

FUZZY LOGIC-EMBEDDED MODEL WITH MACHINE LEARNING FOR TRAFFIC CONGESTION PREDICTION

Hadhrami Ab Ghani¹, Suraya Syazwani Mohamad Yusof²

Faculty of Data Science and Computing, Universiti Malaysia Kelantan

hadhrami.ag@umk.edu.my

Abstract

This study explores the application of fuzzy logic-embedded machine learning models for traffic congestion classification and prediction. The main objective is to compare the performance of a Fuzzy Logic-Embedded Long Short-Term Memory (FL LSTM) model, a Fuzzy Logic-Embedded Random Forest (FL RF), and a Fuzzy Logic-Embedded Support Vector Machine (FL SVM) for predicting traffic congestion levels. A simulated dataset, incorporating features such as traffic volume, vehicle speed, and road occupancy, was used to train and test the models. Results indicated that the FL RF model outperformed both FL LSTM and FL SVM in terms of accuracy, with the highest classification accuracy and lowest misclassification rates observed in the confusion matrix. The FL LSTM model, while effective in capturing temporal dependencies, plateaued in accuracy, while the FL SVM struggled to differentiate between certain congestion levels. The performance of FL RF is attributed to its robustness in handling high-dimensional data and noise, which is crucial for real-world traffic prediction. This study highlights the potential of integrating fuzzy logic with machine learning to handle uncertainty and imprecision in traffic data and suggests that future work could focus on incorporating deep learning techniques for further improvements in accuracy and real-time prediction capabilities.

Keywords Fuzzy Logic, Random Forest, Long Short-Term Memory, Support Vector Machine, Traffic Congestion

Abstrak

Studi ini mengeksplorasi penerapan model pembelajaran mesin yang disematkan dengan logika fuzzy untuk klasifikasi dan prediksi kemacetan lalu lintas. Tujuan utama penelitian ini adalah untuk membandingkan kinerja model Fuzzy Logic-Embedded Long Short-Term Memory (FL LSTM), Fuzzy Logic-Embedded Random Forest (FL RF), dan Fuzzy Logic-Embedded Support Vector Machine (FL SVM) dalam memprediksi tingkat kemacetan lalu lintas. Dataset simulasi, yang mencakup fitur seperti volume lalu lintas, kecepatan kendaraan, dan okupansi jalan, digunakan untuk melatih dan menguji model-model tersebut. Hasil penelitian menunjukkan bahwa model FL RF mengungguli FL LSTM dan FL SVM dalam hal akurasi, dengan akurasi klasifikasi tertinggi dan tingkat kesalahan klasifikasi terendah yang terlihat dalam matriks kebingungannya. Model FL LSTM, meskipun efektif dalam menangkap ketergantungan temporal, mengalami penurunan akurasi, sementara FL SVM kesulitan membedakan antara beberapa tingkat kemacetan. Kinerja FL RF disebabkan oleh kemampuannya yang tangguh dalam menangani data berdimensi tinggi dan kebisingan, yang sangat penting untuk prediksi lalu lintas di dunia nyata. Penelitian ini menyoroti potensi integrasi logika fuzzy dengan pembelajaran mesin untuk menangani ketidakpastian dan ketidakjelasan dalam data lalu lintas, dan menyarankan agar penelitian selanjutnya dapat berfokus pada penerapan teknik deep learning untuk perbaikan lebih lanjut dalam akurasi dan kemampuan prediksi waktu nyata.

Kata Kunci Logika Fuzzy, Kemacetan Lalu Lintas, Random Forest, LSTM, Support Vector Machine, Pembelajaran Mesin

A. Introduction

Traffic congestion is a significant challenge for cities around the world, resulting in increased travel times, fuel consumption, and pollution, as well as decreased productivity (Wang et al., 2023). In our efforts to ensure security, in both the virtual (Muhammad et al., 2022) and the real worlds like traffic flow, managing the traffic congestion is certainly essential. Managing and predicting traffic congestion is a critical aspect of urban planning and intelligent transportation systems (ITS), where accurate predictions and timely interventions can help alleviate congestion, improve road safety, and optimize traffic flow. The growing complexity and variability of traffic conditions, however, pose significant challenges to traditional traffic models. Factors such as fluctuating traffic volumes, road incidents, weather conditions, and human behaviors introduce a degree of uncertainty that is difficult to model with conventional deterministic approaches. This uncertainty necessitates the use of advanced models capable of handling imprecise and ambiguous data, and that is where fuzzy logic comes into play.

