series forecasting, demonstrating the effectiveness of transfer learning for forecasting tasks.
55	Gu et al.	2025	ACM International Conference on Web Search and Data Mining (WSDM)	Developed ProCC , a programmatic reinforcement learning framework for transparent and efficient TCP congestion control with improved interpretability.
56	Zhang et al.	2023	China Communications	Proposed a QoS-aware online learning-based congestion control algorithm that dynamically optimizes Quality of Service under varying network conditions.
57	Welzl et al.	2024	IEEE Access	Proposed a machine learning-based real-time TCP packet loss prediction model that accurately predicts packet loss and improves TCP performance.
58	Wang et al.	2024	arXiv	Introduced a global-local timestamp fusion

				framework that improves robust time-series forecasting by effectively utilizing timestamp information.
59	Li et al.	2023	arXiv	Investigated linear mapping techniques for long-term time-series forecasting, showing that simple linear models can achieve competitive performance.
60	Xia et al.	2021	International Joint Conference on Neural Networks (IJCNN)	Proposed RLCC , a practical reinforcement learning-based Internet congestion control framework that improves adaptability and network performance.
61	Zhang et al.	2024	IEEE International Conference on Sensing, Communication and Networking (SECON)	Developed RTCC , an end-to-end sub-RTT congestion control mechanism designed for next-generation high-speed communication networks.
62	Zhang et al.	2023	IEEE Internet of Things Journal	Proposed SageFormer , a series-aware transformer framework that effectively captures temporal dependencies for long-term multivariate time-series forecasting.
63	Dai et al.	2024	Journal of Supercomputing	Introduced SALSTM , a segmented self-attention Long Short-Term Memory model that enhances long-term forecasting accuracy by integrating attention mechanisms with LSTM.
64	Yang and Gonsalves	2025	Forecasting	Developed SegmentedCrossformer , an enhanced transformer architecture that captures both cross-time and cross-dimensional relationships for multivariate forecasting.

65	Akkineni and Uddagiri	2025	IEEE International Conference on Big Data	Proposed Sig-Patchformer , a path signature-based transformer that improves forecasting efficiency while reducing computational complexity.
66	Zhang et al.	2024	International Joint Conference on Artificial Intelligence (IJCAI)	Introduced Skip-Timeformer , a transformer architecture utilizing skip-time interaction to improve long-sequence time-series forecasting performance.
67	Jia et al.	2025	IEEE Symposium on Computers and Communications (ISCC)	Proposed a deep reinforcement learning-based SLO-targeted congestion control framework that optimizes network performance according to service-level objectives.
68	Tian et al.	2022	ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)	Developed Spine , an efficient deep reinforcement learning-based congestion control algorithm with ultra-low computational overhead and high adaptability.
69	Garcia Remy and	2026	arXiv	Performed a statistical benchmarking of transformer models for low signal-to-noise time-series forecasting, highlighting their strengths and limitations under noisy conditions.
70	Chen et al.	2025	International Conference on Computing, Information and Big Data Applications	Proposed T-Mamba , a hybrid Mamba-Transformer framework that enhances stock price prediction through improved temporal feature extraction and forecasting accuracy.

71	Li et al.	2021	IEEE Journal on Selected Areas in Communications (JSAC)	Proposed TCP-NeuRoc , a neural adaptive TCP congestion control framework with online changepoint detection, enabling dynamic adaptation to changing network conditions and improving congestion control performance.
72	Sun et al.	2025	IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	Developed TCTformer , a dual-attention transformer architecture for long-term time-series forecasting that effectively captures temporal dependencies and improves forecasting accuracy.
73	Li et al.	2025	IEEE Transactions on Mobile Computing	Introduced the Concentrator Method , a learning-based framework that combines classical congestion control algorithms with machine learning modules, achieving improved adaptability and network performance.
74	Ying and Lu	2024	IEEE Access	Proposed TFEformer , a Temporal Feature Enhanced Transformer that strengthens temporal feature extraction for multivariate time-series forecasting, resulting in higher prediction accuracy.
75	Wang et al.	2024	arXiv	Developed TimeMixer , a decomposable multiscale mixing architecture that captures temporal patterns at different scales, improving the accuracy of time-series forecasting.
76	Wang et al.	2024	arXiv	Proposed TimeMixer++ , a universal time-series pattern learning framework capable of performing predictive analysis across diverse forecasting tasks with enhanced generalization.

77	Wu et al.	2022	arXiv	Introduced TimesNet , a temporal 2D-variation modeling framework that transforms one-dimensional temporal sequences into two-dimensional representations for improved time-series analysis and forecasting.
78	Tian et al.	2026	IEEE/ACM Transactions on Networking	Proposed a Multi-Agent Deep Reinforcement Learning (MADRL) -based congestion control framework that achieves fair and efficient bandwidth allocation while improving overall network performance in dynamic environments.

