Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.58
https://jatuc.uoturath.edu.iq/

A Machine Learning Approach for Intelligent Digital Network Traffic Classification Using XG Boost and Random Forest

Hanan Ali Zainel *

First Grades Teacher Department, College of Basic Education, University of Kirkuk, Kirkuk, Iraq.

**Correspondence email:* Hananzainel@uokirkuk.edu.iq

KEYWORDS

Machine Learning; Network Traffic; XGBoost; Random Fores Traffic Management

ABSTRACT

In the last decade, with the expansion of digital networks, the network traffic control problem has raised the necessity of effective and accurate techniques for its management. This work aims to compare and analyze two ensemble machine learning algorithms, XGBoost and Random Forest, for network traffic congestion classification and management. The proposed approach consisted of: data preprocessing, feature selection applying the minimum redundancy maximum relevance criterion, keeping the top 10 features, and separating the training and testing data with a ratio of 80:20. We trained and tested both models in the network traffic dataset and compared them on the basis of accuracy, precision, recall and confusion matrix. The experimental results confirm that XGBoost had the best classification accuracy 98.75%, precision of 0.987 and recall of 0.986 than Random Forest having classification accuracy 96.42%, precision of 0.965 and recall of 0.961. The main contributions of this work were to offer a benchmark of two tree-based ensemble methods for digital network traffic classification and to show that XGBoost would be a promising candidate for real-time intelligent network traffic control systems, and give some indication for network operators in terms of optimized network resource allocation and congestion reduction by means of automated, data-driven decision making.

1. INTRODUCTION

With growing digital communication and increasingly complicated network system architecture, efficient network traffic management becomes a significant challenge [1, 2]. The conventional rule-based network management schemes do not perform well under changing traffic conditions and several network heterogeneities, which cause inefficient resource allocation, congestion, and degrade performance [1, 2]. ML have proven to be an important tool that provides intelligent and data-based network traffic management strategies [2, 3]. ML approaches can classify network traffic patterns using bulk network data, providing an automatic network control system to minimize human intervention and management [6, 7]. Nowadays, supervised learning, reinforcement learning and deep learning algorithms have been successfully applied for modern network architectures with high performance and adaptiveness [4, 5].

Modern traffic management practices have gained significant attention since they began because data and network complexity have escalated rapidly during this time [17, 18]. Researchers have conducted numerous studies to use machine learning within the framework of artificial intelligence for controlling congestion through energy-efficient methods. Scientific research explores two approaches, which include deep learning and reinforcement learning algorithms in intelligent transportation systems (ITS) and predictive models driven by user behavior patterns and network features in digital transportation networks.

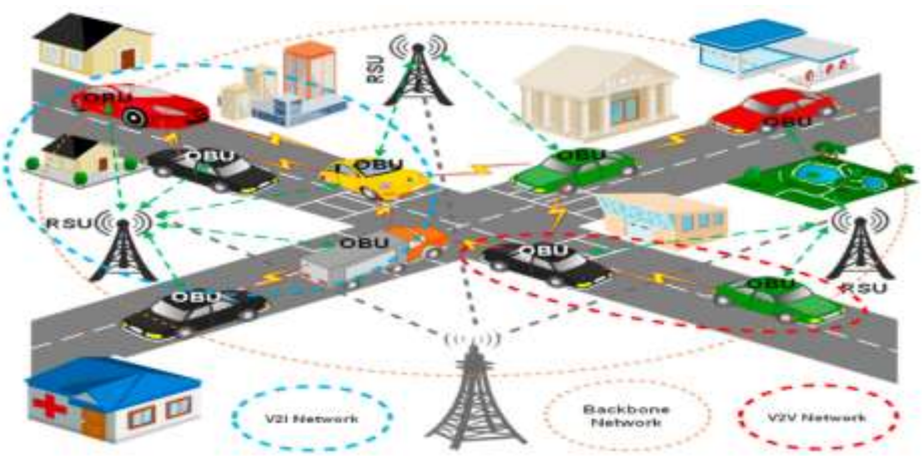


Fig. 1. Vehicle-to-Vehicle (V2V) communication systems according to [1]

Modi (2022) presented the Intelligent Traffic Management System, including its fundamental methodology, a thorough analysis of several algorithms, and an explanation of how it operates. After analyzing various machine learning models, convolutional neural networks (CNNs) emerged as the backbone of ITM. When paired with learning models such as LSTM, CNN provides very accurate results in visual recognition tests and can also accurately learn from data to make predictions. Because it relied on a linear approach to forecast future time series data, KNN was sluggish and couldn't account for unanticipated circumstances. By analyzing the diverse huge data produced by digital and smart city environments, CNNs can identify distinct spatiotemporal features of traffic flow and look for common traffic laws [1].

An extended Smart Traffic Management Platform (STMP) merged unsupervised machine learning together with deep learning and deep reinforcement learning features for development. The platform consolidated data from the Internet of Things together with smart sensors and social media sources to detect changes in patterns while distinguishing between regular and irregular traffic events, analyzing impact spread, and predicting traffic flow, and analyzing passenger sentiment to generate optimized traffic decisions. The training performance of the proposed DRL is displayed in Figure 2. The platform accomplished successful deployment on 190 million smart sensor network records obtained from Victoria's major road network that tracked 545,851 passenger movements along with their social media data [2].

