Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 1, October 2024, pp. 461~468
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i1.pp461-468

An evaluation of multiple classifiers for traffic congestion prediction in Jordan

Mohammad Hassan, Areen Arabiat
Department of Communication and Computer Engineering, Faculty of Engineering, Al Ahliyya Amman University, Amman, Jordan

Article Info

Article history:

Received Feb 1, 2024
Revised May 22, 2024
Accepted Jun 5, 2024

Keywords:

Fuzzy unordered rule induction algorithm
Logistic regression
Machine learning
Multi-layer perceptron
Naive Bayes
Stochastic gradient descent

ABSTRACT

This study contributes to the growing body of literature on traffic congestion prediction using machine learning (ML) techniques. By evaluating multiple classifiers and selecting the most appropriate one for predicting traffic congestion, this research provides valuable insights for urban planners and policymakers seeking to optimize traffic flow and reduce jamming and. Traffic jamming is a global issue that wastes time, pollutes the environment, and increases fuel usage. The purpose of this project is to forecast traffic congestion at One of the most congested areas in Amman city using multiple ML classifiers. The Naïve Bayes (NB), stochastic gradient descent (SGD) fuzzy unordered rule induction algorithm (FURIA), logistic regression (LR), decision tree (DT), random forest (RF), and multi-layer perceptron (MLP) classifiers have been chosen to predict traffic congestion at each street linked with our study area. These will be assessed by accuracy, F-measure, sensitivity, and precision evaluation metrics. The results obtained from all experiments show that FURIA is the classifier that presents the highest predictions of traffic congestion where By 100% achieved Accuracy, Precision, Sensitivity and F-measure. In the future further studies can be used more datasets and variables such as weather conditions; and drivers behavior that could integrated to predict traffic congestion accurately.

Corresponding Author:

Mohammad Hassan
Department of Communication and Computer Engineering, Faculty of Engineering
Al Ahliyya Amman University
Al-Saro, Al-Salt, Amman, Jordan
Email: mhassan@ammanu.edu.jo

1. INTRODUCTION

Increased urban traffic challenges encompass a spectrum of impacts, including reduced productivity, heightened air pollution, and increased fuel consumption, particularly as urbanization trends persist [1]. Within developed nations, such as the United States, traffic congestion stands as a significant economic burden, costing billions annually [2]. Scholars have delved into the multifaceted nature of congestion, commonly defining it as the point where transportation demand exceeds available road capacity. Despite strides in infrastructure development, congestion remains a pressing societal concern, often classified as either recurrent (stemming from high demand) or intermittent (arising from capacity constraints) [3].

Mitigating congestion requires a range of solutions, from improving infrastructure and encouraging public transit to anticipating traffic conditions, but with possible cost and practicality issues [4]. Through the application of internet of things (IoT), (AI), and machine learning (ML)technologies; the use of these technologies through big data can lead to accurate predictions of congestion patterns within the road network making it a better way of controlling urban traffic [5]. Big data and AI have impacted the growth of traffic congestion predictions using ML techniques by analysis of various traffic factors as captured by authors in [6], [7].

The LST model may be exploited in an innovative way AIand IoT traffic data to enhance the prediction of asphalt movement" in smart city settings. The LST model utilizes IoT sensors and deep learning algorithms to generate more accurate real-time vehicle count estimates to predict future traffic data and make better-informed decisions [8]-[11]. The potential benefits of combining IoT with AI are significant, including enhanced urban mobility, decreased traffic and congestion, and more effective traffic management. In this study, long short-term memory (LSTM) accuracy was 91%, whereas linear regression, K-nearest neighbor (KNN), and support vector machine (SVM) accuracy were 41%, 43%, and 46% of total, respectively.

The critical path method was initially introduced in a study done in Helsinki, Finland. Convolutional long short-term memory (CPM-ConvLSTM), a spatiotemporal model that predicts short-term congestion levels in each road segment, outperforms six rivals in traffic data prediction accuracy [10]. A traffic congestion forecast model is created using RF, a reliable and efficient ML approach. The model, which took into consideration the weather, time of year, unique road conditions, road quality, and holidays, had a low generalization error and an accuracy of 87.5% [12]. Congestion matrices for regional traffic networks are generated using a variety of approaches and the relative positions of road nodes. They used a convolutional long-short-term memory network to predict congestion throughout the network. The technique exhibited interpretability for congestion prediction by outperforming baseline models and accurately capturing traffic's temporal and spatial features [13].