6. Comparative Analysis

The rapid evolution of communication networks has encouraged researchers to investigate various artificial intelligence and deep learning models for packet loss prediction. Selecting an appropriate prediction model depends on factors such as prediction accuracy, computational complexity, training time, memory requirements, and the ability to capture temporal dependencies in network traffic. Among the available approaches, Long Short-Term Memory (LSTM) has emerged as one of the most effective models due to its capability to learn long-term sequential patterns. This section compares LSTM with Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), CNN-LSTM hybrid models, Transformer-based architectures, and traditional machine learning techniques.

Recurrent Neural Networks (RNNs) were initially developed to process sequential data by retaining information from previous time steps through recurrent connections. Although RNNs are capable of modelling temporal relationships, they suffer from the vanishing and exploding gradient problems during training, which restrict their ability to learn long-term dependencies. As a result, their prediction performance declines when analysing long and highly dynamic network traffic sequences. LSTM addresses these limitations by incorporating memory cells along with input, forget, and output gates that selectively preserve relevant information. Consequently, LSTM generally achieves higher prediction accuracy and lower prediction errors, including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), compared with conventional RNNs.

Gated Recurrent Unit (GRU) is a simplified version of LSTM that combines the forget and input gates into a single update gate, reducing the number of trainable parameters. GRU models require less memory and shorter training time while providing prediction performance comparable to LSTM in many applications. However, LSTM often demonstrates better stability and superior performance when modelling complex network traffic with long-term temporal dependencies, making it more suitable for large-scale packet loss prediction tasks.

Hybrid CNN-LSTM models integrate Convolutional Neural Networks (CNNs) with LSTM networks to exploit both spatial and temporal characteristics of network traffic. CNN layers automatically extract important traffic features, while LSTM layers learn sequential dependencies. This hybrid architecture often achieves higher prediction accuracy, precision, and recall than standalone LSTM models. Nevertheless, CNN-LSTM models require larger datasets, longer training time, and higher computational resources because of their increased architectural complexity.

Transformer-based models have recently gained attention due to their self-attention mechanism, which enables efficient modelling of long-range dependencies without recurrent processing. Transformers generally outperform LSTM when trained on large-scale datasets and can process sequences in parallel, reducing inference time. However, they require significantly greater computational power, memory, and training data, making them less practical for

resource-constrained environments such as IoT and edge computing. In comparison, LSTM offers a better balance between prediction performance and computational efficiency for many real-world networking applications.

Traditional machine learning algorithms, including Decision Trees, Support Vector Machines (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost), remain popular because of their simplicity and lower computational requirements. These methods perform well on structured datasets but rely heavily on manual feature engineering and often fail to capture complex temporal relationships in network traffic. Compared with these approaches, LSTM automatically learns sequential features from historical traffic data, resulting in improved accuracy, higher precision and recall, and lower prediction errors. Although LSTM requires longer training time than conventional machine learning models, its ability to model nonlinear temporal patterns makes it one of the most reliable approaches for packet loss prediction in modern communication networks.

Table 6.1 Comparative analysis

Model	Accuracy	RMSE	MAE	Precision	Recall	Training Time	Computational Cost
RNN	Moderate	High	High	Moderate	Moderate	Low	Low
LSTM	High	Low	Low	High	High	Moderate	Moderate
GRU	High	Low	Low	High	High	Low–Moderate	Low–Moderate
CNN-LSTM	Very High	Very Low	Very Low	Very High	Very High	High	High
Transformer	Excellent	Very Low	Very Low	Excellent	Excellent	Very High	Very High
Decision Tree	Moderate	High	High	Moderate	Moderate	Very Low	Very Low
SVM	Moderate–High	Moderate	Moderate	High	Moderate	Moderate	Moderate
Random Forest	High	Moderate	Moderate	High	High	Moderate	Moderate
XGBoost	High	Low	Low	High	High	Moderate	Moderate

7. Applications

The ability to accurately predict packet loss has become increasingly important with the rapid evolution of intelligent communication systems. LSTM-based packet loss prediction has been successfully applied across a wide range of networking environments where maintaining reliable data transmission and ensuring service quality are critical. By forecasting future packet loss using historical traffic patterns, LSTM enables proactive network management, congestion avoidance, intelligent routing, and adaptive resource allocation.

In fifth-generation (5G) and emerging sixth-generation (6G) communication networks, packet loss prediction plays a significant role in supporting ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC). These applications require continuous monitoring of network conditions to maintain high reliability and minimise transmission failures. LSTM models help network operators anticipate congestion and dynamically optimise routing decisions, thereby improving throughput, reducing latency, and enhancing overall network reliability.

The Internet of Things (IoT) represents another major application area for LSTM-based packet loss prediction. IoT environments consist of numerous interconnected sensors and smart devices that continuously exchange data over heterogeneous wireless networks. Due to limited bandwidth, constrained energy resources, and fluctuating communication conditions, IoT systems are particularly vulnerable to packet loss. LSTM models analyse historical communication patterns to predict future transmission failures, enabling efficient resource management, energy conservation, and improved communication reliability for smart homes, healthcare monitoring systems, industrial automation, and environmental sensing applications.