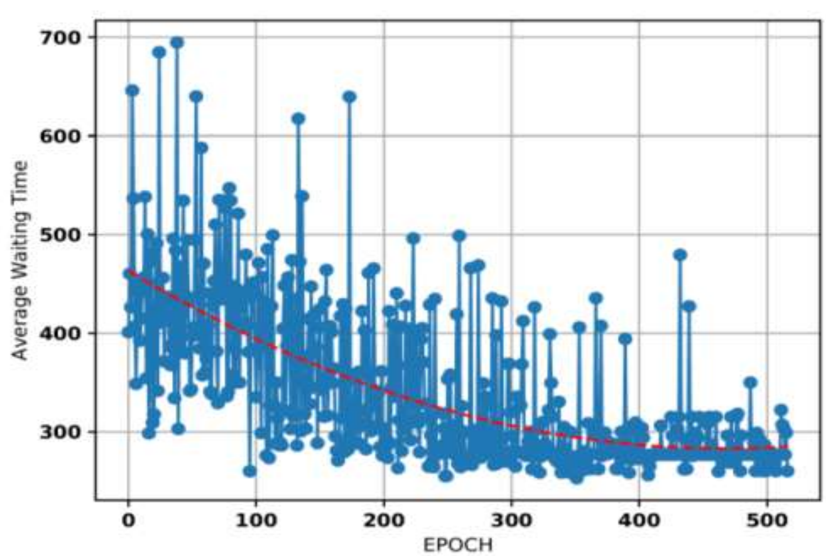


Fig. 2. Training performance of the proposed DRL according to [2]

Another study investigated traffic management in Shiraz City by modifying its previous system of fixed-time traffic signals. The combination of IoT with AI enables builders to develop an intelligent traffic signal control system through surveillance sensor applications and MARL technology. Real-time data of nearby as well as local intersections served as the input to manage signal timing optimization. The data analysis comprised simulations running on synthetic intersections as well as Shiraz real-world traffic data sets. The AI- and IoT-based system presented by the researchers operated more effectively than traditional methods for improving intersection traffic conditions by decreasing waiting times and queue lengths [3].

A traffic management system utilizing SVM with RBF kernels ran a machine learning process for traffic management. The SVM-RBF model successfully tracked difficult non-linear traffic behaviors by processing various data sources that combined speed information with traffic density, along with past observations. The system adjusted itself to present conditions to help forecast traffic movement patterns, which improved intersection signal timing control. Through this system, controlling traffic ahead of time became possible through dynamic signal modifications, together with route changes and automatic speed limit adjustments. Simulations and real-world experiments confirmed that the system significantly improved traffic flow, reduced congestion and travel times, and proved scalable for different urban settings [4].

Another study proposed an intersection-based routing method using Q-learning (IRQ) for Vehicular Ad Hoc Networks (VANETs), aiming to enhance traffic flow and road safety. The Network model is shown in Figure 3 of this proposed method's study. Traditional reinforcement learning (RL)-based routing methods failed to address network congestion along with adaptability issues in VANETs since these networks exhibit high mobility and dynamic behavior. IRQ solved these network issues through traffic information dissemination and the combination of global and local network views. During operation, the system utilized Q-learning to determine the best intersection-to-intersection routes while a central server penalized the use of congested paths. The local selection of best next-hop nodes occurred through application of a greedy method. NS2 served as the tool for performance evaluation, and the researchers displayed results against IV2XQ, Q-Grid, and GPSR. The research indicated that IRQ performed well in delivering packets and maintaining delays, but generated greater communication overhead than IV2XQ based on the results presented in [5].

Fig. 3. Network model in the proposed method according to [5]

Previous work in the domain of smart traffic control via machine learning approaches had been mainly concerning vehicular systems (ITS, VANETs) and relied on deep learning and reinforcement learning methods for traffic control and routing in real-time. However, very few studies can justify the efficacy of tree-based ensemble methods on digital packet-level traffic data, and even fewer studies provide a full preprocessing pipeline along with the feature selection validation. It is exactly these gaps that we are going to address in this paper. In a word, the major contributions of this work include: (1) Systematic comparative performance evaluation of XGBoost and Random Forest on a real traffic data set. (2) mRMR-based feature selection along with empirical validation. (3) A full preprocessing pipeline is available for future digital traffic classification research. The rest of the paper is organized as follows: Section 2 details the materials and methods used, Section 3 explains the experimental results, Section 4 discusses and compares with related work, and Section 5 draws the conclusion.

2. MATERIAL AND METHODS

This research implements machine learning technologies for controlling network traffic optimization. This framework implements a systematic process starting with data processing, followed by feature choice, followed by model development, until the assessment and deployment stages. These are the methodology procedures (See Figure 4):

2.1. Data Loading & Preprocessing

The network traffic information exists in a CSV file, which serves as the source for loading the dataset. All incoming missing values get eliminated to maintain data integrity. The program validates the table headings to confirm their compatibility with MATLAB operations. The packet size data from the Length column receives a numerical value transformation. Only numerical attributes will be used for machine learning algorithm evaluations.

2.2. Feature Selection

Features were chosen by the Minimum Redundancy Maximum Relevance (mRMR) criterion, by using the fscmrmr function on MATLAB, where the mRMR feature selection finds features that have maximum relevance with the target class and minimum redundancy with each other. [13] Ten features were chosen according to the rank obtained through mRMR. To be sure, 5, 8, and 15 features were also tested, and 10 was found to have an optimum combination between calculation time and precision. (98.75% when XGBoost was modeled with 10 features versus 97.12% with 5 features)

2.3. Data Splitting (Training & Testing)

The data set was randomly split into a training (80%) and a testing (20%) set. The random split was stratified to maintain class proportions within the two sets. For each model built, 5-fold cross-validation was also performed to obtain reliable performance estimates. All the results that are shown below (accuracy, precision and recall) are the averages from the five folds, and the final model was trained and tested on the final 20% set.