Using data from the Greater Amman Municipality, this study use AI techniques, namely machine and deep learning, to forecast congestion using selected classifiers. The best classifier for forecasting traffic congestion on all routes to Amman City's eighth roundabout was established using a variety of statistical parameters, including accuracy, specificity, sensitivity, and the F1-measure. The use of large amount of traffic data from four traffic approaches to the research location is a new way to enhance the prediction, Furthermore, the Naïve Bayes (NB), stochastic gradient descent (SGD), fuzzy unordered rule induction algorithm (FURIA), logistic regression (LR), decision tree (DT), random forest (RF), and multi-layer perceptron (MLP) classifiers are used to find the best classifier with the highest performance, followed by a thorough examination and more than twenty experiments.

2. METHOD

2.1. System architecture of traffic congestion prediction

The study's goal is to estimate traffic congestion in Amman, especially at the intersection of the 8th roundabout, using various ML techniques. The Weka tool is used in the study to build ML models. Initially, data was obtained from the Greater Amman Municipality, preprocessed, and converted to a comma-separated value (CSV) format for compatibility with the Weka tool. Subsequently, the dataset is inputted into Weka to generate predictive models utilizing classifiers such as NB, SGD, FURIA, LR, DT, RF, and MLP. To optimize model performance, a random sampling technique with a 70% training and 30% testing set allocation was employed due to the substantial dataset size of 8,640 records per boundary. Additionally, 10-fold cross-validation was implemented to enhance model robustness and accuracy. Following the generation of a confusion matrix within WEKA, the results undergoes assessment based on metrics such as accuracy, f-measure, precision, and recall subsequent to the development of the trained model. Moreover, a comparative evaluation will be carried out to determine the ML model that demonstrates superior performance metrics. The results will elucidate the most effective ML algorithms for predicting traffic congestion, among other insights.

2.2. Dataset

The dataset sourced from the Greater Amman Municipality spanning from January 1, 2019, to December 31, 2019, was selected with the aim of facilitating accurate and effective prediction of traffic congestion. It encompasses variables such as traffic volume per lane, density, speed, occupancy, width, and distance. The determination of traffic volume for each lane approach was achieved through the utilization of detectors and sensors. Specifically, the dataset includes detailed traffic volume data for each approach on a lane over a 24-hour period, every month throughout the entirety of 2019. The attributes utilized for congestion prediction analysis and their visualization are delineated in Figure 1.

The above dataset includes the data of all approaches entering the 8^{th} roundabout, as shown in Figure 2, including westbound, northbound, eastbound, and southbound. The subsequent stage in the development of the traffic congestion prediction system is data cleansing, which involves the correction or deletion of incorrect, corrupted, improperly formatted, duplicate, or incomplete data from the dataset. In this process, redundant data was removed using the remove duplicates feature in Excel. Structural errors were rectified by adding a new rule using conditional formatting to correct incorrect classifications. The third step is data preprocessing, which is a crucial initial step in the generation of an ML classifier, ensuring the data is suiTable for the analysis requirements. The preparation module in Weka handles this process, typically saving the traffic dataset as a CSV file.