Software-Defined Networking (SDN) benefits significantly from intelligent packet loss prediction because of its centralised network control architecture. SDN controllers continuously collect network statistics and utilise predictive

models to optimise traffic engineering and routing decisions. Integrating LSTM into SDN environments enables proactive congestion detection, dynamic load balancing, and adaptive flow management, ultimately improving Quality of Service while reducing network downtime and communication delays.

Cloud computing infrastructures also rely on packet loss prediction to maintain efficient communication among distributed servers, virtual machines, and data centres. Packet loss may lead to service degradation, increased response time, and reduced application performance. LSTM-based prediction models assist cloud service providers in identifying potential communication bottlenecks before performance deteriorates, enabling intelligent workload distribution, resource scheduling, and network optimisation.

Edge computing has emerged as a complementary paradigm to cloud computing by processing data closer to end devices. Since edge computing applications often support latency-sensitive services such as autonomous vehicles, industrial control systems, and smart healthcare, maintaining reliable communication is essential. LSTM-based packet loss prediction enables edge nodes to proactively manage network resources, minimise communication failures, and ensure real-time data delivery while operating under limited computational resources.

Video streaming platforms represent one of the most bandwidth-intensive Internet applications. Packet loss directly affects streaming quality by causing buffering delays, frame freezing, reduced video resolution, and interrupted playback. LSTM prediction models help streaming systems estimate future network conditions and dynamically adjust video bitrates, transmission rates, and buffering strategies, thereby improving viewers' Quality of Experience and reducing playback interruptions.

Smart transportation systems, including intelligent traffic management, connected vehicles, and autonomous driving technologies, depend heavily on continuous and reliable communication among vehicles, roadside infrastructure, and cloud services. Packet loss in these systems may compromise navigation accuracy, safety, and decision-making capabilities. LSTM-based prediction techniques enable proactive communication management by forecasting network disruptions, reducing transmission failures, and supporting reliable vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in intelligent transportation environments.

Overall, the versatility of LSTM has made it one of the most widely adopted deep learning techniques for packet loss prediction across modern communication systems. Its capability to model complex temporal relationships enables proactive network optimisation, enhances Quality of Service and Quality of Experience, and supports the reliable operation of next-generation networking technologies.

8. Future Research Directions

The growth of artificial intelligence, deep learning, and the latest communication networks provides significant benefits for packet loss prediction. Long Short-Term Memory (LSTM) has proven useful for predictive tasks in the field of packet loss prediction, but some problems still need to be solved in the field of interpretability, scalability, computing efficiency, and real-time implementation. Therefore, further research should be directed at creating more intelligent and adaptive prediction models that can meet the needs of communication systems on the rise.

One invitational approach is integrating Explainable Artificial Intelligence (XAI) into packet-loss forecasting systems. Most deep learning models, including LSTM, act as black-box technologies, making it challenging to unearth how forecasting decisions are arrived at. By employing explanation methods such as SHAP, LIME, and attention visualization, models can become more comprehensible, enabling administrators to recognize the most significant predictive traffic attributes that cause packet losses. Also, explainable models can raise users' trust and enable them to be successfully implemented in mission-critical domains such as healthcare, autonomous driving, and industrial automation.

One further new research field worth mentioning is labeled Federated Learning (FL). It can make numerous scattered devices work together to learn predictive algorithms without putting raw data into magazine. Since telecommunication networks produce enormous amounts of sensitive information, federated learning gives much more security while using lower amount of communication channels. Using LSTM combined with federated learning can allow making decentralized predictions about packet losses with the help of the Internet-of-things, mobile communications, and every other type of electronic devices without disclosing any data related to clients

The increased use of Tiny Machine Learning (TinyML) provides opportunities for packet loss prediction in environments that have limited resources. TinyML means that low-power devices with little memory and processing power can use small deep learning models. This means that developers can create compressed and optimized LSTM

models for TinyML platforms and thus be able to predict packet loss in smart homes, IoT devices, and wearables without using cloud computing.

In the future, research could be focused on hybrid models of Transformer and LSTM approaches. Although LSTMs can efficiently handle temporal relations in time series analysis, Transformers can also successfully capture long-range dependencies using their self-attention technology. Hybrid models offer an opportunity to use both temporal memory and worldwide context for improving accuracy and generalization.

The use of Graph Neural Networks (GNNs) is an intriguing area of research. Communication networks are full of graph structures whereby routers, switches, and end devices are connected through several communication links. GNNs can depict all these topological relationships, while LSTM will take care of the temporal aspect of the network, which will help to make better predictions about congestions and packet losses on a larger scale as well as on a multitype network infrastructure.

The role of Digital Twin Networks (DTN) in the management of intelligent networks is increasing. Digital Twin technology enables the creation of virtual replicas of real networks, which facilitates research on network behavior, assessment of different congestion scenarios, and trial of prediction algorithms. Furthermore, combining the model of predicting packet losses based on LSTM with digital twin technology will make it possible to conduct preventive network optimization, identifying failures faster, and performing adaptive traffic engineering.