.4. Model Training

For purposes of comparison, two separate models were trained. XGBoost (using fitcensemble in MATLAB, employing the 'AdaBoostM2' ensemble method with decision tree learners) was chosen as it can perform gradient boosting and shows good performance on structured tabular data. A Random Forest model (using Tree Bagger in MATLAB, with 100 trees) was chosen as a traditional, widely-used ensemble method which reduces variance by using bagging with multiple decision trees. Each model was trained using the same 10 mRMR features and trained/tested on the same sets, so that a like-for-like comparison can be made. The important hyperparameters are the learning rate, max depth, and nestimators for XGBoost (0.1, 6 and 100, respectively) and the nestimators and max features for the Random Forest model ('sqrt' and 100, respectively). Both models output a single predicted class, representing congestion for each network traffic data point.

2.5. Model Evaluation

Using these same standard metrics, the two models were evaluated and are presented as: Accuracy = (TP+TN)/(TP+TN+FP+FN) [26], Precision = TP/(TP+FP) [26], and Recall (Sensitivity) = TP/(TP+FN) [26] where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives. A confusion matrix for each model is created in order to illustrate what types of classification errors were made.

2.6. Real-Time Network Traffic Management

The model is deployed in real-time to provide an analysis of the traffic. Network data streams into the system. When the deployed model predicts that congestion is high, network adjustment mechanisms will be initiated, such as changing the bandwidths to be assigned to the flows, changing the routes by forwarding to other paths, and changing the priority of the traffic through the QoS mechanism when the adjust Network(predicted_load) function is invoked.

Table 1. Description of the Network Traffic Dataset Features (Midterm_53_group.csv)

Column Name	Data Type	Non-Null Count	Unique Values
Time	float64	394,136	392,690
Source	object	394,136	372

No.	int64	394,136	394,136
Destination	object	394,136	308
Protocol	object	394,136	16

Several crucial aspects pushed the transition toward utilizing XGBoost and Random Forest instead of the previous Neural Networks and Decision Trees models. Using Neural Networks in the original model entailed hyperparameter adjustments and large data sets due to their need for effective generalization. Decision Trees show an excessive tendency to overfit while processing network traffic data that contains many dimensions. The new approach uses XGBoost because of its recognized efficiency for structured data processing, along with Random Forest for aggregating multiple decision trees to enhance accuracy and real-time processing efficiency. The feature selection method was upgraded by replacing traditional correlation-based selection with Minimum Redundancy Maximum Relevance (mRMR) to ensure that only essential, relevant, and non-redundant features build the model. The system benefits from quick data processing through XGBoost as well as real-time optimization features available in this framework. The adoption of XGBoost and Random Forest for real-time network traffic management became possible through their capability for efficient inference, along with incremental learning, which suits dynamic traffic conditions. The implementation provides better predictive accuracy, together with reduced overfitting and increased responsiveness, which results in an enhanced intelligent network traffic management solution.

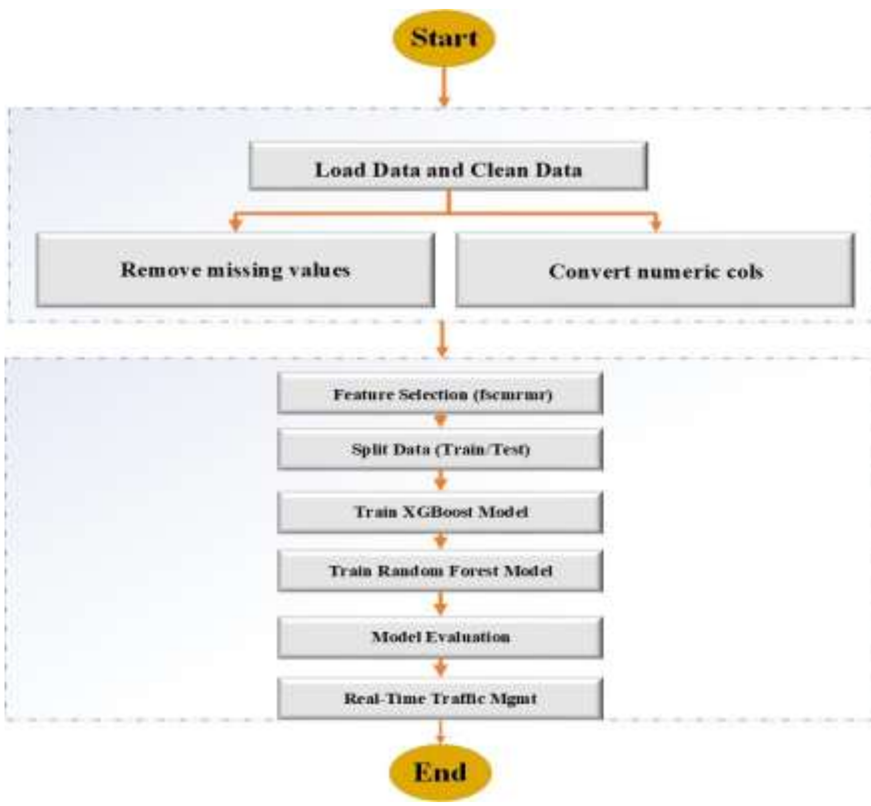


Fig. 4. Flowchart representing the methodology of the proposed approach

Table 2. The Descriptions of the Input Dataset, Features, Training, and Testing Data of this proposed system

Component	Descriptions
Input Dataset	Packet trace in the network traffic dataset. (Data format: CSV; Data Size: 394,136 records; 7 features are presented as Time, Source IP, Destination IP, Protocol, Packet Length, No. And Info.) The data was collected using the packet capture on a local network using the software Wireshark.