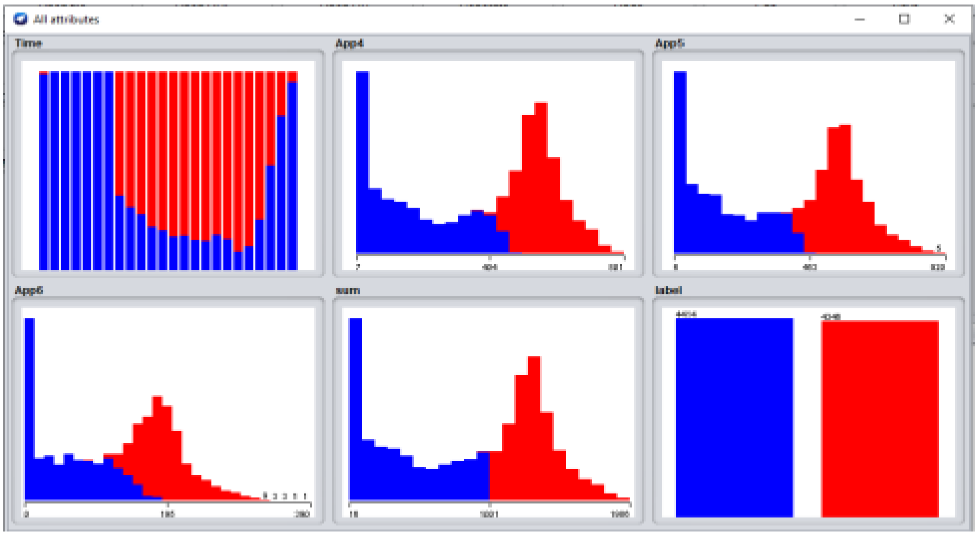

Figure 1. Attributes for congestion prediction analysis and attributes list visualization

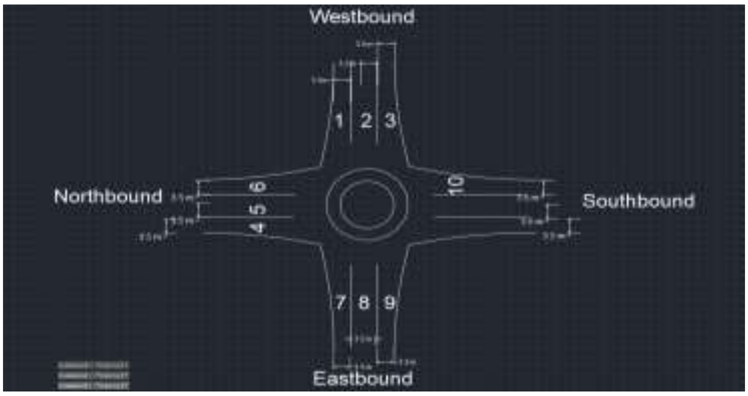

Figure 2. Study area

2.3. Machine learning classifiers

ML is a methodology employed to enhance machines' data processing capabilities. In the realm of smart transportation, ML plays a pivotal role in analyzing intricate relationships among road networks, traffic patterns, environmental factors, and traffic incidents. Transportation systems may be made safer, more sensitive to consumer requirements, and more efficient by utilizing machine learning. For example, traffic congestion may be identified and predicted using ML algorithms, which enables more efficient traffic control and shorter travel times. In order to detect and respond to issues like accidents or road closures, ML may also be used to evaluate sensor data from traffic cameras and other sources. This will make commuter transportation more dependable and seamless. In order to minimize traffic and cut emissions, ML may also be used to improve traffic signal timing and routing. This will assist create a more ecologically friendly and sustainable transportation system.

2.3.1. The Naïve Bayes

Built upon the NB principle, it computes posterior probabilities by leveraging prior probabilities. NB is particularly suiTable for high-dimensional data sets owing to its computational efficiency, robustness against noise, and capability for incremental learning. The Bayesian classification framework establishes the posterior probability through the formulation depicted in (1).

$$P(c|x) == \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

where: x is the feature vector, c is the classification variable, P(x) is the evidence, P(x|c) is the likelihood distribution, P(c|x) is the posterior probability [14].

2.3.2. Stochastic gradient descent (SGD)

SGD is a classifier that merges regularized linear models with SGD. SGD facilitates incremental learning through the partial fit method. However, in its pursuit of reaching the global minimum, SGD adjusts the network topology after each training iteration, reducing errors by approximating gradients for randomly chosen batches. This method involves random sampling and frequent high-variance adjustments, leading to significant variations in the objective function [15], [16].

2.3.3. Fuzzy unordered rule induction algorithm

FURIA, an evolution of the RIPPER algorithm. It excels in generating fuzzy rules instead of traditional ones and incorporates an effective rule stretching mechanism for identified cases. Experimental results demonstrate FURIA's superiority over other classifiers in terms of classification accuracy. FURIA utilizes fuzzy rules and unordered rule sets for classification, employing an unordered one to differentiate each class from the rest efficiently [17], [18].

