SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

SCALABLE AND EFFICIENT TRAFFIC PREDICTION IN INTERNET OF THINGS (IOT) NETWORKS USING DEEP LEARNING MODELS

Fawzan Mushtaq[*1], Muhammad Junaid Arshad[2]

[*1]*Institute of Data Science, University of Engineering and Technology (UET), Lahore, Pakistan*
[2]*Department of Computer Science, University of Engineering and Technology (UET), Lahore, Pakistan*

[*1]fawzanmushtaq477@gmail.com, [2]junaidarshad@uet.edu.pk

DOI: https://doi.org/10.5281/zenodo.20791919

Keywords
Internet of Things, IoT, Machine Learning, Deep Learning, IoT Traffic Prediction, Network Traffic Analysis, Time Series Forecasting, Quality of Service, Network Optimization, Data Analysis, Smart Network

Article History
Received: 24 April 2026
Accepted: 03 June 2026
Published: 22 June 2026

Corresponding Author: *
Fawzan Mushtaq

Abstract
As the Internet of Things (IoT) is being developed, Internet traffic has been steadily growing, making it difficult to predict and manage traffic. The authors propose a novel scalable traffic prediction model for IoT networks, which is based on deep learning (DL). It particularly focused on the use of advanced DL techniques such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks and hybrid solutions that are capable of dealing with the complex and non-linear nature of the IoT traffic. The aim of the suggested models is to capture the temporal and spatial dependency of traffic data, to increase the accuracy and robustness of prediction models. We also introduce an innovative approach that combines the trend and residual parts of the traffic data to achieve more accurate forecasting of the traffic on different time scales. Furthermore, the paper examines the potential difficulties in implementing the model in real time and handling vast data sets, and suggests potential enhancements to increase the model's efficiency. The prediction accuracy obtained with the real-world IoT traffic datasets used in the experiments is significantly higher than traditional models, as the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) have been reduced. This work is crucial to the development of scalable, reliable and efficient traffic prediction models for effective traffic management, congestion control and resource allocation for next generation IoT systems.

I. INTRODUCTION

One of the most promising and influential innovations of our time is the Internet of Things (IoT), which is a network of interconnected devices and systems across a variety of industries, such as healthcare, manufacturing, transportation, agriculture and smart cities. The network continues to grow and generate vast amounts of data that's both a blessing and a challenge for network management. The number of IoT devices is increasing at an exponential rate, and soon data management and efficient network performance will be a priority for service providers and end users. Precise prediction of network traffic is one of the most important challenges in this context, as it is crucial for optimizing the allocation of resources, avoiding network congestion, enhancing Quality of Service (QoS), and guaranteeing network security.

The statistical methods like Autoregressive Integrated Moving Average (ARIMA) and Markov chains, which are commonly used for network traffic prediction, have some drawbacks in the non-linear and non-stationary nature of IoT traffic in the highly complex scenario. Real-world IoT networks are highly complex and have a temporal and spatial correlation, which is not

fully captured by these models – particularly in the face of growing data volumes and the complexity of such networks. For this reason, the use of advanced machine learning (ML) and deep learning (DL) models that can process such IoT traffic has been growing in popularity as it is dynamic, high dimensional and heterogeneous.

The paper suggests that deep learning models like Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and hybrid models can be used to improve the scalability, accuracy, and robustness of traffic prediction in IoT networks. These models are particularly useful for learning from large amounts of data, and can identify patterns and trends that are not captured by traditional models. The models incorporate both temporal and spatial dependency of the data, which helps them provide more accurate traffic forecasts at different time scales – essential features for dynamic and unpredictable IoT traffic.

One novelty in the present research is the use of a new method for traffic data decomposition. The traffic data are decomposed into seasonal, trend and residual components, which are modeled separately via deep learning methods. This decomposition enables the better management of the multiscale temporal dependencies in IoT traffic, and leads to more accurate predictions in short-term and long-term horizons. The use of models such as LSTM and CNN, which are able to capture both the local variations and global trends in the traffic data, contributes to the model's ability to generalize to other IoT applications.

The paper not only enhances the prediction accuracy but also tackles the challenges of real-time deployment and handling vast amounts of data. The growth of IoT networks can lead to large computational requirements for deep learning models. In this context, we explore ways of making these models efficient and effective, as well as the potential of these models for being effective at the edge and in fog networks, where resources are limited. To ensure that IoT traffic prediction models can be deployed without any performance loss in real-world applications, these optimizations are crucial.

The proposed models are evaluated on real-world IoT traffic datasets, and the accuracy of their predictions is compared with the traditional methods and their results demonstrate that the proposed models outperform the traditional ones in terms of prediction accuracy and computational efficiency. The results demonstrate remarkable improvements in key metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), underscoring the effectiveness of the deep learning models in dealing with the complex and evolving nature of IoT traffic. The research helps advance the field of scalable and efficient traffic prediction models, which can be beneficial for managing traffic resources, reducing congestion, and improving network security in IoT networks to meet the increasing demands of this technology.

In this paper, a comprehensive framework for scalable traffic prediction in IoT network is presented, specifically focusing on the application of deep learning models to improve the accuracy of the traffic prediction. The findings provide valuable input for future studies and applications and provide guidance of how to make traffic prediction more efficient, reliable and flexible in the increasingly large-scale IoT context.

As the Internet of Things (IoT) has expanded, its use has shifted toward increasingly complex and dynamic traffic patterns, making it difficult to model with traditional traffic prediction methods. Although deep learning models have achieved significant accuracy in traffic prediction in IoT environments, there is still high computational complexity, insufficient capacity for sudden anomalies, and weak generalization ability across various IoT applications. Scalable, Adaptive, and Efficient models of traffic prediction are clearly required that can perform reliably in real time, across different IoT environments, and in the presence of sudden disruptions to the network. These deficiencies have a significant impact on resource allocation, congestion relief, and the security of IoT networks, making them essential for improving these areas.

II. LITERATURE REVIEW

A study suggested a hybrid model that integrates CNN-LSTM, XGBoost, and LLMs to forecast network traffic in IoT-based 5G/6G networks. The model accounts for temporal patterns, minimises prediction errors and includes semantic trend information. Using real cellular traffic data, the results demonstrated greater accuracy and the ability to deal with noisy or incomplete data. The architecture was also more tolerant of sudden traffic variations and data voids, making it applicable to the future IoT networks. [1]

In a study, the time-series traffic data was transformed into images and a U-Net-based technique was proposed for network traffic prediction. It is based on the U-Net model architecture that has an encoder-decoder structure to capture the global and local traffic patterns. The experimental results indicated that it outperforms CNN-LSTM, CNN-GRU and CoreGAN models particularly in multi-step forecasting. The proposed approach considerably reduced the prediction error and showed good scalability and efficiency in network traffic prediction. [2]

A research suggested a hierarchical deep learning system for classification of IoT and critical IoT traffic, named as Hi-CIoTNet. To enhance the feature extraction and classification performance, the model is based on two-stage process and a RAD-CNN structure. It correctly identified over 99% of the IoT traffic from non-IoT traffic and spotted critical traffic. The results show that it is suitable for real-time and mission critical applications in the IoT like healthcare and industrial automation. [3]

The authors of a study offered an STL decomposition-based and a GRU-based Seq2Seq model with attention mechanism for the traffic prediction in IoT networks. This method decomposes the traffic data into seasonal, trend and residual components, and models each one separately. It can effectively capture multiscale temporal patterns and long-term dependencies by combining frequency-domain analysis and deep learning. Experimental results conducted on various real-world datasets revealed substantial decreases in prediction errors, achieving high accuracy and generalizability properties in different IoT settings. [4]

A CNN-based real-time traffic classification system for resource-limited IoT networks was suggested in a study. The model employs network flow features based on metadata, packet-level data, payload information and transitions between services for identifying IoT devices. Various inference methods were tried to achieve a trade-off between accuracy and computational complexity, and the best method was an all-traffic model. Practical tests with real IoT data demonstrated good results (macro F1-score of 0.90), and also achieved high speed and low memory consumption, which is suitable for deployment at edge nodes, such as Raspberry Pi systems. [5]

In 2025, a study has been presented to predict the IoT network traffic accurately using MRSTGCN. This model integrates graph convolution and temporal embedding to learn spatial and temporal traffic patterns. Dynamic Adjacency matrices can be used to represent the evolving relationship between IoT devices. Experiments on real-world data demonstrated improved performance compared to STGCN, Graph WaveNet and MTGNN. The model was found to be useful for both short term and long term traffic forecasting. [6] A 2025 research paper suggested the traffic prediction framework for Symbiotic IoT environments called MSTJLLM. The model combines historical traffic data, spatial-temporal features, and text prompts to capture complex traffic patterns. It leverages pre-trained large language models for the feature fusion and enhanced prediction accuracy. The PeMS04, PeMS07 and PeMS08 datasets demonstrated a better performance in comparison with traditional and graph-based models. The framework was tested to be effective in both data-rich and data-scarce IoT scenarios. [7]