Lastly, Reinforcement Learning (RL) holds substantial promise for the intelligent management of congestion as well as in the adaptive loss of packets. Unlike supervised approaches of learning, reinforcement learning involves continuous learning about the optimal policies of making decisions through the interaction between the system and its environment. The results of combining LSTM with reinforcement learning could be the development of autonomous congestion control solutions that can adaptively alter their routing, bandwidth, and transmission rates.

To sum up, future research in the field of packet loss prediction should strive for the design of lightweight, easy-to-understand, privacy-preserving, and adaptive deep learning architectures that can work efficiently in various networks. The teamwork of the latest AI methods with LSTM prediction techniques is likely to result in a dramatic improvement in quality, flexibility, and smartness of communication systems of the future.

9. Conclusion

Packet Loss Prediction (PLP) has become a very important field of research that helps to increase performance, reliability, and efficiency of modern networks. The development of the prediction techniques went through several stages starting from classical statistical models like ARIMA and Markovian models to machine learning approaches such as Decision Trees, Support Vector Machines, Random Forest, and XGBoost. While these techniques ensure reasonable performance in case of stable networks, they usually have problems dealing with non-linear nature and time dependency of dynamic network traffic. Deep learning algorithms, especially LSTM, helped to make great improvements in PLP because of their ability to learn long-term dependencies.

As this literature review has shown, LSTM is still considered the best architecture of deep learning used in packet loss prediction because of the memory-based structure, solution to the problem of the vanishing gradient, and better results in working with time series. Compared with machine learning algorithms and classical recurrent neural networks, LSTM demonstrates better prediction accuracy, RMSE and MAE values, and greater precision and recall. In addition, the use of LSTM in 5G/6G networking, Internet of Things, Software-Defined Networking, cloud and edge computing, and intelligent transport systems shows its relevance to current networking problems.

However, there are some drawbacks in addition to the benefits described above. The LSTM neural networks usually require big training sets, lots of computational power, and time for training that is bigger compared to other methods. The black box algorithm also makes it difficult to interpret results, thus increasing difficulties with deploying these models in safety-critical situations.

Accordingly, future research efforts should thus concentrate on the application of Explainable AI, Federated Learning, TinyML, Transformer-LSTM model architectures, Graph Neural Networks, Digital Twin Networks, and Reinforcement Learning in order to address the above-stated drawbacks. With the above-mentioned techniques having a lot of potential to enhance prediction precision, computational efficiency, scalability, privacy, and adaptiveness as well as provide intelligent network management, it is evident that efficient packet loss prediction models will be an essential area for future research. This is especially relevant in the context of the evolution of communication systems to 6G, edge intelligence, and large-scale IoT.