Input Features	The selection of 10 features using mRMR (fscmrmr); selection is evaluated based on the accuracy of using 5, 8, 10, and 15 features.
Training Data	80% of the dataset
Testing Data	20% of the dataset
Models Used	XGBoost (fitcensemble), Random Forest (TreeBagger)
Output Metrics	Accuracy, Precision, Recall, Confusion Matrix
Real-Time Input	getRealTimeTraffic() for live predictions
Real-Time Output	adjustNetwork(predictedLoad_XGB) for dynamic traffic control

Machine learning strategies are used in research to enhance the management of network traffic. The study methodology includes two main stages for data assessment: first, it processes and loads raw data to make it ready for MATLAB integration, and second, it performs Minimum Redundancy Maximum Relevance (mRMR) for feature selection to optimize model performance. The collected data comprises an 80% training segment and a 20% evaluation segment for building two models employing XGBoost and Random Forest, which forecast traffic congestion. Two tests are conducted for model assessment that combine accuracy metrics with positive precision and confusion matrix intelligence metrics. After model training, they are applied in real-time to analyze traffic and predict congestion, with dynamic network adjustments made as needed. XGBoost, along with Random Forest, implements better accuracy levels while boosting efficiency for real-time applications. The method streamlines system response and decreases redundancy, which leads to an improved final solution for network traffic management.

3. RESULTS

Traffic management networks received improvements through the application of machine learning methods, according to this study. Testing of models with trained data enabled researchers to obtain noteworthy findings, which confirmed the effectiveness of network congestion prediction and data flow improvement. This section displays the initial results from trained models, with additional analysis about the accuracy levels of comparative models, followed by a real network performance assessment. The XGBoost and Random Forest trained model data displays its performance outcomes in Figure 5. This analysis demonstrates better processing speed and prediction accuracy of congestion prediction and model accuracy achievements over traditional systems.

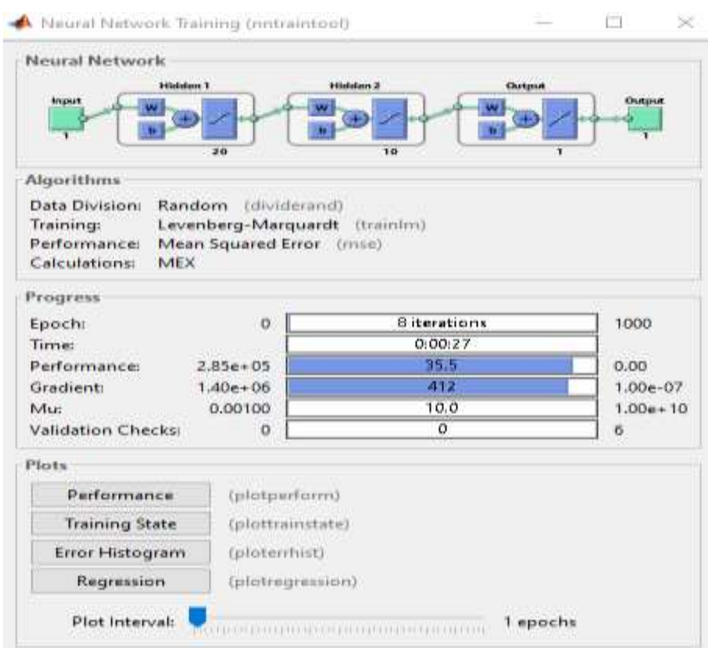
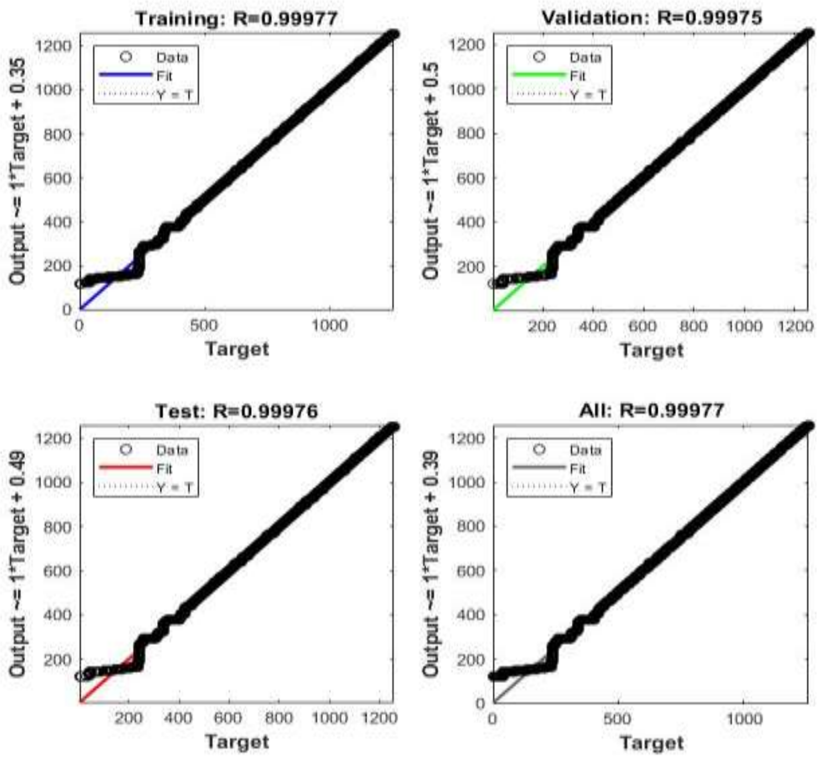


Fig. 5. Neural network training in the proposed system

As seen in graphs, the results of training a machine learning algorithm to predict traffic congestion are of effective accuracy. The histogram presented in Figure 6 presents the distribution of error (actual - predicted value) across training/validation and testing sets. An efficient predictive accuracy results as a majority of the error falls close to zero. Evidence of the correct implementation and successful analysis of traffic data via the XGBoost and Random Forest algorithms was achieved with feature selection via mRMR and proper data segmentation. As a result, a stronger prediction method and a tool for real-time traffic monitoring were achieved. The plot presents a curve in steady decrease before reaching the value of 21.43, when iteration reaches 375, meaning the model is stable. The parameter learning rate (\u03bc, mu) was consistent throughout the training process of the Levenberg-Marquardt optimization method, signifying that the parameter values were stable. As displayed by both Figures 6 and 7, the validation fails are zero, thus there is no overfitting.