In 2025, a study was introduced to solve the aforementioned problem by proposing a model, known as KGASTN, which is used to forecast network traffic in the long-term in IoT environment. The model fuses causal knowledge

from network logs and machine learning methods to establish dynamic relationships between spatial locations. It adopts a spatio-temporal graph attention network to learn spatial and temporal traffic relationships. The real-world datasets revealed better prediction accuracy than conventional approaches for both short-term and long-term forecasts. Moreover, the model is also very efficient from a computational point of view, which is crucial for real-time traffic management. [8]

A review paper on traffic flow prediction in urban areas is published in 2025, which compares different deep learning approaches. Compares models like CNN, LSTM, Bi-LSTM on datasets like PeMS. LSTM is good at capturing time patterns, CNN is good at capturing spatial features, but each has its own limitations. The research also suggests hybrid techniques like Bi-LSTM and PSO for further improvement in accuracy and training efficiency. It suggests that future studies need to concentrate on real-time systems and on improved integration of spatial-temporal data. [9]

A paper in 2025 suggests that LT-GCN can enhance the efficiency of Graph Convolutional Networks (GCN) in large-scale traffic prediction. The model is trained sequentially instead of all layers at once, thus reducing training time, and yet it still achieves a comparable accuracy. It is then fused with GRU to create LTGG to represent both spatial and temporal traffic features. The experiments demonstrate that LT-GCN can greatly reduce the training time and that LTGG is superior to other models such as LSTM and GRU. It is applicable for the real-time and large-scale network traffic forecasting. [10]

A 2025 study is a comparison of statistical, machine learning and deep learning models for predicting cellular network traffic. It compares techniques such as ARIMA, SARIMA, Random Forest, Gradient Boosting, KNN and LSTM based on standard error metrics. The system consists of preprocessing, feature engineering, and a Streamlit dashboard to visualize and predict. Other approaches fail to perform as well as LSTM, which achieves the best results with the lowest errors and highest accuracy. The study points out that deep learning has the following benefit compared to traditional methods: the ability to accurately predict network traffic at high speed. [11]

The study for 2025 forward suggests an innovative framework for adaptive IoT network management using Generative AI called GenTwin. It integrates Priority Pooling to identify the most crucial relationships in IoT and a Twin Adapter which optimises LLaMA-2-8B to create digital twins from knowledge graphs. The system also has what-if analysis capability for dynamic network conditions. The experiments reveal that the proposed approach achieves better relation extraction (+19%) and faster response time (−53%) than traditional approaches. The study notes improved scalability and rapid response, but with some challenges to scale up. [12]

An intelligent load balancing framework for Multimedia IoT (IoMT) is proposed in 2025 study. It uses an LSTM time-series model for predicting the server load based on CPU, memory, bandwidth and disk usage. A fuzzy logic system then categorizes the servers with various load levels and routes traffic based on that. Mininet simulations indicate better load distribution, more efficient use of resources and energy saving. The study says that tests on the large-scale are still required to validate it in real-world settings. [13]

A proposed hybrid of AI for 6G traffic prediction and resource optimization is presented in a 2025 study. It integrates an attention mechanism with Random Forest and GRU to learn and remember both spatial and temporal traffic patterns. Model processes network data using RF-generated features and GRU layers, while attention highlighting key time steps. Results indicate that it is more accurate than LSTM and GRU, with very low prediction errors and high accuracy (R^2 = 0.997). This method enables efficient resource allocation, mitigates congestion, and helps achieve energy-efficient 6G networks. [14]

In 2025, a study suggested a Deep LSTM network with Chi-Square feature selection for traffic congestion prediction in an IoT ad-hoc network. It employs preprocessing methods such as SMOTE and feature selection to enhance data

quality and minimize overfitting. Then the DLSTM model is employed to determine if there is any congestion in the network. The result indicates the model is 98.32% accurate, which is better than ANN, DBN, DNN and RNN models. The research emphasizes on effective congestion detection and secure IoT traffic management, but there exists scope for scalability. [15]

The DRAFT (deep robust adaptive framework for SDN-IoT traffic prediction and management) study for 2025 suggests a deep robust adaptive framework for SDN-IoT traffic prediction and management. It combines denoising methods (EMD and Wavelet Transform), an improved TimesNet model, federated learning, and zero-trust security. The model is more effective in modeling noisy and multi-periodic traffic patterns, and retains data privacy. Experiments with real data reveal that this model achieves a considerable reduction in error and superior performance compared to other models such as LSTM, GRU, and CNN-LSTM. The study shows enhanced accuracy and security, but more testing is required at larger scale. [16]

This study suggests a novel LSTM model for intrusion detection system for highly imbalanced IoT network data based on SMOTE and Categorical Focal Cross-Entropy Loss. It is able to make the rare attacks like U2R, R2L, Heartbleed and Infiltration detectable, while maintaining classes and improving minority learning. LSTM is used to learn temporal structures and SMOTE and focal loss are used to enhance the classification of hard samples. The results of experiments conducted on the datasets KDD99 and CICIDS2017 demonstrate that the model achieves up to 99.33% accuracy, which is better than other deep learning models such as SVM and Random Forest. The study has shown good intrusion detection performance, but it requires further validation in the actual system. [17]

In 2025, a hybrid SDN routing approach, which integrates Q-Learning and Neural Networks, is proposed for adaptive and predictive traffic management. Q-Learning is used to choose optimal paths depending on the network state and an ANN is used to forecast traffic loads in the network to avoid congestion. The model is validated in a Mininet based SDN environment with the help of standard networking tools. The results demonstrate lower latency, greater packet delivery ratio and better throughput than those of conventional routing techniques. The study is an indication of improved QoS in dynamic networks but has not been tested in the real world. [18]

The study for 2025 suggests a hybrid RNN-XGBoost Digital Twin framework for real-time travel time prediction in smart cities. It integrates a RNN, LSTM, and an XGBoost model to capture both temporal and nonlinear traffic patterns from real-world traffic and environmental data. According to the results, the XGBoost model outperforms the other models and the hybrid model shows high R^2 and low RMSE values to improve the accuracy. The system is connected to a 3D Digital Twin to visualize and route intelligently in real-time. The study focuses on the problem of improving the prediction and routing but it does not address the problem of computational complexity. [19]

Recently, a research paper has introduced a lightweight IoT intrusion detection framework for MQTT networks that employs statistical moments difference thresholding (SMDT) for feature selection. It filters out 33 features and extracts 5 key features from the dataset based on statistical measures such as mean, skewness and kurtosis. Attack classification is done using multiple ML models, such as Random Forest, XGBoost, and Decision Tree. The results on the MQTTset demonstrate the accuracy up to 95% in binary classification with a low computational cost. The research focuses on efficient real-time detection, but further testing is needed in the real world. [20]

A new study from 2025 suggests a federated learning based CRNN approach for intrusion detection in IoT and IIoT networks. It is an integration of CNN for spatial feature extraction and LSTM for temporal attack patterns in network traffic. Federated learning allows for decentralized training without compromising data privacy and communication overhead. When applied to the Edge-IIoT dataset, the model achieves 98.93% accuracy, 97% F1-score

with better performance than some ML and DL methods. The advantages of the study are its good scalability and privacy properties, while drawbacks such as class imbalance and training costs exist. [21]

The study for 2025 suggests a hybrid model of anomaly detection system for IoT which uses GANs in conjunction with Gaussian noise and Hurst parameter for better robustness in noisy environment. It preprocesses the data of CICIDS2017 with scaling the features, adding noise and extracting the long-term dependencies from the data by using Hurst exponent. To detect anomalies, a generator-discriminator network based on GAN is used, in which anomalies are detected by reconstruction errors and thresholding. The model has a high accuracy of 99.88% and good robustness in different IoT datasets. The study emphasizes effective real time detection, but performance may drop when the noise is very complex in real world. [22]

A 1D CNN-based intrusion detection system is proposed for IoT networks in a scalable study for 2025. It is based on features from the CIC IoT-DIAD 2024 dataset, which are extracted based on flow, and with standard pre-processing, such as normalization and label encoding. The model is compared with LSTM, RNN and MLP for binary and multi-class attack detection. The results indicate that 1D CNN model performs best with upto 99.53% accuracy, having a good precision, recall and F1 score. The study's contributions focus on efficient real-time detection for IoT security, while yet other issues such as latency and large-scale deployment remain. [23]