2.3.4. Decision tree

DT classifiers are extensively employed in diverse domains, this classifier is structured with nodes and branches, utilizing a range of classification algorithms to handle missing values and both continuous and categorical attributes. Decision trees are favored for their interpretability in categorization and decision-making tasks, their inherent myopic induction algorithms can result in suboptimal predictive performance and inherent biases. Decision Trees' simplicity and transparency make them a desirable choice for a variety of applications, particularly those requiring model interpretability. For example, in medical diagnosis, decision trees may be used to discover the key factors that influence a certain disease, allowing for more targeted therapy and better patient outcomes. Furthermore, decision trees may be utilized in marketing to determine the most successful advertising techniques and consumer categories, allowing for more focused and efficient marketing campaigns [19], [20].

2.3.5. Random forest

Is a stochastic technique that generates multiple decision trees (DTs) by utilizing a random vector to reduce correlations and enhance accuracy. Each DT is partitioned into a subset of features, with the number of attributes considered influencing the diversity of the tree. At each split, the optimal split function is determined to promote similarity among trees. The objective is to construct an ensemble of diverse decision trees for varied predictions [21], [22]

2.3.6. Logistic regression

It is a linear classifier that employs probabilities to assign data into binary categories. This approach is characterized by its simplicity and efficiency in data analysis, facilitating a straightforward interpretation of results. LR is predominantly utilized for binary classification tasks [21], [22]. The fundamental formulation of the LR model is represented as follows in (2):

$$\frac{Prob(Yi=1)}{Prob(Yi=0)} = \frac{Pi}{1-Pi} = e^{(\beta_0 + \beta_1 X_{1+} \cdots\cdots + \beta_k X_{ki})} \quad (2)$$

where e is the exponential constant, (1-Pi) is the chance that Y takes a value of 0, and Pi is the probability that Y takes a value of 1 [23], [24]

2.3.7. Multi-layer perceptron

MLP is a neural network deep learning characterized by a three-layer architecture, where each neuron is connected to every other neuron in the layer above. MLP leverages back-propagation and error gradient propagation methods for data transmission and training emphasis on error gradient propagation. It is known for generating high-quality models efficiently, yet necessitates a modular design for handling multiple output values [25]-[27].

2.4. WEKA tool

WEKA, a Java-based open-source ML toolkit developed by Waikato University, offers support for various learning methods, pre- and post-processing techniques, and data transformation methods. It is distributed under the GNU General Public License and is compatible with a range of devices. WEKA encompasses functionalities for data preprocessing, classification, regression, clustering, association rules, and visualization. WEKA includes features like the Explorer for accessing different tools, the Experimenter for comparing predictive performance of learning algorithms at scale, the knowledge flow interface for interactive arrangement of components like filters, classifiers, and evaluations, the workbench that integrates all other GUIs within WEKA, and the simple CLI for direct execution of WEKA commands [28], [29].

3. CLASSIFICATION AND RESULT IMPLEMENTATION

The 8th Roundabout in Amman serves as a junction linking four primary streets, each treated as an individual experimental scenario. The initial experiment involves three approaches originating from the Westbound Street (1, 2, and 3). Subsequently, the second experiment incorporates three approaches stemming from the Northbound Street (4, 5, 6). The third experiment encompasses three approaches originating from the Eastbound Street (7, 8, and 9), while the fourth focuses on a single approach originating from the airport (10).

3.1. Performance matrices

The effectiveness of the traffic congestion prediction model is evaluated using key metrics including true positive (TP), true negative (TN), false positive (FP), and false negative (FN) as illustrated in Table 1. Accuracy, precision, and recall are employed as techniques to summarize and assess the outcomes derived from the confusion matrix [30], [31]. Sample confusion matrix of FURIA using the WEKA interface is shown in Figure 3.

Table 1. Confusion matrix

		Predicted congested	Predicted uncongested
Actual	Congested	True Positive (TP)	False Negative (FN)
	Uncongested	False Positive (FP)	True Negative (TN)