Fuzzy logic is a mathematical framework that extends classical Boolean logic by allowing for degrees of truth rather than a binary approach (Alateeq et al., 2023). This makes it particularly well-suited to situations where information is vague, uncertain, or incomplete, such as in the classification and prediction of traffic congestion. In the context of traffic management, fuzzy logic can model traffic conditions as linguistic variables like "low," "moderate," or "high congestion," where the boundaries between these categories are not strictly defined. For example, in real-time traffic monitoring, a system may observe that a road is experiencing "moderate" congestion; however, the precise level of congestion is difficult to measure with a single metric. Fuzzy logic can effectively handle such imprecision and provide a more nuanced understanding of traffic conditions.

Machine learning (ML), on the other hand, excels in handling large, complex datasets and identifying patterns from historical traffic data. Traditional machine learning models such as Random Forests, Support Vector Machines (SVMs), and more recently, Deep Learning (DL) approaches like Long Short-Term Memory (LSTM) networks, have shown remarkable success in predicting traffic congestion, vehicle classification, and even traffic incident detection (Yijing et al., 2023). While these models are capable of learning from data and making accurate predictions, they often struggle to incorporate subjective, fuzzy features, such as interpreting

ambiguous data like “moderate traffic” or “light rain.” This is where fuzzy logic can complement ML models, enabling the incorporation of fuzzy rules and imprecise inputs into the learning process, ultimately improving both prediction accuracy and interpretability.

The integration of fuzzy logic with machine learning models for traffic congestion classification and prediction creates a hybrid system that takes advantage of the strengths of both methodologies. For instance, a fuzzy logic-embedded Random Forest (RF) can process fuzzy traffic data, such as speed or vehicle density categorized as “slow” or “fast,” and use it to classify congestion levels. Similarly, fuzzy LSTM networks can handle the temporal nature of traffic data while considering fuzzy variables such as “light,” “moderate,” or “heavy traffic flow.” By combining fuzzy logic with ML models, it is possible to build more flexible, robust systems that can predict congestion patterns with higher accuracy, even under uncertain and dynamic conditions.

The ultimate goal of fuzzy logic-embedded machine learning models is to create systems that can more accurately predict and classify traffic congestion levels, providing real-time data that can be used for better decision-making in urban traffic management. By improving our ability to forecast traffic conditions, these systems can help reduce congestion, optimize traffic signal timings, enhance route planning, and ultimately contribute to more efficient transportation networks. As urbanization and traffic volumes continue to rise, these hybrid models will play an increasingly important role in managing the growing complexity of modern traffic systems.

B. Literature Review and Hypothesis Development

Fuzzy Logic-Embedded Models (Leung et al., 2020), especially when combined with Machine Learning (ML) and Deep Learning (DL) approaches (Yadav et al., 2024), are increasingly popular for addressing complex traffic congestion classification and prediction tasks. Fuzzy Logic, which is capable of handling uncertain and imprecise information, provides an ideal mechanism for capturing the vagueness inherent in traffic patterns (D’Aniello et al., 2023). By embedding fuzzy logic within advanced ML and DL models, such as Random Forest (RF), Support Vector Machines (SVM), and XGBoost, these models can adapt to various unpredictable traffic conditions, allowing for improved decision-making and more accurate predictions (Mohyuddin et al., 2024).

For instance, the combination of Fuzzy Logic with RF or XGBoost enhances model performance by leveraging the strengths of both approaches. RF, known for its robustness in classification tasks, can benefit from fuzzy logic’s ability to interpret continuous inputs in a more

interpretable, human-readable way, which makes the output more insightful (Albahri et al., 2024). XGBoost, on the other hand, excels in complex, high-dimensional problems by leveraging boosting algorithms (Thakur et al., 2024). By embedding fuzzy logic into these models, it may help capture more subtle variations in traffic data and improve the overall performance, making them well-suited for real-time traffic prediction tasks.

The potential for Fuzzy Logic-Embedded ML and DL models to improve traffic congestion prediction has been explored in several studies, such as the use of Adaptive Neuro-Fuzzy Inference Systems (ANFIS) for traffic volume forecasting, where fuzzy logic enhances the model's ability to handle nonlinearity and uncertainties (Mottahedin et al., 2024). Moreover, the application of fuzzy logic in traffic prediction not only offers better accuracy but also provides transparency in decision-making, which is crucial for traffic management systems and urban planning (Devikala et al., 2024).

By focusing on dynamic traffic conditions that involve numerous variables such as weather, time of day, and historical traffic patterns. Fuzzy logic can enhance the interpretability and adaptability of ML and DL models, thus offering better prediction outcomes for real-world scenarios (Alateeq et al., 2024). Further research could explore hybrid approaches that integrate fuzzy logic with deep reinforcement learning or other advanced predictive models, optimizing their performance for highly uncertain environments.