10. References

- [1] J. Yao *et al.*, “A Data-Driven Framework for TCP to Achieve Flexible QoS Control in Mobile Data Networks,” *2023 IEEEACM 31st Int. Symp. Qual. Serv. IWQoS*, vol. null, pp. 1–11, 2023, doi: 10.1109/IWQoS57198.2023.10188765.
- [2] A. Das, W. Kong, R. Sen, and Y. Zhou, “A decoder-only foundation model for time-series forecasting,” 2023, doi: 10.48550/arXiv.2310.10688.
- [3] J. Zhou *et al.*, “A Machine Learning-Based Framework for Dynamic Selection of Congestion Control Algorithms,” *IEEEACM Trans. Netw.*, vol. 31, pp. 1566–1581, 2023, doi: 10.1109/TNET.2022.3220225.
- [4] Z. Xia *et al.*, “A Multi-objective Reinforcement Learning Perspective on Internet Congestion Control,” *2021 IEEEACM 29th Int. Symp. Qual. Serv. IWQoS*, vol. null, pp. 1–10, 2021, doi: 10.1109/IWQoS52092.2021.9521291.
- [5] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, “A Time Series is Worth 64 Words: Long-term Forecasting with Transformers,” *ArXiv*, vol. abs/2211.14730, p. null, 2022, doi: 10.48550/arXiv.2211.14730.
- [6] Z. Du, J. Zheng, H. Yu, L. Kong, and G. Chen, “A unified congestion control framework for diverse application preferences and network conditions,” *Proc. 17th Int. Conf. Emerg. Netw. Exp. Technol.*, vol. null, p. null, 2021, doi: 10.1145/3485983.3494840.
- [7] J. Yao *et al.*, “A Unified Framework for DRL-Based Congestion Control to Optimize QoS Over Mobile Networks,” *2025 IEEEACM 33rd Int. Symp. Qual. Serv. IWQoS*, vol. null, pp. 1–10, 2025, doi: 10.1109/IWQoS65803.2025.11143480.
- [8] H. Tian *et al.*, “Achieving Fairness Generalizability for Learning-based Congestion Control with Jury,” *Proc. Twent. Eur. Conf. Comput. Syst.*, vol. null, p. null, 2025, doi: 10.1145/3689031.3696065.
- [9] Q. Hong, F. Meng, and F. Maldonado, “Advancing Long-Term Multi-Energy Load Forecasting with Patchformer: A Patch and Transformer-Based Approach,” *ArXiv*, vol. abs/2404.10458, p. null, 2024, doi: 10.48550/arXiv.2404.10458.
- [10] S. Rahmatinia and S. Hosseini-Seno, “An Attention-Based Deep Learning Model for Multi-Horizon Prediction of Load, Price, and Wind Power Generation in Smart Grids,” *2025 29th Int. Electr. Power Distrib. Conf. EPDC*, vol. null, pp. 1–9, 2025, doi: 10.1109/EPDC67173.2025.11278281.
- [11] F. Han, Y. Wang, Y. Zhang, Q. Li, D. Zhao, and Y. Jiang, “Anole: A Pragmatic Blend of Classic and Learning-Based Algorithms in Congestion Control,” *IEEE Trans. Comput.*, vol. 74, pp. 2501–2514, 2025, doi: 10.1109/TC.2025.3566872.
- [12] J. Zhou *et al.*, “Antelope: A Framework for Dynamic Selection of Congestion Control Algorithms,” *2021 IEEE 29th Int. Conf. Netw. Protoc. ICNP*, vol. null, pp. 1–11, 2021, doi: 10.1109/ICNP52444.2021.9651912.
- [13] J. Lu, X. Han, and S. Yang, “ARM: Refining Multivariate Forecasting with Adaptive Temporal-Contextual Learning,” *ArXiv*, vol. abs/2310.09488, p. null, 2023, doi: 10.48550/arXiv.2310.09488.
- [14] X. Liao, H. Tian, C. Zeng, X. Wan, and K. Chen, “Astraea: Towards Fair and Efficient Learning-based Congestion Control,” *Proc. Ninet. Eur. Conf. Comput. Syst.*, vol. null, p. null, 2024, doi: 10.1145/3627703.3650069.
- [15] P. Niu, T. Zhou, X. Wang, L. Sun, and R. Jin, “Attention as Robust Representation for Time Series Forecasting,” *ArXiv*, vol. abs/2402.05370, p. null, 2024, doi: 10.48550/arXiv.2402.05370.
- [16] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting,” 2021, [Online]. Available: <https://www.semanticscholar.org/paper/fc46ccb83dc121c33de7ab6bdeab7d970780b2f>
- [17] B. Wu, T. Li, C. Luo, C. Ouyang, X. Du, and F.-J. Wang, “AutoPlex: inter-session multiplexing congestion control for large-scale live video services,” *Proc. ACM SIGCOMM Workshop Netw.-Appl. Integr.*, vol. null, p. null, 2022, doi: 10.1145/3538401.3546596.
- [18] H. Xu, L. Li, L. Wang, and F. Song, “Congestion Control as a Service: Towards Low Latency Mobile Uploading,” *2024 IEEEACM 32nd Int. Symp. Qual. Serv. IWQoS*, vol. null, pp. 1–10, 2024, doi: 10.1109/IWQoS61813.2024.10682903.
- [19] W. Wei, H. Gu, and B. Li, “Congestion Control: A Renaissance with Machine Learning,” *IEEE Netw.*, vol. 35, pp. 262–269, 2021, doi: 10.1109/MNET.011.2000603.
- [20] W. Jiang *et al.*, “Copa+: Analysis and Improvement of the Delay-based Congestion Control Algorithm Copa,” *IEEE INFOCOM 2022 - IEEE Conf. Comput. Commun.*, vol. null, pp. 920–929, 2022, doi: 10.1109/INFOCOM48880.2022.9796913.