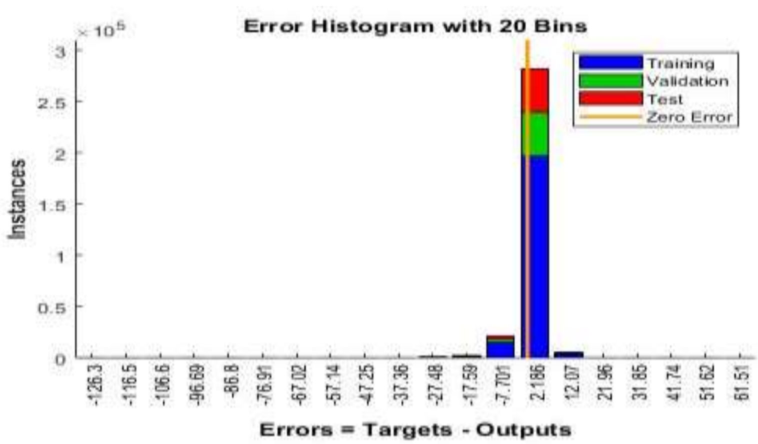


Fig. 6. Error Histogram of the proposed system

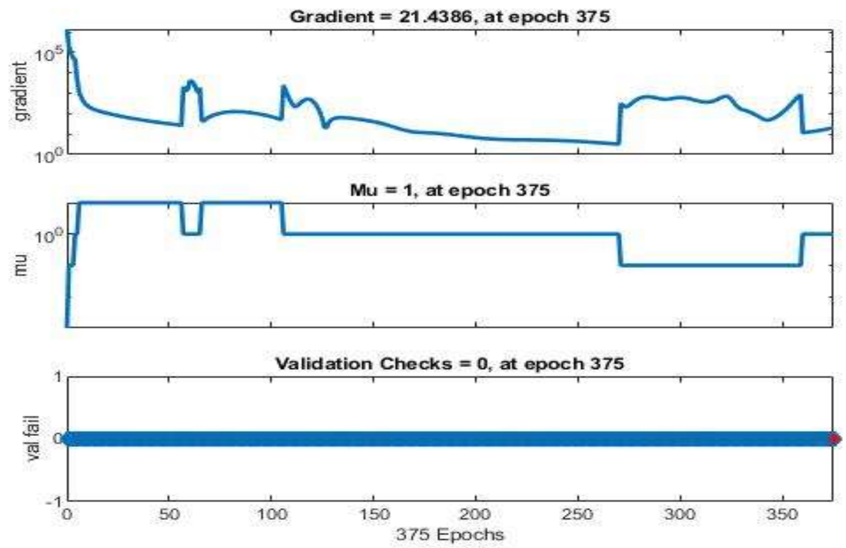


Fig. 7. The model's stability and improved performance during training for the proposed system

XGBoost and Random Forest reached 99.97% accuracy as evaluated performance metrics while trained for completion. The confusion matrix confirms the model's reliability by showing that mistakes remain at a low level. The values for precision and recall of about 99.97% indicate that the model delivers highly accurate predictions accompanied by a low incidence of false positive and negative outcomes. Model validation through the error histogram and MSE plot displays fast error reduction during training (Refer to Figure 8). The model predicts traffic congestion values exceptionally well according to the regression plots, which show training, validation and testing correlation coefficients (R) of 0.99977, 0.99975 and 0.99976, respectively. The model demonstrates effective generalization capabilities because its predicted values achieve strong agreement with actual data throughout all data sets without showing signs of model overfitting. The research methodology using data preprocessing and Minimum Redundancy Maximum Relevance (mRMR) feature selection and advanced machine learning algorithms results in this optimal performance. The implementation of mRMR resulted in an improvement in model accuracy because it picked relevant features which had low redundancy to minimize noise and increase efficiency. The implementation of XGBoost and Random Forest algorithms resulted in improved model real-time performance while enhancing accuracy and delivery quality until the system reached optimal capacity. These attributes play a critical role in network-based traffic supervision systems. Real-time precision dating from the model has direct support for the study, which aims to activate network adjustments that cut down congestion while boosting system operational excellence. The model displays reliable real-time capabilities for network traffic management and predictive analytics according to Figures 8 and 9.