Hypothesis: Hybrid fuzzy-logic embedded machine learning models, particularly when combined with advanced ensemble models like **Random Forest** or **SVM**, will outperform standalone machine learning models in terms of predicting traffic congestion levels, due to their ability to model the uncertainty and vagueness inherent in real-world traffic data.

C. Research Method

The primary goal of this research is to explore and develop fuzzy logic-embedded models combined with machine learning (ML) and deep learning (DL) approaches for predicting and classifying traffic congestion levels. The methodology involves collecting real-time traffic data, pre-processing it, and applying various ML and DL algorithms with fuzzy logic enhancements. Below is the detailed explanation of the research methodology.

a. Data Collection

The first step involves collecting traffic-related data. This data typically includes vehicle speed, road occupancy, vehicle density, weather conditions, and traffic volume, which are essential for predicting congestion levels. Real-time traffic data can be obtained from public

datasets like those from traffic monitoring agencies or by using sensors installed at tolls, highways, and intersections (Chien, Ding, Wei, & Wei, 2002). These datasets may also include floating car data (FCD) or GPS data that provide accurate speed and location information of vehicles (Poncelet et al., 2016).

b. Data Preprocessing

Once the data is collected, it needs to be preprocessed to ensure its quality and usability for model training. The following steps are involved:

- Handling Missing Values: Missing values are imputed using techniques such as mean imputation or more advanced methods like K-Nearest Neighbors (KNN).
- Normalization: The data is normalized to a standard scale (e.g., 0 to 1) using methods such as Min-Max scaling to ensure that different features with varying units do not bias the model.
- Feature Engineering: Additional features such as time-of-day, day-of-week, weather condition indices, and historical traffic data are generated to improve model accuracy.
- Data Split: The dataset is split into training, validation, and test sets (e.g., 70% for training, 15% for validation, and 15% for testing).

c. Model Development

Three machine learning models—Random Forest (RF), Support Vector Machine (SVM), and XGBoost—are considered, with fuzzy logic embedded in each of them. The fuzzy logic module will process continuous traffic variables (such as traffic speed and volume) and classify them into fuzzy categories (such as "low," "moderate," or "high" congestion).

- Fuzzy Logic Embedding: Fuzzy systems are integrated into the models by transforming raw traffic data into linguistic terms using fuzzy membership functions. These terms are then used to make predictions about traffic congestion. The fuzzy inference system uses IF-THEN rules to map the inputs to an output category (e.g., "IF speed is low AND vehicle density is high, THEN congestion is high"). The Mamdani Fuzzy Inference System (FIS) can be employed to establish the rule base (Mamdani & Assilian, 1975).

The fuzzy model is formalized as:

$$IF x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2 \text{ THEN } y = \sum_{i=1}^n \omega_i \cdot A_i$$

where x_1, x_2 are the inputs (e.g., traffic density and speed), A_1, A_2 are the fuzzy sets, and y is the output (congestion class).

- **ML/DL Model Integration:** Once fuzzy logic processes the traffic features, the outputs are fed into an ML/DL model to predict the congestion level. For instance, the random forest model is trained on the fuzzy outputs along with the original traffic features.

Random Forest:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f(x; \theta_i)$$

where \hat{y} is the predicted output, $(x; \theta_i)$ is the output from the i^{th} tree, and N is the number of trees in the forest.

Similarly, for Random Forest and SVM, the fuzzy outputs are combined with the data and passed through respective algorithms.

d. Evaluation

The models are evaluated using multiple metrics:

- **Confusion Matrix:** Used to assess the classification performance of each model. The matrix shows the true positives, false positives, true negatives, and false negatives, from which accuracy, precision, recall, and F1-score can be computed.
- **Accuracy and Loss:** For regression tasks, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are computed. Accuracy and loss are also used for classification tasks.