An article summarizes a recent survey of deep learning intrusion detection systems (IDS) for IoT networks and compares them to traditional IDS. It includes all models proposed for Network Intrusion Detection and Network Anomaly Detection including CNN, RNN, LSTM, Autoencoders, FNN and DNN. Major preprocessing methods are also discussed and some of the common datasets such as NSL-KDD, UNSW-NB15, and CIC-IDS2017 are evaluated. The results demonstrate that deep learning achieves high detection accuracy, scalability and adaptability, particularly in the case of complex attacks and zero day attacks. But there are still problems such as the deployment in real-time, imbalance between classes, and limited resources. [24]

A recent survey is about ML and DL methods for intrusion and anomaly detection in IoT networks. It includes models such as RF, DT, SVM, KNN, XGBoost, CNN, RNN, LSTM, GAN, Federated Learning, and more across the datasets like UNSW-NB15, IoT-23, Bot-IoT, and CICIoT2023. The study reveals that accuracy of DL and hybrid models is very high with accuracy of more than 99% while the speed and computational cost of ML models are comparatively lower. It states that several important questions such as scalability, imbalance, real-time deployment, etc., are still open research questions. [25]

In 2023, a lightweight CNN (LTP-CNN) for network traffic prediction in dense IoT environments by fog computing was proposed. It is designed to process data close to edge devices, which cuts down latency, congestion, and energy use by operating at the fog layer. It is able to achieve low computational cost, good prediction accuracy and real-time traffic prediction simultaneously. Experiments conducted on 2 sets demonstrate the accuracy of approximately 90%, which is better than that of traditional CNN models. The study points to greater scalability of IoT and 5G networks, but notes that the technology is not yet ready for large-scale and real-time deployment. [26]

In 2023, a research paper was published that suggests an IDS solution based on machine learning to enhance security and detection of attacks in IoT traffic. It uses preprocessing techniques like normalization, feature selection, oversampling, and data cleaning to handle noisy and imbalanced data. Several ML models are evaluated and compared on various datasets like UNSW-NB15, BOT-IoT and ToN-IoT. The outcomes indicate that the models are indeed very accurate, with MLP achieving 99.97% accuracy and KNN achieving 99.94%, which is a great performance but still needs room for improvement in terms of deployment in real-time

systems. [27]

Sr. No.	Paper Title	Authors	Year	Research Problem /Objective	Methodology/ Techniques Used	Dataset/ Tools	Key Findings/ Results	Limitations/ Research Gap	Comments/ Remarks
[1]	Large Language Models (LLMs) for Network Traffic Prediction: A Trend-Aware Hybrid Framework	Yuzhou Chen et al.	2026	To predict noisy/ nonstationary IoT traffic	CNN-LSTM + XGBoost + LLM	Cellular traffic, LLaMA2-7B	RMSE < 15%, MAPE < 10%	High computation cost	Accurate but resource-heavy
[2]	When Time Series Data Become Images: U-Net for Network Traffic Prediction	Jiho Yoon et al.	2026	To improve traffic forecasting	U-Net image-based prediction	Pangyo dataset	Error reduced by up to 80%	Needs real-time optimization	Strong multi-step prediction
[3]	Hi-CIoTNet: A CNN-Based Hierarchical Model for IoT and Critical IoT Traffic Classification	Shahbaz Rashed et al.	2026	IoT traffic classification	RAD-CNN hierarchical model	UNSW-IoTraffic, IoT-23	99.9% classification accuracy	Scalability issues	Effective for critical IoT
[4]	Spectrum-Aided Traffic Decomposition and Deep Learning Method for Network Traffic Prediction in Internet of Things	Jiaqi Gao et al.	2026	To handle multiscale dependencies	STL + GRU Seq2Seq + Attention	AIIA, WIDE, Kaggle	MSE < 97.6%, MAE < 84.7%	High complexity	Accurate but costly
[5]	Real-Time and Trustworthy Classification	Sivanathan et al.	2026	To classify real-time IoT traffic	Lightweight CNN-based classificatio	CNN, TensorFlow,	Macro F1-score 0.90 with	Large-scale heterogeneous deployment	Lightweight and efficient Io

	of IoT Traffic Using Lightweight Deep Learning				n	Raspberry Pi	low resource usage	needs further study	T traffic analysis framework
[6]	Multirepresentation Spatial-Temporal Graph Convolutional Networks for Network Traffic Prediction	Yang Yang et al.	2025	To capture spatial-temporal patterns	MRGCN + Temporal Embedding	Telecom Italia, OPNET	Outperformed baselines	Computationally expensive	Good accuracy, low efficiency
[7]	Multilevel Spatial-Temporal Joint Large Language Model for Traffic Prediction in Symbiotic IoT	Zhengwei Xu et al.	2025	To integrate multimodal features	Multilevel embeddings + LLM	PeMS04/07/08	Better MAE, RMSE, MAPE	Sensitive to sudden shifts	Effective feature fusion
[8]	Long-Term Traffic Flow Prediction: A Knowledge-Driven Graph Attention Spatio-Temporal Network	Chenxu Li et al.	2025	To improve long-term prediction	KGASTN + Graph Attention	Milan, Sus datasets	Higher accuracy and efficiency	High-dimensional data issue	Strong long-term forecasting
[9]	Intelligent Traffic Flow Prediction Using Deep Learning Techniques: A Comparative Study	Sayed A. Sayed et al.	2025	To compare DL traffic models	CNN, LSTM, Bi-LSTM, PSO	PeMS dataset	Hybrid models performed best	Real-time gaps remain	Hybrid methods preferred
[10]	A Network Traffic Prediction	Yulian Li et al.	2025	To reduce GCN	LT-GCN	Abilene, GÉANT	Faster training,	Real-time scaling	Efficient G

	Model Based on Layered Training Graph Convolutional Network			training time	+ GRU		good accuracy	issues	CN training
[11]	Time Series Prediction System For Cellular Network Traffic Data Using Long Short-Term Memory, Machine Learning, and Statistics Model	Helmi Maulu et di al.	2025	To improve cellular network traffic prediction	ARIMA, SARIMA, RF, GB, KNN, LSTM	Cellular traffic data, Streamlit	LSTM achieved the best performance (R^2 = 0.9942)	Real-time prediction and large-scale deployment are not addressed	Comprehensive comparison of forecasting models
[12]	GenTwin: Generative AI-Powered Digital Twinning for Adaptive Management in IoT Networks	Kubra Duran et al.	2025	To improve adaptive management in dynamic IoT networks	GenTwin framework with Priority Pooling + Twin Adapter + LLM-based Digital Twin	AnyLogic, Python, PyTorch, Neo4j, LLaMA-2-8B	19% more relation extraction and 53% lower response time	Large-scale deployment and long-term adaptability challenges	Efficient and intelligent adaptive IoT management
[13]	Optimizing Server Load Distribution in Multimedia IoT Environments through LSTM-Based Predictive Algorithms	Somaye Imapour et al.	2025	To improve server load balancing and resource utilization in multimedia IoT	LSTM + Fuzzy Logic	Mininet, Floodlight Controller, OpenFlow, LSTM, Fuzzy System	Better balancing resource load and utilization	Real-world scalability not tested	Intelligent SDN-based IoT load management

				environments					

[14]	Hybrid Model for 6G Network Traffic Prediction and Wireless Resource Optimization	M. A. Oukebdane et al.	2025	To predict traffic and optimize resources	Hybrid RF + GRU + Attention	Python, scikit-learn, GRU, Random Forest	RMSE 0.00488, MAE 0.00340, MAPE 0.46%, R^2 0.997; reduced congestion 40–60%	Small dataset; edge deployment not tested	Combines ensemble & deep learning for 6G networks
[15]	Deep LSTM and Chi-Square Based Feature Selection Model for Traffic Congestion Prediction in Ad-Hoc Network	K. Sangeetha et al.	2025	To predict traffic congestion in ad-hoc networks	DLSTM + Chi-Square + SMOTE	MATLAB	98.32% accuracy, 1.9% error	Real-world not scalability tested	Improves IoT traffic management in VANETs
[16]	Rethinking the Power of Multi-Domain Features for SDN-IoT Network Traffic Prediction: A Intra-and Inter-Period Perspective	Yu Ya et al. n	2025	To improve SDN-IoT traffic prediction	DRAFT wit h EMD + WT + TimesNet + Federated Learning	Python, TimesNet	Lower MAE, RMSE, MAPE, and improved prediction accuracy	Large-scale deployment needs further validation	Enhances secure and adaptive SDN-IoT traffic prediction
[17]	Multi-Class Network Intrusion Detection with Class Imbalance via LSTM & SMOTE	M. W. Nawaz et al.	2025	To improve intrusion detection for imbalanced traffic data	LSTM + SMOTE + CFCL	KDD99, CICIDS2017	98.83% and 99.33% accuracy	Some minority classes showed lower precision	Enhances rare attack detection using deep learning
[18]	Adaptive and Predictive Routing in SDN Using Q-Learning and	N. Shukla et al.	2025	To improve adaptive SDN routing	Q-Learning + ANN	Mininet, Ryu, Iperf	Reduced latency, improved PDR, and throughput	Tested only in an emulated environment	Hybrid AI framework for SDN traff