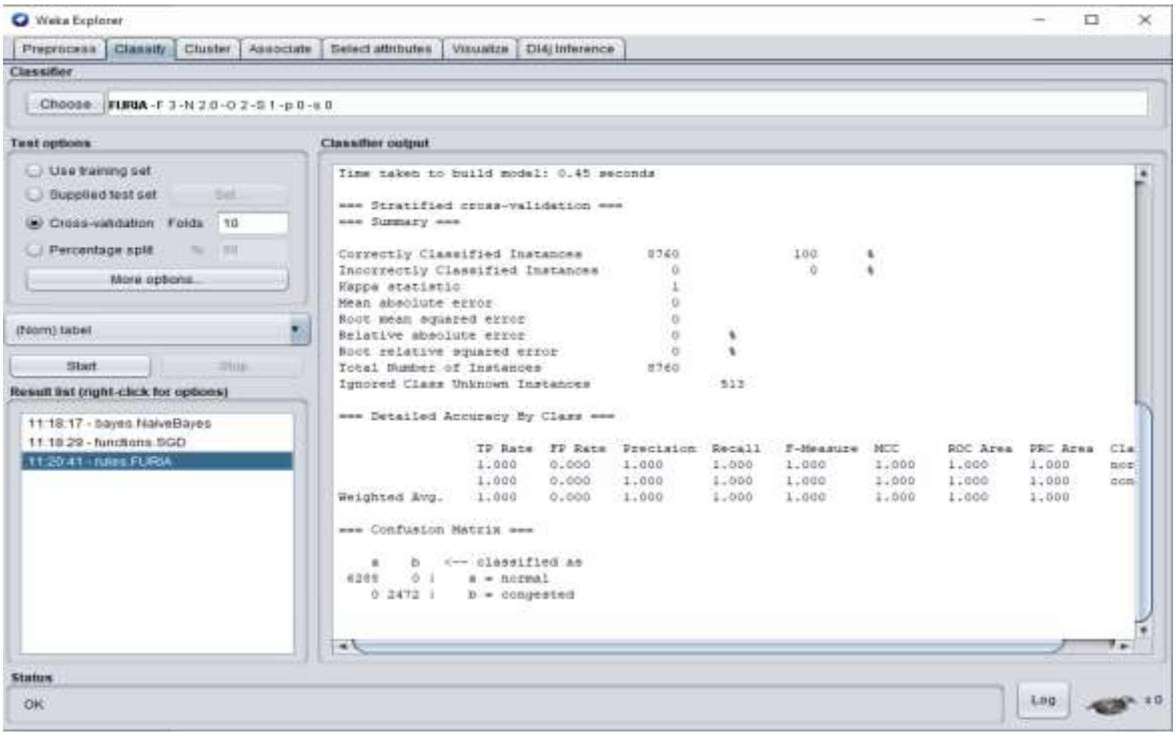

Figure 3. Result of FURIA confusion matrix using WEKA

The classification accuracy statement is a pivotal measure used to evaluate the appropriateness of a classification system for its designated task. Such accuracy statements find application in diverse scenarios, including classifier assessment, where significant focus is placed on detecting fluctuations in the precision of data classification. It is computed by dividing the number of correct predictions by the total number of predictions, as depicted in (3) [32], [33].

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Precision is calculated as the ratio of correctly identified positive instances (congested) to the total positive instances (congested or uncongested). Sensitivity, another evaluation metric utilized in this study, is determined by the ratio of correctly identified positive values to the total number of positive values. The sensitivity metric assesses how effectively the classifier can detect positive values. Lastly, the F-measure, employed in this paper for result evaluation, is computed by taking the harmonic mean of precision and sensitivity, assigning equal importance to each metric [34]. Table 2 presents the outcomes of the comparison for all experiment across various ML classifiers. Additionally, Figure 4 illustrates the accuracy representation of all classifiers for all experiment.

Table 2. Comparison of all classifiers for all experiments

	DT	MLP	RF	LR	NB	SGD	FURIA
Accuracy	97.20%	95.40%	96.20%	97.60%	87.54%	97.91%	100.00%
Precision	96.90%	98.40%	97.00%	97.70%	91.10%	97.90%	100.00%
Sensitivity	97.60%	92.50%	95.40%	97.50%	87.50%	97.90%	100.00%
F-measure	97.20%	95.30%	96.20%	97.60%	88.10%	97.90%	100.00%