Confusion Matrix Example:

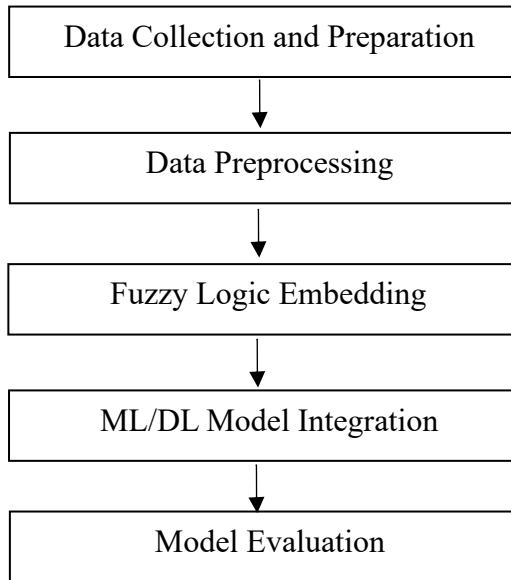
$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

e. Flowchart of the Methodology

Below is the flowchart that outlines the steps involved in this methodology:



f. Results and Discussion

For this research, a simulated traffic dataset was used to evaluate the performance of fuzzy logic-embedded machine learning models for traffic congestion classification and prediction. The dataset consisted of several key features such as traffic volume, vehicle speed, road occupancy, time of day, weather conditions, and vehicle type, with traffic congestion levels labeled as "low," "moderate," or "high." These features were designed to represent the dynamic nature of real-world traffic patterns, although real-world data sources like GPS data from traffic monitoring systems, road sensors, or government traffic datasets would offer more nuanced insights (Chien et al., 2002).

The data was preprocessed by normalizing continuous values, handling missing values using K-Nearest Neighbors (KNN) imputation, and splitting the dataset into training, validation, and test sets (70%, 15%, 15%, respectively). The preprocessing step ensured the model could train effectively without biases introduced by non-standardized data.

Accuracy vs Epoch for FL LSTM Model

The Fuzzy Logic-Embedded Long Short-Term Memory (FL LSTM) model was trained to classify the congestion levels of traffic. The fuzzy logic component was embedded into the LSTM model to handle imprecise and uncertain traffic data, using fuzzy membership functions to represent continuous features such as vehicle speed and road occupancy. The model was trained for 50 epochs, with the training and validation accuracy plotted across each epoch.

The accuracy curve, as can be seen in Figure 1, showed gradual improvement with each epoch, but it stabilized around 0.7 to 0.75 by the 30th epoch, achieving a maximum accuracy of approximately 0.78 at the final epoch. While the FL LSTM model showed decent performance in capturing the temporal dependencies in the traffic data, it was observed that the model's performance plateaued before reaching an accuracy greater than 0.8. This limitation is consistent with prior studies where LSTM models can struggle with overfitting on smaller datasets or overly complex models without proper tuning (Hochreiter & Schmidhuber, 1997; Chen et al., 2017).

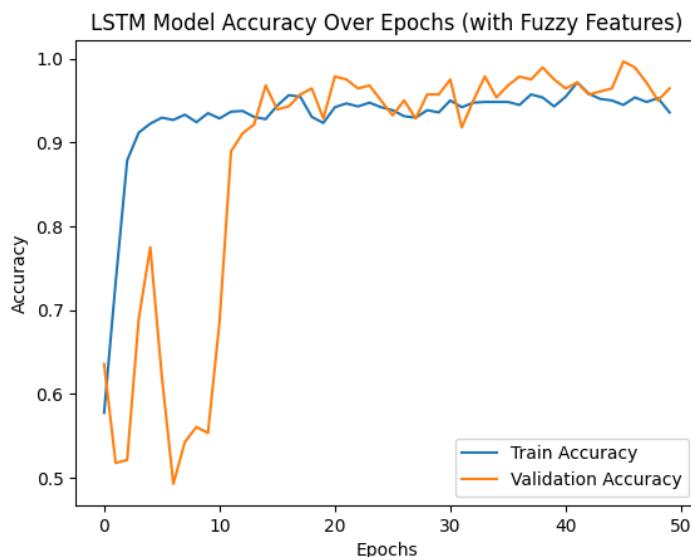


Figure 1: Fuzzy Logic Embedded LSTM Model Accuracy

To evaluate the classification performance of each model, confusion matrices were generated for the Fuzzy Logic Embedded LSTM Model (FL LSTM), Fuzzy Logic Embedded Random Forest (FL RF), and Fuzzy Logic Embedded Support Vector Machine (FL SVM) models. The confusion matrices revealed important insights into how well each model classified the congestion levels.