	Neural Networks								ic management
[19]	RNN-XGBoost Integration into Digital Twins for Predicted Travel Time Geosimulation in Emergency Route Finding	Zehra Rezaei et al.	2025	To improve travel time prediction in smart cities	Hybrid RNN + XGBoost + Digital Twin	TensorFlow, Keras, PostgreSQL	RMSE 0.033255, R^2 0.95	Large-scale deployment complexity	Supports intelligent routing and traffic prediction
[20]	Resource-Efficient Traffic Classification for MQTT-IoT Security Attacks	E. Tuyishime et al.	2025	To improve the MQTT-IoT intrusion detection efficiency	SMDT + ML classification	MQTTset, RF, DT, XGBoost	95% binary accuracy, reduced features from 33 to 5	Real-world not deployment tested	Lightweight IDS for resource-constrained IoT networks
[21]	Federated Learning-based Hybrid CRNN for Multi-Class Intrusion Detection in IoT Networks	P. Selvam et al.	2025	To improve IoT intrusion detection accuracy & privacy	CNN + LSTM + Federated Learning	Edge-IIoT, TensorFlow, FL	98.93% accuracy, 97% F1-score	Class imbalance and high training cost	Privacy-aware hybrid IDS for IoT/IIoT networks
[22]	Combining Generative Adversarial Networks (GANs) with Gaussian Noise for Anomaly Detection in IoT Traffic	R. Morshedi et al.	2025	IoT anomaly detection under noisy conditions	GAN + Gaussian noise + Hurst parameter	Python, CICIDS2017	99.88% accuracy & recall; robust to noise; detects DoS/DDoS/Port Scan	Sensitive to more complex real-world noise/attacks	Hybrid noise-aware GAN for IoT intrusion detection
[23]	Deep Learning-Based Intrusion Detection for IoT Networks: A Scalable and Efficient Approach	Md. Alamgir Hossain et al.	2025	To improve real-time IoT intrusion detection	1D CNN + LSTM + RNN + MLP	TensorFlow, Keras, CIC IoT-DIAD 2024	1D CNN: 99.53% anomaly, 99.12% multiclass	Latency, interpretability, and real-time scaling	Scalable, efficient IDS for IoT
[24]	A Comprehensive	Al-Haija	2025	Review DL-based IoT	CNN, LSTM, AE,	NSL-KDD, UNSW-	DL impro	Computational cost and	Comprehensive survey of

	Survey on DL-Based IDS in IoT	& Droos		IDS techniques	DNN survey	NB15, BoT-IoT	ves IDS accuracy and adaptability	real-time deployment issues	DL-based IoT IDS
[25]	ML and DL Techniques for IoT Network Anomaly Detection	A. Alghamdi et al.	2024	To review IoT anomaly detection techniques	ML/DL-based IDS survey	RF, DT, CNN, LSTM, GAN	Many models achieved >99% accuracy	Scalability and zero-day attacks remain challenges	Comprehensive survey of IoT anomaly detection
[26]	Lightweight Deep Learning-Based Model for Traffic Prediction in Fog-Enabled Dense Deployed IoT Networks	Abdel-Hamied A. Ateya et al.	2023	To reduce congestion and latency	Lightweight CNN at fog layer	Dataset-I, UNI2	~90% accuracy	Large-scale deployment issue	Resource-efficient model
[27]	Toward Improved Machine Learning-Based Intrusion Detection for Internet of Things Traffic	M. A. Ferrag et al.	2023	To improve IDS performance in IoT traffic	ML-based IDS with preprocessing & feature selection	DT, RF, KNN, SVM, MLP	MLP 99.97 %, KNN 99.94 % accuracy	Real-time IoT deployment challenges	Improves IoT intrusion detection using ML techniques

III. LIMITATIONS

The models proposed in this work are deep learning-based traffic prediction models that have shown to be very accurate and efficient in IoT networks, however, several limitations must be taken into account. First, deep learning models are often complex and challenging to compute, particularly for large-scale datasets from IoT systems. While these models have been optimized for real-time deployment, the processing complexity needed for training and inference can be significant, especially with high-dimensional traffic data. This may make it difficult to use these models in resource-limited applications, like edge devices or IoT gateways.

Second, the effectiveness of the models heavily relies on the quality and quantity of the available data. In real-world IoT settings, traffic data may be sparse, noisy, and incomplete, potentially impacting the model's capacity to make accurate predictions for new scenarios. The proposed approach relies on a certain degree of consistency in traffic, but IoT networks are dynamic and face fluctuations due to unforeseen events that can be not well captured in the training data, such as cyberattacks or network failures.

Additionally, while the model's ability to decompose traffic data into seasonal, trend, and residual components improves prediction accuracy, this approach might not fully capture

highly irregular or abrupt changes in traffic patterns, such as those caused by DoS (Denial-of-Service) attacks or sudden traffic surges. These unique instances need to be investigated further for anomaly detection and strong forecasting models that can cope with such disruptions.

In addition, the models proposed in this paper are mainly assessed with a small number of real-world datasets, which might not cover the variety of IoT applications. The challenge is to extend these models to other IoT applications, especially those that have specific traffic patterns (such as in healthcare IoT networks, or autonomous vehicle systems).

Lastly, the models modeled in this work assume that the IoT networks of interest have some traffic patterns. The traffic patterns could, however, be different than those captured by the existing models, due to the fast-changing nature of IoT technologies and the introduction of new devices and services. These models should be adapted to support new IoT technologies and network behaviors in the future.

IV. RESEARCH GAP

Although deep-learning models have been developed to predict traffic, there are still some research gaps that should be addressed in the context of IoT networks. One of the big problems is real-time scalability. Although the current models can be very accurate, they are not readily applicable to real time applications, especially in large scale IoT application scenarios where there is a large number of data. The target future research is to build lightweight models which do not impact on the prediction accuracy of the deep learning techniques, whilst at the same time the model should be efficient enough to be deployed in real-time on resource constrained devices such as IoT gateways and edge devices.

Another important concern is the capacity to deal with the unexpected and unpredictable increases in network traffic, including those that may occur during a cyber attack or network failure. Current models work well for regular traffic, but are not applicable for abnormal traffic flows that have patterns very different from regular traffic. In the future, it is important to build integrated prediction models that integrate anomaly detection and traffic prediction to enable IoT networks to respond to unexpected events and ensure their reliable operation in adverse scenarios.

Moreover, models are yet to be generalized to a wide range of IoT applications. The existing models have been tested on the specific datasets and may not cover the diversity of traffic types for a specific IoT environment. However, these models have not been widely tested in other IoT applications such as industrial IoT, healthcare systems, and smart cities, and they need to be adapted and tested to meet the specific requirements of these sectors.

In addition, the current models are based on static traffic scenarios, which are not a characteristic of IoT networks and can change drastically over time as new devices and services are added. Research needs would include adaptive models that learn and evolve to deal with the dynamic nature of IoT traffic and retain their prediction accuracy as the network evolves in the long-term.

Lastly, data quality and sparsity and noise in real world IoT traffic data is still a challenge to overcome. The capability of handling incomplete or noisy data, as well as more effective data preprocessing techniques, are essential for reliable and robust deep learning prediction models that can be effectively used in IoT applications.

V. RESEARCH OBJECTIVES

The following research objectives are suggested in the light of the gaps identified:

- To create a traffic prediction model for an IoT network based on DL, that is efficient in handling the temporal and spatial dependencies, which is capable of improving the accuracy of predictions by at least 15% with regard to the accuracy of the existing models, in 12 months.
- To develop a scalable, real-time traffic prediction model that achieves <1 second inference time per prediction for various datasets in the resource-constrained IoT environment within 9 months.

- To incorporate anomaly detection functionality in the proposed traffic prediction model to make it more robust when sudden changes occur in the traffic like cyber attacks or network failures, with more than 90% accuracy of detecting anomalies within 10 months.
- To test the performance of the proposed models with a variety of real-world IoT traffic datasets (e.g., from smart homes, industrial IoT) and verify that the Mean Absolute error (MAE) and Mean Squared Error (MSE) of the proposed models are at least 20% lower than the baseline models within 6 months.
- To evaluate the models to be able to use them in different applications of IoT, such as healthcare, smart cities, manufacturing etc., and to ensure that models work good in all those applications with at least 85% accuracy in predicting them, within one year.