- [21]Y. Zhang and J. Yan, "Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting," 2023, [Online]. Available: <https://www.semanticscholar.org/paper/fb45d31cc89207aec392dbac8908cc24db2df871>
- [22]K. Mishra, N. Jain, and S. Bhutani, "Delay Driven Machine Learning Scheme for TCP Data Flow Performance Enhancement," *Adv. Nonlinear Var. Inequalities*, vol. null, p. null, 2025, doi: 10.52783/anvi.v28.4488.
- [23]S. Emara, B. Li, and Y. Chen, "Eagle: Refining Congestion Control by Learning from the Experts," *IEEE INFOCOM 2020 - IEEE Conf. Comput. Commun.*, vol. null, pp. 676–685, 2020, doi: 10.1109/infocom41043.2020.9155250.
- [24]Y. Fan, B. Hou, P. Xie, L. Liu, and H. Ma, "EchoCC: Refining Learning-Based Congestion Control With WordBook," *IEEE Trans. Netw.*, vol. 33, pp. 2553–2568, 2025, doi: 10.1109/TON.2025.3566495.
- [25]H. Tian, X. Liao, C. Zeng, D. Sun, J. Zhang, and K. Chen, "Efficient DRL-Based Congestion Control With Ultra-Low Overhead," *IEEEACM Trans. Netw.*, vol. 32, pp. 1888–1903, 2024, doi: 10.1109/TNET.2023.3330737.
- [26]J. Xing, X. Xu, C. Liu, and D. Hu, "Electricity Load Forecasting Based on Patch Channel-Mixing Transformer Model," *2024 6th Int. Conf. Electron. Eng. Inform. EEI*, vol. null, pp. 450–454, 2024, doi: 10.1109/EEI63073.2024.10696901.
- [27]S. Rahmatinia and S. A. H. Seno, "Enhancing NTMA with Simultaneous Multi-QoS Parameter Prediction using Transformer-Based Deep Learning," *2024 11th Int. Symp. Telecommun. IST*, vol. null, pp. 53–58, 2024, doi: 10.1109/IST64061.2024.10843472.
- [28]T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting," 2022, [Online]. Available: <https://www.semanticscholar.org/paper/563bac1c5cdd5096e9dbf8d4f3d5b3c4f7284e06>
- [29]C. Li, B. Xiao, and Q. Yuan, "FPN-fusion: Enhanced Linear Complexity Time Series Forecasting Model," *ArXiv*, vol. abs/2406.06603, p. null, 2024, doi: 10.48550/arXiv.2406.06603.
- [30]Y. Lan and E. Oermann, "Gateformer: Advancing Multivariate Time Series Forecasting through Temporal and Variate-Wise Attention with Gated Representations," *ArXiv*, vol. abs/2505.00307, p. null, 2025, doi: 10.48550/arXiv.2505.00307.
- [31]W. Yang, Y. Liu, C. Tian, J. Jiang, and L. Guo, "Gemini: Divide-and-Conquer for Practical Learning-Based Internet Congestion Control," *IEEE INFOCOM 2023 - IEEE Conf. Comput. Commun.*, vol. null, pp. 1–10, 2023, doi: 10.1109/INFOCOM53939.2023.10229069.
- [32]J. Zhao, J. Zhang, D. Niu, T. Zhu, C. Zhang, and J. Yang, "GLoSyformer: Global-Local Synergistic Enhanced Transformer for Multivariate Time Series Forecasting," *2025 Int. Jt. Conf. Neural Netw. IJCNN*, vol. null, pp. 1–8, 2025, doi: 10.1109/IJCNN64981.2025.11227793.
- [33]T. Li, C. Wang, and F. Gao, "Hadaformer: Efficient Transformers with Hadamard Attention for Long-Term Series Forecasting," *2025 10th Int. Semin. Comput. Technol. Mech. Electr. Eng. ISCME*, vol. null, pp. 220–226, 2025, doi: 10.1109/ISCME66795.2025.11281479.
- [34]N. Khanal, H. Li, and Y. Zhuang, "InfoST: An Information-Theoretic Framework for Evaluating Spatio-Temporal Prediction," *2025 IEEE Int. Conf. Data Min. Workshop ICDMW*, vol. null, pp. 248–254, 2025, doi: 10.1109/ICDMW69685.2025.00035.
- [35]H. Cao, Z. Huang, T. Yao, J. Wang, H. He, and Y. Wang, "InParformer: Evolutionary Decomposition Transformers with Interactive Parallel Attention for Long-Term Time Series Forecasting," 2023, doi: 10.1609/aaai.v37i6.25845.
- [36]S. Abbasloo, "Internet Congestion Control Benchmarking," *ArXiv*, vol. abs/2307.10054, p. null, 2023, doi: 10.48550/arXiv.2307.10054.
- [37]Y. Liu *et al.*, "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting," *ArXiv*, vol. abs/2310.06625, p. null, 2023, doi: 10.48550/arXiv.2310.06625.
- [38]H. Xu, L. Wang, and F. Song, "KerDqn: Deep Reinforcement Learning Enhanced Congestion Control in Kernel," *ICC 2024 - IEEE Int. Conf. Commun.*, vol. null, pp. 479–484, 2024, doi: 10.1109/ICC51166.2024.10622900.
- [39]Z. Du, J. Zheng, H. Yu, H. Zhang, and G. Chen, "Libra: A Congestion Control Framework for Diverse Application Preferences and Network Conditions," *IEEE Trans. Netw.*, vol. 33, pp. 510–525, 2025, doi: 10.1109/TNET.2024.3492096.