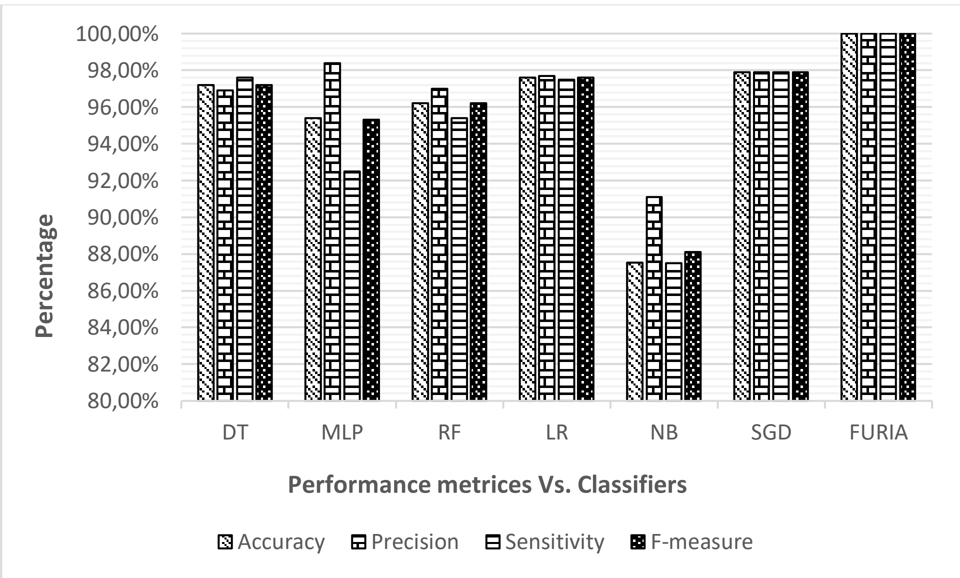

Figure 4. Comparison of all classifiers for all experiments

4. DISSCUSSION

Utilizing ML for traffic congestion prediction offers significant advantages in reducing time wastage, fuel consumption, and cost savings. This study employed NB, SGD, FURIA, LR, DT, RF, and MLP classifiers for prediction purposes. The results indicate that FURIA achieved the highest classification accuracy and precision score of 100%. Additionally, FURIA exhibited the highest sensitivity and F-measure ratings. By considering all metrics, FURIA demonstrated accurate predictions for both positive and negative classes, achieving 100% accuracy across all metrics. Furthermore, the accuracy presented in this study surpasses that of previous research efforts presented in Table 3.

Table 3. Comparison with other studies

#	Study	Accuracy rate
1	Majumdar *et al.* [6]	84–95%
2	Bai *et al.* [13]	87.5%
3	Moumen *et al.* [8]	91%
4	Najm *et al.* [9]	91%
5	Our study	100%

5. CONCLUSION

This research presents a thorough investigation of several classification methods for predicting traffic congestion in the 8th Roundabout in Amman City, Jordan. Four datasets, each with around 8,640 samples, were explored and processed separately in each street section. The data was provided by the municipal authorities of Greater Amman to predict traffic congestion in all intersections of the 8th Roundabout. The datasets were placed in WEKA mining software through which a confusion matrix was calculated to see which classifier would be the best for traffic congestion prediction. The following classifiers were applied in our research: NB, SGD, FURIA, LR, DT, RF, and MLP. Forefathers used accuracy, precision, and recall and F-measure for the classifiers: All classifiers' performance was evaluated using accuracy, precision, sensitivity, and the F-measure assessment criteria. The data revealed that FURIA was the top classifier for anticipating traffic congestion across all segments, scoring 100% accuracy across all assessment matrices.

REFERENCES

[1] M. R. Jabbarpour, H. Zarrabi, R. H. Khokhar, S. Shamshirband, and K. K. R. Choo, "Applications of computational intelligence in vehicle traffic congestion problem: a survey," *Soft Computing*, vol. 22, no. 7, pp. 2299–2320, Jan. 2018, doi: 10.1007/s00500-017-2492-z.

[2] H. Xiong, A. Vahedian, X. Zhou, Y. Li, and J. Luo, "Predicting traffic congestion propagation patterns: A propagation graph approach," in *IWCTS 2018 - Proceedings of the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, Nov. 2018, pp. 60–69, doi: 10.1145/3283207.3283213.

[3] T. Afrin and N. Yodo, "A survey of road traffic congestion measures towards a sustainable and resilient transportation system," *Sustainability (Switzerland)*, vol. 12, no. 11, p. 4660, Jun. 2020, doi: 10.3390/su12114660.