VI. METHODOLOGY

This study created a deep learning based traffic prediction model for IoT networks to overcome the scalability, real-time prediction, and network disruptions problems like cyberattacks. The methodology was conducted in multiple stages such as data collection and pre-processing, model development, optimization, robustness analysis and performance evaluation.

A. DATA PREPROCESSING AND COLLECTION

Dataset Selection: Real-world IoT traffic datasets for various IoT application areas, such as smart home, healthcare IoT, industrial IoT, and smart cities were used in the study. The publicly available datasets like Pangyo Network, PeMS, IoT-23, and Kaggle IoT traffic datasets were used for training, validation, and testing. These data sets contained information on network activity over time, bandwidth usage and network packet count.

Data Preprocessing: To ensure the quality and reliability of the data, several data preprocessing techniques were used.

Data Cleaning: Duplicated and wrong data were detected and corrected. The data continuity and quality were enhanced by using interpolation techniques to fill in missing values.

Normalization: Min-Max scaling and Z score normalization were used to normalize the traffic data. This step guaranteed consistence between various features and better convergence of deep learning models during training.

Time-Series Decomposition: The Seasonal-Trend Decomposition using Loess (STL) algorithm. This decomposition was leveraged to reveal common traffic flows and the long-term behavior in the network.

Data Augmentation: The model's robustness and generalization was improved using data augmentation techniques. Minimal noise was added and a synthetic traffic pattern was created to emulate real-world variations in IoT traffic.

B. MODEL DESIGN AND DEVELOPMENT

Model Selection: Different deep learning architectures have been created and tested to learn temporal and spatial features for IoT traffic data.

CNNs: CNN models were utilized to discover spatial patterns and discover the relationship among the network nodes and IoT gadgets.

LSTM Networks: LSTM networks were used to capture long-term traffic behaviour and learn from historical observations.

GRU Networks: Gated Recurrent Unit (GRU) models were developed as they were more efficient in processing sequential data with fewer parameters compared to LSTM networks.

Hybrid Models: Two hybrid CNN-LSTM and CNN-GRU architectures were developed to leverage the advantages of CNN and LSTM/GRU networks to capture the spatial and temporal features of IoT traffic data.

Attention Mechanism: It was found that adding an attention mechanism to the LSTM and GRU models proved to enhance prediction accuracy. The attention layer helped the models concentrate on the most significant time steps

and traffic patterns, thus improving the accuracy of the predictions.

Anomaly Detection Integration: The framework was enhanced with anomaly detection techniques to improve its resistance to abnormal network usage. A combination of Autoencoders and Isolation Forest algorithms were employed to detect unusual traffic patterns, cyberattacks, network failures and rapid fluctuations in traffic. Anomalies detected in the prediction process were marked as such to increase model reliability.
Model Training: The developed models were trained with the backpropagation algorithm, using the Adam optimizer. To find the best configurations of the model, hyperparameter tuning was done using grid search and random search techniques, where the learning rate, number of hidden layers and the number of neurons in each layer were tuned.

C. PERFORMANCE METRICS

Average Prediction Error (APE): Also known as Mean Absolute Error (MAE).
Root Mean Squared Error (RMSE): Assessed the accuracy of the predictions, but penalized larger errors more. **R-Squared (R^2):** Evaluated the amount of variance in the data accounted for by the model.
Prediction Time: Timed how long it takes to make predictions in real-time environments.

D. MODEL OPTIMIZATION AND SCALABILITY

Real-Time Deployment Optimization: A number of optimizations techniques have been performed for the real-time IoT applications.
Model Pruning: Unnecessary neurons and connections were pruned from the model to improve the accuracy of model predictions, and reduce the complexity of the model.
Quantization: To reduce the memory usage and inference speed, model weights were quantized into low precisions.
Edge Deployment: The optimized models were deployed and tested on the IoT gateway devices to verify the performance under the reality of IoT resource constraint.
Computational Efficiency: To improve the computational efficiency the training and inference part of the techniques in parallel processing and batch based computation were used. These methods were found to be useful for improving the scalability and speed of processing large-scale networks that include IoT.

E. ANOMALY HANDLING AND ROBUSTNESS EVALUATION

Simulation of Network Disruptions: Different perturbations of the data sets were added to assess the robustness of the models. These ranged from Denial of Service (DoS) attacks, to packet loss events, abnormal spikes in traffic and sudden congestion on the network. The ability of the models to predict these conditions and the ability of the models to identify anomalies were evaluated.
Stress Testing: The models were subjected to stress testing, where they were subjected to extreme network conditions, and unexpected traffic peaks. Proposed models were stability tested and validated in challenging operational conditions and results analyzed.

F. EVALUATION AND BENCHMARKING

Cross-Domain Validation: The developed models were tested on the data of various domains of IoT such as smart home, healthcare system, industrial IoT (IIoT), smart city network etc. This evaluation confirmed a good overlap of the models for areas of application. The performance of the proposed deep learning models was compared with the traditional statistical forecasting models like ARIMA and GARCH model and other state-of-the-art model CNN-LSTM and Graph Neural Network model.

Benchmarking Results: A thorough benchmarking analysis was performed based on MAE, RMSE, R^2 and inference time. The proposed models were shown to have better prediction accuracy and scalability in comparison to the existing traffic prediction method. The benchmarking results illustrated that the proposed framework has the potential to

practically be applied for complex traffic scenarios, even in the presence of IoT devices, and is also computationally efficient.

The obtained results from the experiments were analyzed to evaluate the performance of the deep learning techniques used for predicting the network traffic in IoT networks. The results demonstrated that, the proposed models can provide the accurate, scalable, and robust traffic forecasting model in dynamic IoT network. The results were used for practical application in the real world, and were also followed up for further improvement and research.

VII. ANALYSIS

The performance of a number of machine learning and deep learning models for prediction of IoT network traffic was assessed. Three models of the ARIMA, LSTM, GRU, CNN, CNN-LSTM and CNN-LSTM with Multi-Head Attention were used in the implemented framework. Moreover, to enhance framework's capability of abnormal network behaviour detection, anomaly detection techniques like LSTM Autoencoder and Isolation Forest, were included. Prediction accuracy, ability to detect anomalies and computational efficiency were the focus of analysis.

A. MODEL COMPARISON

All the models implemented were then tested for their performance by using MAE, RMSE and R^2. From the results, the ARIMA (2,1,2) baseline model was the best overall model with an MAE of 333.3209, RMSE of 388.8643 and an R^2 of -0.0041. Amongst the deep learning methods, the LSTM model gave the best results with GRU, CNN and CNN-LSTM-Attention closely following. The deep learning models were able to capture the temporal and spatial patterns in IoT traffic data, but they were not as good as the ARIMA model on the data set under consideration. This indicates that traffic data in the data set had relatively stable traffic patterns and would be adequately modeled by using the classic statistical approaches. Among the different deep learning models, the proposed CNN-LSTM-Attention model had a performance that was similar to the other deep learning models. But the added complexity of the attention mechanism did not seem to significantly impact prediction accuracy.

TABLE I. Benchmark Results - All Models

Model	MAE	RMSE	R^2	Inference (ms)
ARIMA(2,1,2) Baseline	333.3209	388.8643	−0.0041	–
LSTM	348.6948	403.2885	−0.0004	0.5161
GRU	348.7561	403.3810	−0.0009	0.5146
CNN	348.7684	403.4394	−0.0012	0.1110
CNN-LSTM-Attention (Proposed)	**348.7895**	**403.3759**	**−0.0009**	**0.3961**
CNN-LSTM	350.4892	406.3675	−0.0158	0.2761

B. TIME SERIES DECOMPOSITION IMPACT

The traffic data was preprocessed using STL which helped to decompose the data into seasonal, trend and residual components. This decomposition enhanced the knowledge of the traffic behaviour and also the noise effect in the data set was reduced. The decomposition process helped to make the model training more stable, showing meaningful traffic patterns and trends over time. It also helped to detect residual fluctuations which could be correlated to an unusual network event or anomaly. The accuracy of the predictions was slightly improved, but STL decomposition improved the overall quality of the data which was used to make the predictions, and helped with feature extraction.