- [40]X. Huang, J. Li, C. Yang, R. Liao, and L. Jiang, "LWaveNet: A Time Series Forecasting Model with Learnable Wavelet Transform and Differential Attention," *2025 6th Int. Conf. Mach. Learn. Comput. Appl. ICMLCA*, vol. null, pp. 1–6, 2025, doi: 10.1109/ICMLCA66850.2025.11336703.
- [41]X. Dong, L. Jiang, Z. Xu, H. Han, C. Li, and Z. Wang, "MAI-Unet: A Multi-Scale Attention Interactive Network for Multivariate Long-Term Time Series Forecasting," *2024 IEEE Smart World Congr. SWC*, vol. null, pp. 632–639, 2024, doi: 10.1109/SWC62898.2024.00116.
- [42]L. Jia, C. Zhou, T. Huang, C. Li, and L. Sun, "Meet Challenges of RTT Jitter, A Hybrid Internet Congestion Control Algorithm," *Proc. ACM Web Conf. 2024*, vol. null, p. null, 2024, doi: 10.1145/3589334.3645338.
- [43]H. Bai *et al.*, "MetaCon: Revitalizing Internet Congestion Control with Meta-Reinforcement Learning," *ICASSP 2025 - 2025 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*, vol. null, pp. 1–5, 2025, doi: 10.1109/ICASSP49660.2025.10888362.
- [44]C. Klotergens, T. Dervede, L. Schmidt-Thieme, and V. K. Yalavarthi, "Mixing It Up: Exploring Mixer Networks for Irregular Multivariate Time Series Forecasting," 2025, [Online]. Available: <https://www.semanticscholar.org/paper/42a5be3c42554e4b3dd22f89df782a38d9cc4475>
- [45]D. Luo and X. Wang, "ModernTCN: A Modern Pure Convolution Structure for General Time Series Analysis," 2024, [Online]. Available: <https://www.semanticscholar.org/paper/4ed9e4401a0388fcc0950a0449dcef00805f5f2c>
- [46]Q. Liu, P. Yang, F. Lyu, N. Zhang, and L. Yu, "Multi-Objective Network Congestion Control via Constrained Reinforcement Learning," *2021 IEEE Glob. Commun. Conf. GLOBECOM*, vol. null, pp. 1–6, 2021, doi: 10.1109/GLOBECOM46510.2021.9685180.
- [47]Z. Zhong *et al.*, "Muses: Enabling Lightweight Learning-Based Congestion Control for Mobile Devices," *IEEE INFOCOM 2022 - IEEE Conf. Comput. Commun.*, vol. null, pp. 2208–2217, 2022, doi: 10.1109/INFOCOM48880.2022.9796880.
- [48]Y. Liu, H. Wu, J. Wang, and M. Long, "Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting," *Adv. Neural Inf. Process. Syst.* 35, vol. null, p. null, 2022, doi: 10.52202/068431-0718.
- [49]W. Yue *et al.*, "OLinear: A Linear Model for Time Series Forecasting in Orthogonally Transformed Domain," *ArXiv*, vol. abs/2505.08550, p. null, 2025, doi: 10.48550/arXiv.2505.08550.
- [50]T. Zhou, P. Niu, X. Wang, L. Sun, and R. Jin, "One Fits All: Power General Time Series Analysis by Pretrained LM," *Adv. Neural Inf. Process. Syst.* 36, vol. null, p. null, 2023, doi: 10.52202/075280-1877.
- [51]Z. Xiao, A. Harwood, and M. A. Rodriguez, "Optimal Rate Control for Latency-constrained High Throughput Big Data Applications," *2022 IEEE Int. Conf. Big Data Big Data*, vol. null, pp. 234–241, 2022, doi: 10.1109/BigData55660.2022.10021076.
- [52]A. Sacco, M. Flocco, F. Esposito, and G. Marchetto, "Partially Oblivious Congestion Control for the Internet via Reinforcement Learning," *IEEE Trans. Netw. Serv. Manag.*, vol. 20, pp. 1644–1659, 2023, doi: 10.1109/TNSM.2022.3215669.
- [53]D. Wei, J. Zhang, X. Zhang, and C. Huang, "Plume: Lightweight and generalized congestion control with deep reinforcement learning," *China Commun.*, vol. 19, pp. 101–117, 2022, doi: 10.23919/JCC.2022.00.019.
- [54]T. Zhou, P. Niu, X. Wang, L. Sun, and R. Jin, "Power Time Series Forecasting by Pretrained LM," *ArXiv*, vol. abs/2302.11939, p. null, 2023, doi: 10.48550/arXiv.2302.11939.
- [55]Y. Gu, K. Zhang, Q. Liu, R. Yu, X. Lin, and X. Sun, "ProCC: Programmatic Reinforcement Learning for Efficient and Transparent TCP Congestion Control," *Proc. Eighteenth ACM Int. Conf. Web Search Data Min.*, vol. null, p. null, 2025, doi: 10.1145/3701551.3703585.
- [56]L. Zhang, K. Zhu, Y. Cui, and Y. Jiang, "QoS-aware congestion control with online learning," *China Commun.*, vol. 20, pp. 168–178, 2023, doi: 10.23919/JCC.2023.02.010.
- [57]M. Welzl, S. Islam, and M. V. Stephanides, "Real-Time TCP Packet Loss Prediction Using Machine Learning," *IEEE Access*, vol. 12, pp. 159622–159634, 2024, doi: 10.1109/ACCESS.2024.3488511.
- [58]C. Wang *et al.*, "Rethinking the Power of Timestamps for Robust Time Series Forecasting: A Global-Local Fusion Perspective," *ArXiv*, vol. abs/2409.18696, p. null, 2024, doi: 10.48550/arXiv.2409.18696.
- [59]Z. Li, S. Qi, Y. Li, and Z. Xu, "Revisiting Long-term Time Series Forecasting: An Investigation on Linear Mapping," *ArXiv*, vol. abs/2305.10721, p. null, 2023, doi: 10.48550/arXiv.2305.10721.
- [60]Z. Xia *et al.*, "RLCC: Practical Learning-based Congestion Control for the Internet," *2021 Int. Jt. Conf. Neural Netw. IJCNN*, vol. null, pp. 1–8, 2021, doi: 10.1109/IJCNN52387.2021.9533754.