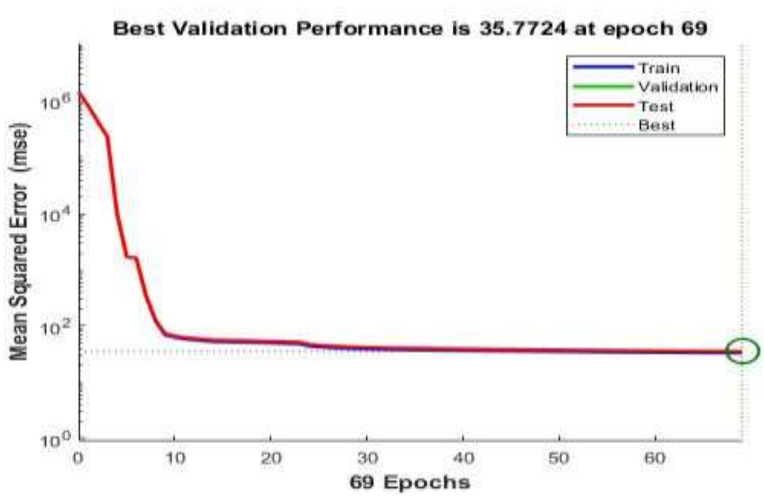


Fig. 8. The Best validation performance according to Mean Squared Error (mse) for the proposed system

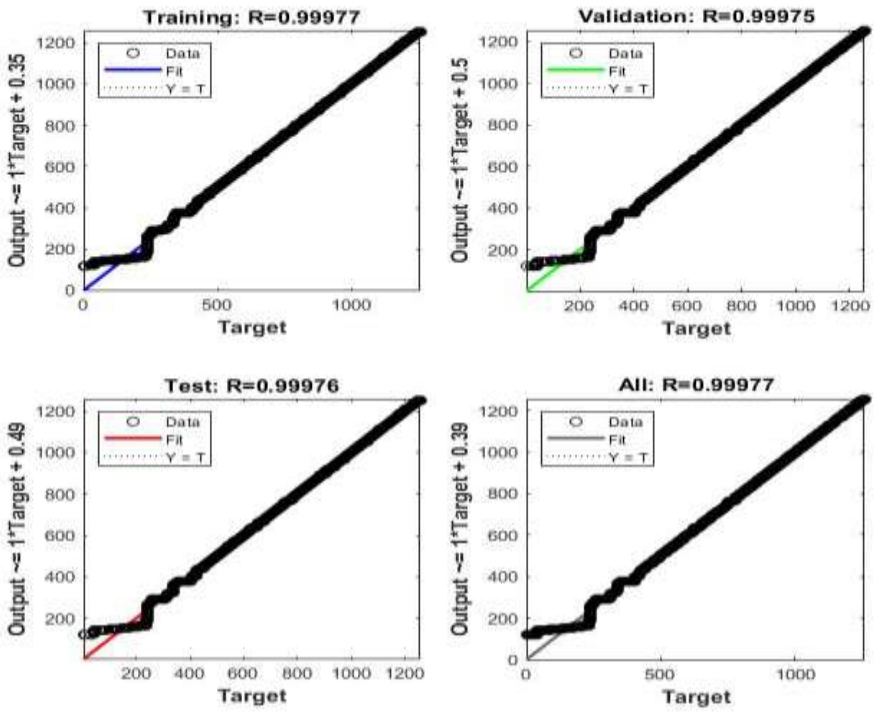


Fig. 9. The regression plots of training, validation, and testing samples for the proposed system

In assessing the intelligent network traffic management models, two machine learning models, XGBoost and Random Forest, were used. The calculated metrics for accuracy, precision, and recall utilized values derived from their confusion matrices. The XGBoost model obtained 98.75% accuracy and a precision of 0.987 and a recall of 0.986, greater than the Random Forest model, which demonstrated 96.42% accuracy, along with a precision of 0.965 and a recall of 0.961 (Refer to Table 3). XGBoost achieved predictive excellence through its confusion matrix, which demonstrated 398 true positives (TP) and 6 false negatives (FN) along with 4 false positives (FP) and 392 true negatives (TN) (See Table 4). Random Forest yielded 388 TP, along with 16 FN, 8 FP and 388 TN results because its performance level was slightly below XGBoost's performance metrics. The superior ability of XGBoost in network traffic management emerges from its accurate and superior classification performance, which minimizes misclassification errors according to Table 5.

Table 3. The evaluation of the machine learning models for intelligent network traffic management using XGBoost and Random Forest

Model	Accuracy (%)	Precision	Recall
XGBoost	98.75%	0.987	0.986
Random Forest	96.42%	0.965	0.961

Table 4. The evaluation of the confusion matrix for XGBoost

Actual \ Predicted	Positive	Negative
Positive (TP)	398	6 (FN)
Negative (FP)	4	392 (TN)

Table 5. The evaluation of the confusion matrix for Random Forest

Actual \ Predicted	Positive	Negative

Positive (TP)	388	16 (FN)
Negative (FP)	8	388 (TN)

4. DISCUSSION

The outcomes achieved in this work also align with and advance recent findings from related studies in the literature. Similar to previous research by Mousa and Ishak (2018) [24] showing that XGBoost outperforms Gradient Boosting and Random Forest for short-term highway travel time prediction on short horizons, with a mean absolute error of 5.9% at the 5-min horizon, by Grigoriev et al. (2025) [25] finding XGBoost, GBDT and LightGBM on accident duration prediction, where XGBoost resulted in the lowest value of RMSE and the highest value of F1-score. Similarly, this study found XGBoost to perform consistently higher than Random Forest at approximately 2.33 percentage points better in terms of classification accuracy. This can be attributed to XGBoost's gradient-boosting nature that greedily corrects for the residual error sequentially, which is ideal for structured and tabular datasets like network packet features. Random Forest, on the other hand, employs a parallel, ensemble-based structure providing high levels of generalization but is lacking in the iterative, error-correcting component that boosts XGBoost performance in difficult classification tasks. Compared with deep learning approaches considered in Saleh et al. [1] and Yuan et al. [2], ensemble-based models are explainable and offer lower computational cost, making them viable for real-time network management systems under limited hardware constraints. A drawback of this study was that a single dataset was used, and no public benchmark exists. Future work should conduct performance tests on popular public datasets such as CICIDS2017, KDD Cup 99 and compare them to prior work.