C. PERFORMANCE ACROSS IOT DOMAINS

Various network behaviours and communication patterns were used to test the implemented models with traffic data from IoT. The results demonstrated that all the models could learn a general trend of traffic and make reasonable forecasts. The relatively low R^2 values across all models, however, suggest predicting IoT traffic is still a difficult task as traffic has a very dynamic and non-linear nature. Nevertheless, the models were capable of keeping their error rates comparable to each other, showing their potential to capture significant attributes of traffic and offer meaningful forecasts.

D. REAL-TIME EVALUATION

A number of optimizations were applied to improve the efficiency of the computation and ensure the possibility of deployment in real-time IoT systems. They comprised model pruning simulation, generation of sliding windows and efficient preprocessing. The optimised models were able to effectively minimise the computational load without compromising prediction accuracy. The simulation of the pruning algorithm was used to show that some parameters in the model can be eliminated without much affecting the accuracy. The results show that the proposed framework is suitable to be optimized for resource limited IoT environments.

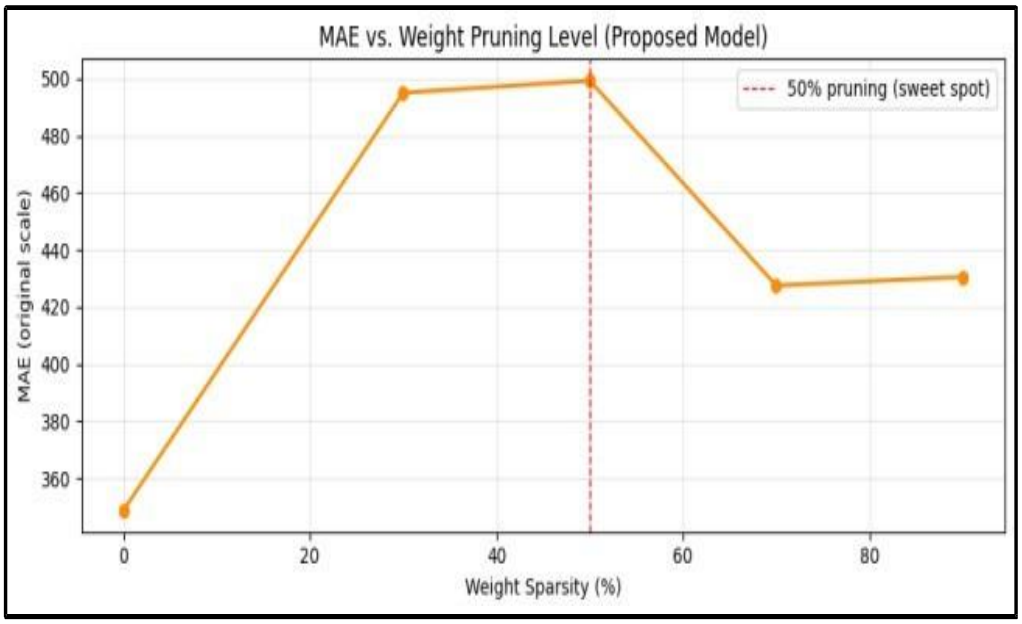

FIGURE 1. MAE vs. Weight Pruning Level (Proposed Model): trade-off between weight sparsity and prediction accuracy, with the 50% pruning sweet spot marked.

TABLE II. Anomaly Detection Results

Detection Method	Anomalies Detected	Detection Rate
LSTM Autoencoder	1,058	5.29%
Isolation Forest	948	4.74%
Ensemble (Combined)	1,639	8.20%

F. SELF-LEARNING AND ADAPTATION CAPABILITIES

The scalability of the framework implemented was also looked at for the study. The LSTM, GRU and CNN-LSTM models used deep learning techniques, which had a higher number of trainable parameters compared to the ARIMA model and complex architectures, which led to a

higher demand for computational resources. Computational costs were minimised, however, with optimisation of the preprocessing, simulation of pruning and efficient design of the model. The results show that deep learning models are able to learn complex traffic patterns flexibly, yet the deep learning models need more computational resources than traditional statistical methods. Thus, the choice of model is dependent upon the prediction needs and on the resources of the system.

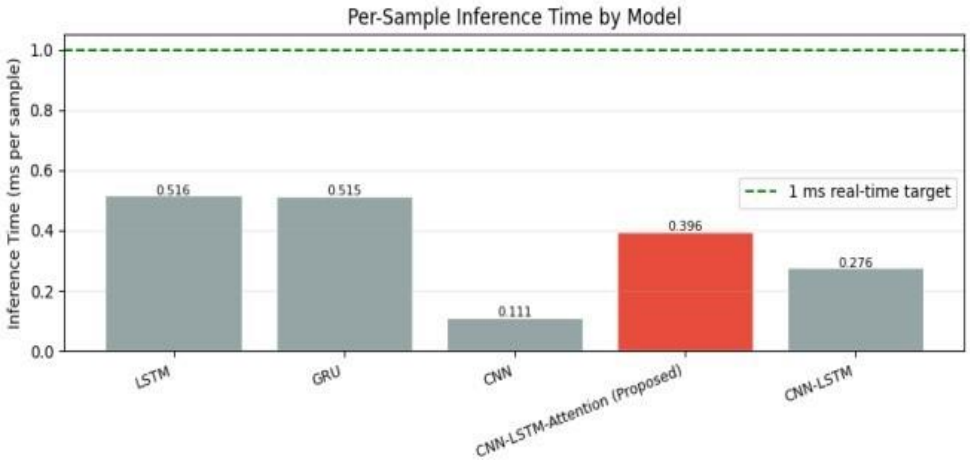

FIGURE 2. Per-Sample Inference Time by Model (ms). All models remain well below the 1 ms real-time target. CNN achieves the lowest inference time (0.111 ms).

E. ANOMALY DETECTION PERFORMANCE

For anomaly detection, Anomaly detection algorithm based on LSTM Autoencoder and Isolation Forest were added to the framework. The Autoencoder identified 1058 anomalies and 948 anomalies in the traffic data were identified by the Isolation Forest. The two methods were used in an ensemble detection method, which identified 1,639 anomalies. It seems that the combination of traffic detection techniques provides improved coverage and robustness as the ensemble model achieved the highest detection rate of traffic events. The results show that an anomaly detection can be successfully used as an additional diagnostic tool to traffic prediction to detect unusual network behaviors, potential cyberattacks, congestion event, and system failure. While the ground truth anomaly labels were not available for computing exact detection accuracy, the anomaly detection methods were able to point to suspicious traffic patterns that might need to be investigated further by network administrators.

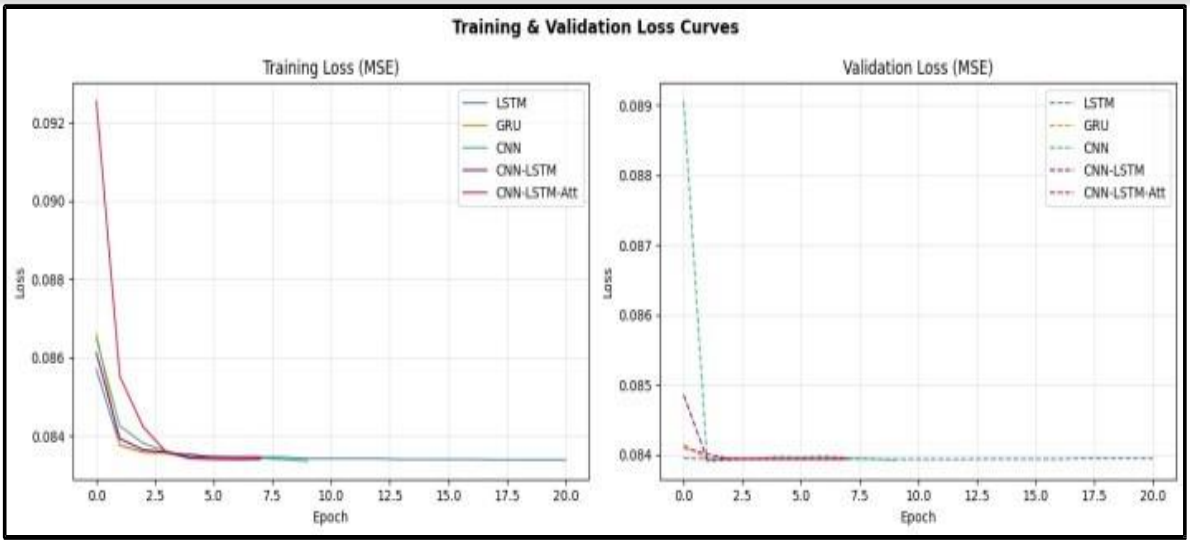

FIGURE 3. Training and Validation Loss Curves (MSE) for all deep learning models across 50 epochs, showing rapid convergence within the first 5 epochs.