[4] S. Kwoczek, S. Di Martino, and W. Nejdl, "Predicting and visualizing traffic congestion in the presence of planned special events," *Journal of Visual Languages and Computing*, vol. 25, no. 6, pp. 973–980, Dec. 2014, doi: 10.1016/j.jvlc.2014.10.028.

[5] F. Lécué, R. Tucker, V. Bicer, P. Tommasi, S. Tallevi-Diotallevi, and M. Sbodio, "Predicting severity of road traffic congestion using semantic Web technologies," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8465 LNCS, Springer International Publishing, 2014, pp. 611–627.

[6] S. Majumdar, M. M. Subhani, B. Roullier, A. Anjum, and R. Zhu, "Congestion prediction for smart sustainable cities using IoT and machine learning approaches," *Sustainable Cities and Society*, vol. 64, p. 102500, Jan. 2021, doi: 10.1016/j.scs.2020.102500.

[7] M. Akhtar and S. Moridpour, "A review of traffic congestion prediction using artificial intelligence," *Journal of Advanced Transportation*, vol. 2021, pp. 1–18, Jan. 2021, doi: 10.1155/2021/8878011.

[8] I. Moumen, J. Abouchabaka, and N. Rafalia, "Enhancing urban mobility: integration of IoT road traffic data and artificial intelligence in smart city environment," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 985–993, Nov. 2023, doi: 10.11591/ijeecs.v32.i2.pp985-993.

[9] I. A. Najm, A. K. Hamoud, J. Lloret, and I. Bosch, "Machine learning prediction approach to enhance congestion control in 5G IoT environment," *Electronics (Switzerland)*, vol. 8, no. 6, p. 607, May 2019, doi: 10.3390/electronics8060607.

[10] X. Di, Y. Xiao, C. Zhu, Y. Deng, Q. Zhao, and W. Rao, "Traffic congestion prediction by spatiotemporal propagation patterns," in *Proceedings - IEEE International Conference on Mobile Data Management*, Jun. 2019, vol. 2019-June, pp. 298–303, doi: 10.1109/MDM.2019.00-45.

[11] T. Ito and R. Kaneyasu, "Predicting traffic congestion using driver behavior," *Procedia Computer Science*, vol. 112, pp. 1288–1297, 2017, doi: 10.1016/j.procs.2017.08.090.

[12] Y. Liu and H. Wu, "Prediction of road traffic congestion based on random forest," in *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, Dec. 2018, vol. 2, pp. 361–364, doi: 10.1109/ISCID.2017.216.

[13] M. Bai, Y. Lin, M. Ma, P. Wang, and L. Duan, "PrePCT: Traffic congestion prediction in smart cities with relative position congestion tensor," *Neurocomputing*, vol. 444, pp. 147–157, Jul. 2021, doi: 10.1016/j.neucom.2020.08.075.

[14] F. S. Alenazi, K. El Hindi, and B. AsSadhan, "Complement-class harmonized naïve bayes classifier," *Applied Sciences (Switzerland)*, vol. 13, no. 8, p. 4852, Apr. 2023, doi: 10.3390/app13084852.

[15] A. Angelopoulos *et al.*, "Impact of classifiers to drift detection method: a comparison," in *Proceedings of the 22nd Engineering Applications of Neural Networks Conference*, Springer International Publishing, 2021, pp. 399–410.

[16] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 4, pp. 2715–2743, 2021.

[17] J. Hühn and E. Hüllermeier, "FURIA: An algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, Apr. 2009, doi: 10.1007/s10618-009-0131-8.

[18] A. Alzayed, W. Almayyan, and A. Al-Hunaiyyan, "Diagnosis of obesity level based on bagging ensemble classifier and feature selection methods," *International Journal of Artificial Intelligence & Applications*, vol. 13, no. 02, pp. 37–54, Mar. 2022, doi: 10.5121/ijaia.2022.13203.

[19] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[20] O. Sagi and L. Rokach, "Approximating XGBoost with an interpretable decision tree," *Information Sciences*, vol. 572, pp. 522–542, Sep. 2021, doi: 10.1016/j.ins.2021.05.055.

[21] M. G. Raj, C. Pradip, N. Saju, and S. V. Tresa Sangeetha, "Random forest-based method for micro grid system in energy consumption prediction," *Journal of Physics: Conference Series*, vol. 1964, no. 5, p. 52002, Jul. 2021, doi: 10.1088/1742-6596/1964/5/052002.