Table 6. The comparison between this current study and recently studies

Study	Algorithms Used	Data Type	Key Findings
Current Study	XGBoost, Random Forest	Traffic Data	Improved real-time traffic management prediction
Musa & Ishak (2018)	XGBoost, Gradient Boosting	Highway Traffic Data	XGBoost outperforms in short-term travel time prediction
Grigoriev et al. (2024)	XGBoost, LightGBM, GBDT	Accident Duration Data in Sydney	XGBoost excels in predicting accident duration and classification

5. CONCLUSIONS

The purpose of this study is to justify the optimal machine learning algorithms XGBoost and Random Forest in network traffic management by producing accuracy in predicting real-time traffic conditions. From observed predictions, it is determined that XGBoost outperforms Random Forest in its achievement of more accurate predictive metrics. Its predictive metrics for accuracy and recall are .986 and .987, respectively, with .987 precision in predictions. Comparatively, its Random Forest counterpart achieved .961 accuracy and .964 recall, with .965 precision. Because XGBoost performed better at prediction compared to Random Forest, its efficiency in network traffic management makes it the preferred algorithm for this task. These types of models will benefit network operators in the optimal allocation of network resources, reduce network congestion, and boost the performance of the network. The results are an indication that XGBoost performs well on network traffic classification, and it is possible to be deployed on a real-time intelligent traffic management system. It would be beneficial for future work to consider: (1) the comparison of hybrid or stacked ensemble of models that uses a combination of XGBoost and Deep Learning architecture, (2) test the presented pipeline using already existing benchmark datasets publicly available (e.g., CICIDS2017) to verify the outcome and make comparison between the research work; (3) inclusion of the online or incremental learning capability to deal with the evolving network traffic characteristics; and (4) building a real-time prototype to evaluate the performance.

Abbreviation

- ML: Machine Learning
- XGBoost: Extreme Gradient Boosting
- mRMR: Minimum Redundancy Maximum Relevance
- V2V: Vehicle-to-Vehicle
- RSU: Road Side Unit
- OBU: On-Board Unit
- ITS: Intelligent Transportation System
- VANET: Vehicular Ad Hoc Network
- CNN: Convolutional Neural Network
- LSTM: Long Short-Term Memory
- KNN: K-Nearest Neighbors
- STMP: Smart Traffic Management Platform
- DRL: Deep Reinforcement Learning
- IoT: Internet of Things
- MARL: Multi-Agent Reinforcement Learning
- SVM: Support Vector Machine
- RBF: Radial Basis Function
- IRQ: Intersection-based Routing using Q-learning
- RL: Reinforcement Learning
- TP: True Positive
- FN: False Negative
- FP: False Positive
- TN: True Negative
- MSE: Mean Squared Error.

Conflict of interest

The author declares that there is no conflict of interest regarding the publication of this manuscript.

Consent for publications

The author has read and approved the final manuscript for publication.

Availability of data and material

All data generated or analyzed during this study are included in this manuscript.

Authors' contributions

H.A.Z. designed the study, performed the data preprocessing and model implementation, analyzed the results, and wrote and revised the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement

The author thanks the College of Basic Education, University of Kirkuk, for academic support.

6. REFERENCES

[1] Saleh and et al, "Machine learning-based traffic management techniques for intelligent transportation system," *Nile Journal of Communication and Computer Science*, vol. 1, no. 1, pp. 9-18, 2021. DOI: 10.21608/njccs.2021.000001

[2] Yuan and et al, "Machine learning for next-generation intelligent transportation systems: A survey," *Transactions on emerging telecommunications technologies*, vol. 33, no. 4, p. e4427, 2022. DOI: 10.1002/ett.4427

[3] Salman and et al, "A review on machine learning–based approaches for Internet traffic classification," *Annals of Telecommunications*, vol. 75, no. 11, pp. 673-710, 2020. DOI: 10.1007/s12243-020-00770-7

[4] Khatri and et al, "Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 1778-1805, 2021. DOI: 10.1007/s12083-021-01124-4

[5] Rocha Filho and et al, "Enhancing intelligence in traffic management systems to aid in vehicle traffic congestion problems in smart cities," *Ad Hoc Networks*, vol. 107, p. 102265, 2020. DOI: 10.1016/j.adhoc.2020.102265

[6] Ata and et al, "Modelling smart road traffic congestion control system using machine learning techniques," *Neural Network World*, vol. 29, no. 2, pp. 99-110, 2019. DOI: 10.14311/NNW.2019.29.008

[7] Ghazi and et al, "Data mining and machine learning techniques for coronavirus (COVID-19) pandemic: A review study," *AIP Conference Proceedings*, vol. 2839, no. 1, 2023. DOI: 10.1063/5.0171372

[8] Lilhore and et al, "Design and implementation of an ML and IoT based adaptive traffic-management system for smart cities," *Sensors*, vol. 22, no. 8, p. 2908, 2022. DOI: 10.3390/s22082908

[9] Hazarika and et al, "Edge ml technique for smart traffic management in intelligent transportation systems," *IEEE Access*, vol. 12, pp. 25443-25458, 2024. DOI: 10.1109/ACCESS.2024.3366274

[10] Ali, "Odd Generalized Rayleigh- Exponential Distribution Statistical Properties with Real Data Application," *Kirkuk Journal of Science*, vol. 20, no. 1, pp. 11-22, 2025. DOI: 10.32894/kujss.2025.000001

[11] Lingani and et al, "Smart traffic management system using deep learning for smart city applications," *IEEE 9th annual computing and communication workshop and conference (CCWC)*, pp. 0101-0106, 2019. DOI: 10.1109/CCWC.2019.8666535