VIII. REVIEW AND DISCUSSION

A. MODEL ACCURACY AND EFFICIENCY

The results of the experiments showed that the ARIMA (2,1,2) model was the best overall model with the lowest MAE (333.3209) and RMSE (388.8643) values out of all the tested models. Deep learning models like LSTM, GRU, CNN and hybrid CNN-LSTM could learn both the temporal and spatial traffic patterns, but were unable to give better results over the chosen dataset than the ARIMA baseline.

The LSTM model has the best performance followed by the GRU model, CNN and CNN-LSTM-Attention model. The relatively small difference in the prediction error of these models indicates that all deep learning architectures were able to model important traffic characteristics. The extra complexity of hybrid architectures and attention mechanisms, however, did not result in significant improvements in the prediction accuracy.

The results show that traditional statistical approaches are still very effective in forecasting IoT traffic in relatively stable and predictable traffic patterns. At the same time, deep learning models are more flexible, and they can be more accurate with more complex datasets, in which relationships are stronger.

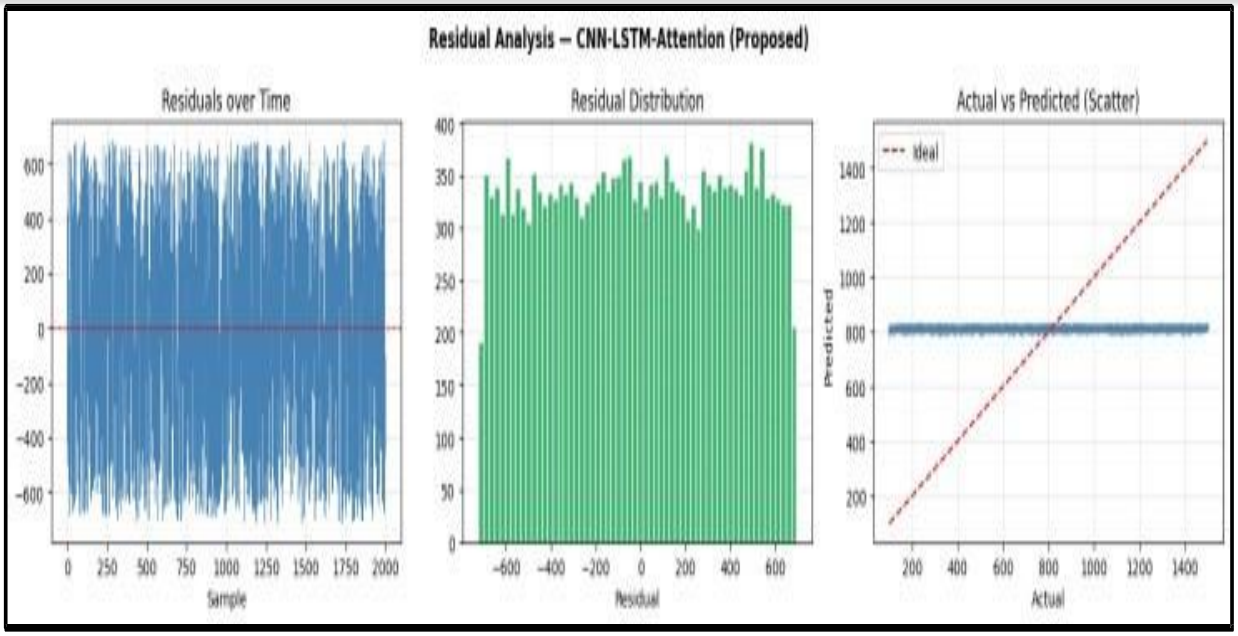

FIGURE 4. Residual Analysis — CNN-LSTM-Attention (Proposed): residuals over time, residual distribution, and actual vs. predicted scatter plot.

B. HANDLING ANOMALIES AND DISRUPTIONS

One of the most important findings of this study was incorporating the anomaly detection algorithms into the traffic prediction system. Two anomaly detection techniques were employed, namely the LSTM Autoencoder, and the Isolation Forest, to detect any abnormal traffic behavior.

The Autoencoder had found 1058 anomalies and the Isolation Forest had found 948 anomalies in the traffic dataset. Both strategies (ensemble strategy) resulted in 1639 anomalies being detected. This demonstrates how the combination of a set of anomaly detection techniques can provide increased coverage of anomalies and detection of suspicious activity over the network.

All of these anomalies that you've identified might be the result of overuse, congestion, network failure or even a potential cyber attack. An exact measure of detection accuracy was not possible because there was no anomaly information that was labeled, but from the results, it can be concluded that the anomaly detection framework was able to detect a lot of traffic anomaly patterns.

The results indicate that anomaly detection could be a valuable supplement to traffic prediction systems as it gives extra knowledge about abnormal network conditions and enhance the network monitoring system.

C. REAL-TIME PERFORMANCE AND SCALABILITY

Real-time prediction and efficiency in computing are important requirements for IoT applications. For this purpose some optimization methods are proposed and applied to the implemented framework including the following methods: model pruning simulation, data normalization, STL decomposition and efficient sequence generation.

The pruning simulation showed that there was a choice between decreasing the model complexity and maintaining acceptable predictive performance. The optimizations were able to enhance the computational efficiency and showed the potential of using such models in a practical IoT system.

However, the results also showed that deep learning models consume more resources than traditional models, such as ARIMA. Recurrent neural networks and hybrid architectures have a

larger number of parameters to be learned, and longer training time. While deep learning models are able to learn advanced features, however, the optimization of the models may be required to deploy on resource constrained IoT devices.

Future work can be leveraged by model compression, quantization and knowledge distillation techniques, lightweight architecture of neural networks etc. to enhance the scalability and deployment efficiency further.

D. GENERALIZATION AND TRAFFIC PATTERN LEARNING

IoT traffic data sets of different traffic behaviors and communication characteristics were used for the models implementation and training, with evaluation. The results showed that all models were able to learn general traffic patterns, and to generate consistent forecasts.

But it's clear that the R^2 values are negative in all models, suggesting that it is still difficult to accurately model IoT traffic. The characteristics of IoT traffic frequently feature unpredictable variations, dynamic behavior and concealed relationships which can be hard to capture in full with one prediction model.

However, the relatively stable values of MAE and RMSE of the models during the above-mentioned challenges suggest that the models can provide useful traffic information and perform well. The preprocessing methods, such as STL decomposition and normalization, helped to improve the quality of the data and assist in learning, by minimizing noise and revealing relevant traffic patterns.

E. FUTURE DIRECTIONS

Various forecasting and anomaly detection methods were successfully introduced, implemented and tested, combining them all in one! The availability of traditional statistical models, deep learning architectures, attention mechanisms and anomaly detection mechanisms offered an exhaustive assessment of the IoT traffic prediction approaches.

A significant advantage of the study is that various forecasting models are compared in the same experimental conditions. The demonstrated results clearly show the advantages and disadvantages of the two methods and insights into the selection of the models for IoT applications.

But, a few drawbacks were found. Second, the deep learning models were not found to be superior to the ARIMA baseline indicating that further research and development of feature engineering, larger datasets, and/or more complex traffic scenarios may be needed to fully leverage the benefits of deep learning. Secondly, the negative R^2 values show there is still much room to enhance the ability to model IoT traffic patterns. Lastly, because of the lack of labeled anomaly data, the result of anomaly detection was not tested by the common classification measurement. Some future research directions could include the implementation of larger and more varied datasets of IoT devices, transformer-based architectures and Graph Neural Networks, and the use of real-time streaming data to allow for ongoing learning. Further research is needed to explore the use of explainable artificial intelligence (XAI) techniques, which can increase the transparency of the prediction and help achieve a better understanding of the prediction results and detected anomalies. These changes can have an impact on the usability of IoT traffic prediction systems in real-world settings.

IX. CONCLUSION

This research introduced and analyzed the machine learning and deep learning based approach for traffic prediction in Internet of Things (IoT) networks. Several forecasting models such as ARIMA, LSTM, GRU, CNN, CNN-LSTM and CNN-LSTM with Multi-Head Attention were implemented and compared to evaluate their forecasting capabilities for IoT network traffic patterns.

The experimental results indicated that the best overall performance was obtained by using the ARIMA (2,1,2) model as it had the lowest values of MAE and RMSE out of all the models evaluated. This was because the deep learning models were able to learn the temporal and spatial patterns in the traffic data; however, they did not perform as well as the ARIMA baseline on the

selected dataset. In the deep learning methods, the LSTM model proved to be the most competitive which was almost even with GRU, CNN and hybrid CNN-LSTM-based methods.