- [61]Y. Zhang, Z. Jia, Q. Li, X. Zhang, and Z. Guo, "RTCC: Enable End-to-end Sub-RTT Congestion Control for Next-generation Network," *2024 21st Annu. IEEE Int. Conf. Sens. Commun. Netw. SECON*, vol. null, pp. 19–27, 2024, doi: 10.1109/SECON64284.2024.10934858.
- [62]Z. Zhang, L. Meng, and Y. Gu, "SageFormer: Series-Aware Framework for Long-Term Multivariate Time-Series Forecasting," *IEEE Internet Things J.*, vol. 11, pp. 18435–18448, 2023, doi: 10.1109/IIOT.2024.3363451.
- [63]Z.-Q. Dai, J. Li, Y. Cao, and Y. Zhang, "SALSTM: segmented self-attention long short-term memory for long-term forecasting," *J. Supercomput.*, vol. 81, p. null, 2024, doi: 10.1007/s11227-024-06493-z.
- [64]Z. Yang and T. Gonsalves, "SegmentedCrossformer—A Novel and Enhanced Cross-Time and Cross-Dimensional Transformer for Multivariate Time Series Forecasting," *Forecasting*, vol. null, p. null, 2025, doi: 10.3390/forecast7030041.
- [65]R. Akkineni and C. Uddagiri, "Sig-Patchformer: A Path Signature Based Transformer for Efficient Time Series Forecasting," *2025 IEEE Int. Conf. Big Data BigData*, vol. null, pp. 796–801, 2025, doi: 10.1109/BigData66926.2025.11401048.
- [66]W. Zhang, H. Wang, and F. Zhang, "Skip-Timeformer: Skip-Time Interaction Transformer for Long Sequence Time-Series Forecasting," 2024, doi: 10.24963/ijcai.2024/608.
- [67]Z. Jia, C. Chen, L. Guan, and J. Woodward, "SLO-Targeted Congestion Control with Deep Reinforcement Learning," *2025 IEEE Symp. Comput. Commun. ISCC*, vol. null, pp. 1–6, 2025, doi: 10.1109/ISCC65549.2025.11325988.
- [68]H. Tian, X. Liao, C. Zeng, J. Zhang, and K. Chen, "Spine: an efficient DRL-based congestion control with ultra-low overhead," *Proc. 18th Int. Conf. Emerg. Netw. Exp. Technol.*, vol. null, p. null, 2022, doi: 10.1145/3555050.3569125.
- [69]C. Garcia and G. Remy, "Statistical benchmarking of transformer models in low signal-to-noise time-series forecasting," *ArXiv*, vol. abs/2602.09869, p. null, 2026, doi: 10.48550/arXiv.2602.09869.
- [70]J. Chen, B. Fan, and Q. Zhang, "T-Mamba: A Hybrid Mamba-Transformer Framework for Stock Price Prediction," *Proc. 2025 6th Int. Conf. Comput. Inf. Big Data Appl.*, vol. null, p. null, 2025, doi: 10.1145/3746709.3746715.
- [71]W. Li, S. Gao, X. Li, Y. Xu, and S. Lu, "TCP-NeuRoc: Neural Adaptive TCP Congestion Control With Online Changepoint Detection," *IEEE J. Sel. Areas Commun.*, vol. 39, pp. 2461–2475, 2021, doi: 10.1109/JSAC.2021.3087247.
- [72]L. Sun, X. Gao, K. Xia, X. Hu, and Y. Feng, "TCTformer: Long-term forecasting with dual attention transformers," *ICASSP 2025 - 2025 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*, vol. null, pp. 1–5, 2025, doi: 10.1109/ICASSP49660.2025.10888630.
- [73]H. Li *et al.*, "Teaching to Fish Rather Than Giving a Fish: The Concentrator Method of Teaching Classic Congestion Control With Learning-Based module," *IEEE Trans. Mob. Comput.*, vol. 24, pp. 10042–10054, 2025, doi: 10.1109/TMC.2025.3567582.
- [74]C. Ying and J. Lu, "TFEformer: Temporal Feature Enhanced Transformer for Multivariate Time Series Forecasting," *IEEE Access*, vol. 12, pp. 153694–153708, 2024, doi: 10.1109/ACCESS.2024.3480953.
- [75]S. Wang *et al.*, "TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting," *ArXiv*, vol. abs/2405.14616, p. null, 2024, doi: 10.48550/arXiv.2405.14616.
- [76]S. Wang *et al.*, "TimeMixer++: A General Time Series Pattern Machine for Universal Predictive Analysis," *ArXiv*, vol. abs/2410.16032, p. null, 2024, doi: 10.48550/arXiv.2410.16032.
- [77]H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis," *ArXiv*, vol. abs/2210.02186, p. null, 2022, doi: 10.48550/arXiv.2210.02186.
- [78]H. Tian *et al.*, "Towards Fair and Efficient Congestion Control Through Multi-Agent Deep Reinforcement Learning," *IEEE Trans. Netw.*, vol. 34, pp. 2966–2980, 2026, doi: 10.1109/TON.2026.3653773.