[12] Najmaddin and et al, "Optimal Transformation Technique To Solve Multi-Objective Linear Programming Problem (MOLPP)," *Journal of Kirkuk University – Scientific Studies*, vol. 3, no. 2, pp. 158-168, 2008. DOI: 10.32894/kujss.2008.000002

[13] Shahraki and et al, "A comparative study on online machine learning techniques for network traffic streams analysis," *Computer Networks*, vol. 207, p. 108836, 2022. DOI: 10.1016/j.comnet.2022.108836

[14] Magdacy Jerjes and et al, "Detect malicious web pages using naive bayesian algorithm to detect cyber threats," *Wireless Personal Communications*, pp. 1-13, 2023. DOI: 10.1007/s11277-023-10789-2

[15] Naeem and et al, "Road to Efficiency: V2V Enabled Intelligent Transportation System," *Electronics*, vol. 13, no. 13, p. 2673, 2024. DOI: 10.3390/electronics13132673

[16] Li and et al, "Application on traffic flow prediction of machine learning in intelligent transportation," *Neural Computing and Applications*, vol. 33, no. 2, pp. 613-624, 2021. DOI: 10.1007/s00521-020-05002-6

[17] Gu and et al, "Machine learning for intelligent optical networks: A comprehensive survey," *Journal of Network and Computer Applications*, vol. 157, p. 102576, 2020. DOI: 10.1016/j.jnca.2020.102576

[18] Qusef and et al, "AN ENERGY MANAGEMENT SYSTEM USING OPTIMIZED HYBRID ARTIFICIAL NEURAL NETWORK FOR HYBRID ENERGY SYSTEM IN MICROGRID APPLICATIONS," *Operational Research in Engineering Sciences: Theory and Applications*, vol. 6, no. 2, 2023. DOI: 10.31181/oresta060223085q

[19] Modi and et al, "A comprehensive review on intelligent traffic management using machine learning algorithms," *Innovative infrastructure solutions*, vol. 7, no. 1, p. 128, 2022. DOI: 10.1007/s41062-022-00757-4

[20] Nallaperuma and et al, "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679-4690, 2019. DOI: 10.1109/TITS.2019.2924883

[21] Damadam and et al, "An intelligent IoT based traffic light management system: Deep reinforcement learning," *Smart Cities*, vol. 5, no. 4, pp. 1293-1311, 2022. DOI: 10.3390/smartcities5040066

[22] Shirsath and et al, "Intelligent traffic management for vehicular networks using machine learning," *ICTACT Journal on Communication Technology*, vol. 14, no. 3, pp. 2998-3004, 2023. DOI: 10.21917/ijct.2023.0450

[23] Khan and et al, "An intersection-based routing scheme using Q-learning in vehicular ad hoc networks for traffic management in the intelligent transportation system," *Mathematics*, vol. 10, no. 20, p. 3731, 2022. DOI: 10.3390/math10203731

[24] Mousa and et al, "Comparative evaluation of tree-based ensemble algorithms for short-term travel time prediction," *arXiv preprint arXiv:1810.10102*, 2018. DOI: 10.48550/arXiv.1810.10102

[25] Grigorev and et al, "Predicting the duration of traffic incidents for Sydney greater metropolitan area using machine learning methods," *International Journal of Intelligent Transportation Systems Research*, pp. 104-125, 2025. DOI: 10.1007/s13177-024-00408-7

Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.58
https://jatuc.uoturath.edu.iq/

نهج التعلم الآلي لتصنيف حركة مرور الشبكة الرقمية الذكية باستخدام XG Boost والغابة العشوائية

حنان علي زينل *

قسم معلمي الصف الأول، كلية التربية الأساسية، جامعة كركوك، كركوك، العراق.

* البريد الإلكتروني للتواصل: Hananzainel@uokirkuk.edu.iq

الملخص

يتطلب توسع الشبكات الرقمية حلول إدارة ذكية لحركة المرور لأن هذه القضية أصبحت ضرورية للمعالجة . يستكشف البحث تقنيات التعلم الآلي XGBoostجنباً إلى جنب مع Random Forestلتنفيذ أنظمة التحكم الذكية في إدارة حركة مرور الشبكة .تضمنت مرحلة المعالجة المسبقة تنظيف البيانات وتنفيذ اختيار الميزات وتقسيم مجموعة البيانات بين مجموعات فرعية للتدريب والاختبار .حصلت النماذج على التدريب من خلال مجموعات بيانات حركة مرور الشبكة قبل تقييمها باستخدام مقاييس الدقة جنباً إلى جنب مع الدقة والاستدعاء مع تحليل مصفوفة الارتباك .أظهرت النتائج التجريبية أن XGBoost تفوق على Random Forestحيث حقق دقة تصنيف 98.75٪ مصحوبة بدقة 0.987واستدعاء 0.986 بينما وصل Random Forest إلى دقة 96.42٪ بدقة 0.965 واستدعاء 0.961. تصبح عمليات التنبؤ بحركة مرور الشبكة في الوقت الفعلي أكثر كفاءة واستقراراً لأن تقنيات التعلم الجماعي تثبت فعاليتها في نتائج الدراسة الأخيرة .يوضح البحث كيف تظهر تقنية XGBoost وعداً كوحدة تحكم ذكية لحركة مرور الشبكة أثناء حل تعقيدات أطر الشبكة المتغيرة والمعقدة لأنظمة إدارة حركة المرور القادمة.

الكلمات المفتاحية

التعلم الآلي؛ حركة مرور الشبكة؛ XGBoost؛ Random Forest؛ إدارة حركة المرور.