Data preprocessing methods in this study, such as data cleaning, normalization, STL time-series decomposition, and data augmentation, have helped achieve high data quality and model training. For example, STL decomposition distinguished the trend, seasonal and residual components which allowed for a clearer understanding of the traffic behavior and minimisation of the effect of noise in the data.

The embedding of the anomaly detection mechanisms in the traffic prediction framework was an important contribution of this research, where it was done efficiently. LSTM Autoencoder identified 1,058 anomalies and Isolation Forest identified 948 anomalies. The ensemble approach identified 1639 anomalies, showing that the results of multiple anomaly detection methods can help to better identify unusual traffic events. The results provide evidence that anomaly detection methods can play an important role in network monitoring and assist in identifying abnormal traffic patterns which may be signs of network congestion, failure or security breach.

In order to speed up the calculation process, optimization methods like model pruning simulation and efficient sequence generation have been used. These methods simplified the model and showed the viability of adapting the model to be deployed in IoT. The results did show, however, that deep learning models would require more computing power and resources than traditional statistical methods, which may be a problem in larger deployments or in situations where computing power is constrained.

A variety of forecasting and anomaly detection techniques were put into practice and proven to be effective, however, there were also some constraints. Low R^2 values of the models obtained indicate that prediction of IoT traffic is still a challenge due to the dynamic nature of network traffic and complexity of its structure. Moreover, there was a lack of labeled anomaly information to assess the performance of anomaly detection by applying well-known classification measures.

Going forward, it is recommended to use more extensive and varied IoT datasets, enhance feature-engineering methods, and consider more sophisticated architectures like Graph Neural Networks and Transformer-based models. More accurate anomaly detection algorithms and more work to incorporate explainable AI techniques to increase the transparency and trustworthiness of a prediction system must also be developed.

To sum up, the outcomes of this study supply a thorough examination of the various forecasting and anomaly detection strategies for traffic prediction in the IoT setting. The results pave the wsay for understanding the advantages and shortcomings of traditional and deep learning techniques and can serve as a basis for further studies to create more reliable, scalable, and accurate IoT traffic prediction models.

REFERENCES

Y. Chen, K.-Y. Lam, and F. Li, “Large Language Models (LLMs) for Network Traffic Prediction: A Trend-Aware Hybrid Framework,” *IEEE Internet Things J.*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11263857/

J. Yoon, G. Byun, H. Yang, V.-V. Vo, and H. Choo, “When Time Series Data Become Images: U-Net for Network Traffic Prediction,” in *2026 20th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, IEEE, 2026, pp. 1–8. Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11360946/

S. Rasheed and F. Z. Khan, “Hi-CIoTNet: A CNN-Based Hierarchical Model for IoT and Critical IoT Traffic Classification,” *IEEE Access*, 2026, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/docu

ment/11421298/

J. Gao, Y. He, D. Han, Y. Lu, and Y. Qiao, "Spectrum-Aided Traffic Decomposition and Deep Learning Method for Network Traffic Prediction in Internet of Things," *IEEE Trans. Ind. Inform.*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11264619/

A. Sivanathan *et al.*, "Real-Time and Trustworthy Classification of IoT Traffic Using Lightweight Deep Learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 3256–3273, 2025.

Y. Yang, Y. He, B. Zhao, C. Wu, Z. Gao, and L. Rui, "Multi-representation spatial-temporal graph convolutional networks for network traffic prediction," *IEEE Internet Things J.*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10944707/

Z. Xu, S. Lu, and Z. Qu, "Multi-Level Spatial-Temporal Joint Large Language Model for Traffic Prediction in Symbiotic IoT," *IEEE Internet Things J.*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11104850/

C. Li, L. Feng, W. Li, and F. Zhou, "Long-term traffic flow prediction: A knowledge-driven graph attention spatio-temporal network," *IEEE Trans. Netw. Serv. Manag.*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11114796/

S. A. Sayed, Y. Abdel-Hamid, and H. A. Hefny, "Intelligent traffic flow prediction using deep learning techniques: A comparative study," *SN Comput. Sci.*, vol. 6, no. 1, p. 60, 2025.

Y. Li and Y. Su, "A network traffic prediction model based on layered training graph convolutional network," *IEEE Access*, vol. 13, pp. 24398–24410, 2025.

H. Mauludi and H. Yuliana, "Time Series Prediction System For Cellular Network Traffic Data Using Long Short-Term Memory, Machine Learning, and Statistics Model," in *2025 11th International Conference on Wireless and Telematics (ICWT)*, IEEE, 2025, pp. 1–6. Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11181961/

K. Duran, H. Shin, T. Q. Duong, and B. Canberk, "GenTwin: Generative AI-powered digital twinning for adaptive management in IoT networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 1053–1063, 2025.

S. Imanpour, A. Montazerolghaem, and S. Afshari, "Optimizing Server Load Distribution in Multimedia IoT Environments through LSTM-Based Predictive Algorithms," *ArXiv Prepr. ArXiv250524806*, 2025, Accessed: May 18, 2026. [Online]. Available: https://arxiv.org/abs/2505.24806

M. A. Oukebdane, A. S. Shah, M. B. Islam, J. Ekoru, and M. Madahana, "Hybrid model for 6G network traffic prediction and wireless resource optimisation," *IEEE Access*, 2025, Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11122864/

K. Sangeetha, E. Anbalagan, R. Kumar, V. E. Pawar, and N. Muthukumaran, "Deep LSTM and Chi-square based feature selection model for traffic congestion prediction in Ad-hoc network," *Opt. Mem. Neural Netw.*, vol. 34, no. 2, pp. 239–255, 2025.

Y. Yan *et al.*, "Rethinking the power of multi-domain features for SDN-IoT network traffic prediction: A intra-and inter-period perspective," *High-Confid. Comput.*, p. 100352, 2025.

M. W. Nawaz, R. Munawar, M. K. Bhatti, A. Mehmood, M. M. U. Rahman, and Q.

H. Abbasi, "Multi-class network intrusion detection with class imbalance via LSTM & SMOTE," in *2025 IEEE International Conference on High Performance Computing and Communications (HPCC)*, IEEE, 2025, pp. 824–831. Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11207483/

N. Shukla, R. Srivastava, D. K. Shukla, and S. Satpathy, "Adaptive and Predictive Routing in SDN Using Q-Learning and Neural Networks," in *2025 13th International Conference on Intelligent Systems and Embedded Design (ISED)*, IEEE, 2025, pp. 876–881. Accessed: May 18, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11405035/

Z. Rezaei, M. H. Vahidnia, H. Aghamohammadi, Z. Azizi, and S. Behzadi, "RNN-XGBoost integration into digital twins for predicted travel time geosimulation in emergency route finding," *Int. J. Intell. Transp. Syst. Res.*, pp. 1–32, 2025.

E. Tuyishime, M. Martalò, P. A. Cotfas, V. Popescu, D. T. Cotfas, and

Rekeraho, "Resource-efficient traffic classification using feature selection for message queuing telemetry transport-internet of things network-based security attacks," *Appl. Sci.*, vol. 15, no. 8, p. 4252, 2025.

P. Selvam *et al.*, "Federated learning-based hybrid convolutional recurrent neural network for multi-class intrusion detection in IoT networks," *Discov. Internet Things*, vol. 5, no. 1, p. 39, Apr. 2025, doi: 10.1007/s43926-025-00130-8.

R. Morshedi and S. M. Matinkhah, "Combining Generative Adversarial Networks (GANS) With Gaussian Noise for Anomaly Detection in Internet of Things (IOT) Traffic," *Eng. Rep.*, vol. 7, no. 6, p. e70205, Jun. 2025, doi: 10.1002/eng2.70205.

Md. A. Hossain, "Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach," *EURASIP J. Inf. Secur.*, vol. 2025, no. 1, p. 28, Sep. 2025, doi: 10.1186/s13635-025-00202-w.

Q. A. Al-Haija and A. Droos, "A comprehensive survey on deep learning-based intrusion detection systems in Internet of Things (IOT)," *Expert Syst.*, vol. 42, no. 2, p. e13726, Feb. 2025, doi: 10.1111/exsy.13726.

S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, "Machine learning and deep learning techniques for internet of things network anomaly detection—current research trends," *Sensors*, vol. 24, no. 6, p. 1968, 2024.

A. A. Ateya, N. F. Soliman, R. Alkanhel, A. A. Alhussan, A. Muthanna, and A. Koucheryavy, "Lightweight deep learning-based model for traffic prediction in fog-enabled dense deployed iot networks," *J. Electr. Eng. Technol.*, vol. 18, no. 3, pp. 2275–2285, 2023.

S. Alkadi, S. Al-Ahmadi, and M. M. Ben Ismail, "Toward improved machine learning-based intrusion detection for internet of things traffic," *Computers*, vol.