[22] P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine learning technique to prognosis diabetes disease: random forest classifier approach," in *Lecture Notes in Networks and Systems*, vol. 218, Springer Singapore, 2022, pp. 219–244.

[23] M. Hossain *et al.*, "A predictive logistic regression model for periodontal diseases," *Saudi Journal of Oral Sciences*, vol. 8, no. 3, p. 150, 2021, doi: 10.4103/sjoralsci.sjoralsci_123_20.

[24] E. Y. Boateng and D. A. Abaye, "A review of the logistic regression model with emphasis on medical research," *Journal of Data Analysis and Information Processing*, vol. 07, no. 04, pp. 190–207, 2019, doi: 10.4236/jdaip.2019.74012.

[25] A. Guezzaz, A. Asimi, Y. Asimi, Z. Tbatou, and Y. Sadqi, "A global intrusion detection system using pcapsocks sniffer and multilayer perceptron classifier," *International Journal of Network Security*, vol. 21, no. 3, pp. 438–450, 2019.

[26] Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak, "Modeling the spread of COVID-19 infection using a multilayer perceptron," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–10, May 2020, doi: 10.1155/2020/5714714.

[27] P. Mohapatra *et al.*, "Artificial neural network based prediction and optimization of centelloside content in Centella asiatica: A comparison between multilayer perceptron (MLP) and radial basis function (RBF) algorithms for soil and climatic parameter," *South African Journal of Botany*, vol. 160, pp. 571–585, Sep. 2023, doi: 10.1016/j.sajb.2023.07.019.

[28] B. Saleh, A. Saedi, A. Al-Aqbi, and L. Salman, "A review paper: analysis of weka data mining techniques for heart disease prediction system," *International journal of medical reviews*, vol. 7, no. 1, pp. 15–24, 2020.

[29] K. P. S. Attwal and A. S. Dhiman, "Exploring data mining tool-Weka and using Weka to build and evaluate predictive models," *Advances and Applications in Mathematical Sciences*, vol. 19, no. 6, pp. 451–469, 2020.

[30] B. J. Erickson and F. Kitamura, "Magician's corner: 9. performance metrics for machine learning models," *Radiology: Artificial Intelligence*, vol. 3, no. 3, p. e200126, May 2021, doi: 10.1148/ryai.2021200126.

[31] O. Almomani, A. Alsaaidah, A. A. A. Shareha, A. Alzaqebah, and M. Almomani, "Performance evaluation of machine learning classifiers for predicting denial-of-service attack in internet of things," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 263–271, 2024, doi: 10.14569/IJACSA.2024.0150125.

[32] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label classifier performance evaluation with confusion matrix," in *Computer Science & Information Technology*, Jun. 2020, pp. 01–14, doi: 10.5121/csit.2020.100801.

[33] J. R. M. Navin and R. Pankaja, "Performance analysis of text classification algorithms using confusion matrix," *International Journal of Engineering and Technical Research (IJETR)*, vol. 6, no. 4, 2016.

[34] H. Yun, "Prediction model of algal blooms using logistic regression and confusion matrix," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2407–2413, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2407-2413.

BIOGRAPHIES OF AUTHORS

Mohammad Hassan has completed his Ph.D. from Baku State University, Azerbaijan. He is an Associate Professor in the Computer Engineering Department at the Faculty of Engineering at Al-Ahliyya Amman University. Hassan is a member of the Jordanian Engineering Association. He has published numerous research papers in various journals and conferences, covering topics such as machine learning, computer networks, intelligent transportation systems, and mobile learning adaptation models. He can be contacted at email: mhassan@ammanu.edu.jo.

Areen Arabiat obtained B.Sc. in Computer Engineering in 2005 from al Balqa Applied University (BAU), and her M.Sc. in Intelligent Transportation Systems (ITS) from Al Ahliyya Amman University (AAU) in 2022. She is currently a computer lab supervisor at the Faculty of Engineering/Al-Ahliyya Amman University (AAU) since 2013. Her research interests are focused on the following areas: machine learning, data mining, artificial intelligence and image processing. Email: a.arabiat@ammanu.edu.jo.