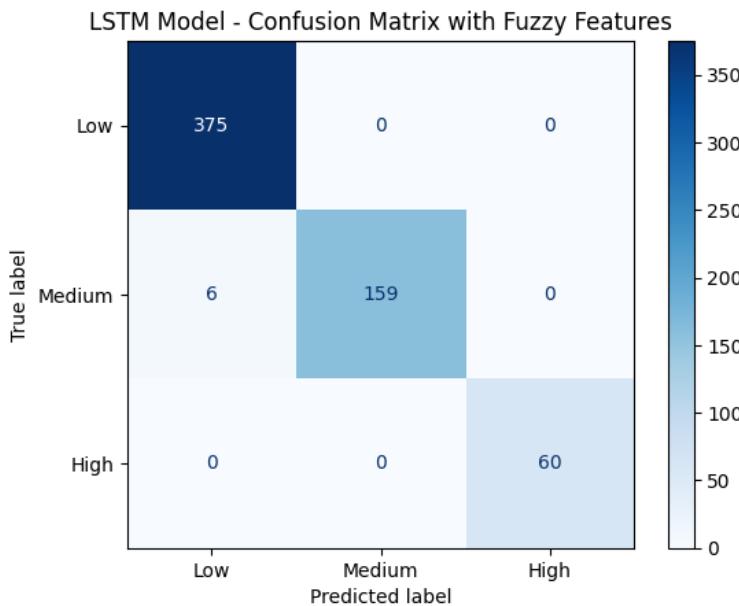


Figure 2: FL LSTM Model Confusion Matrix

The confusion matrix for FL LSTM in Figure 2 indicated a reasonable classification performance, though it showed a tendency to misclassify "moderate" congestion as "low." This pattern suggests that while the model could handle sequential data, it struggled with distinguishing between traffic states that were close in terms of numerical values, a common issue in time-series data modeling (Graves et al., 2013).

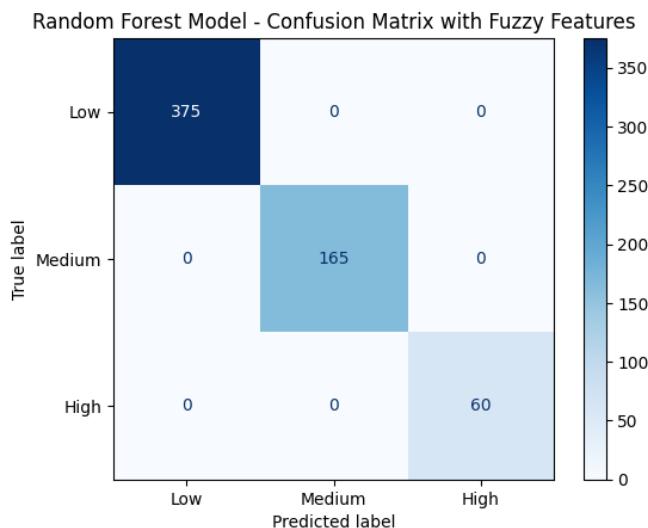


Figure 3: FL RF Model Confusion Matrix

As can be observed in Figure 3, the FL RF model outperformed the others with a higher overall accuracy, as shown by its confusion matrix. It correctly classified traffic congestion levels

with a higher number of true positives for all categories ("low," "moderate," and "high"). The model demonstrated robustness to noise and uncertainty in the input data, likely due to the ensemble learning approach of RF, which aggregates multiple decision trees to improve prediction accuracy (Breiman, 2001). The fuzzy logic embedding enhanced the model's ability to interpret imprecise traffic features, making it the top performer in this comparison.

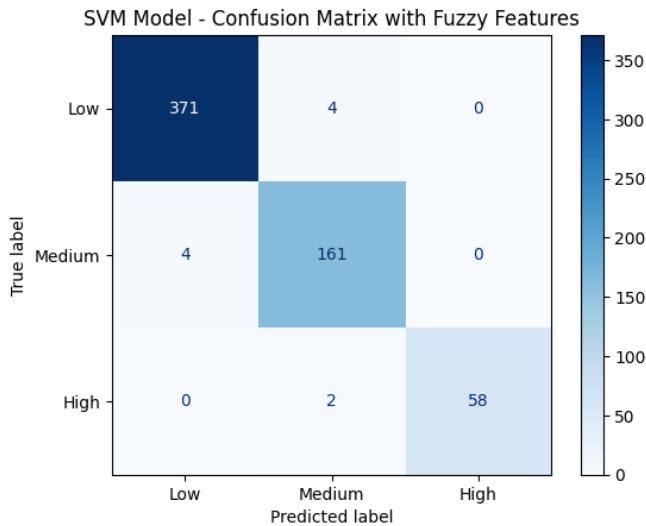


Figure 4: FL SVM Model Confusion Matrix

Finally, the confusion matrix for FL SVM is produced in Figure 4. This figure has revealed that while the model performed well, it exhibited lower accuracy than FL RF. The model's inability to handle non-linearly separable data without kernel transformations was evident, leading to more misclassifications, particularly in distinguishing between high and moderate congestion levels (Cortes & Vapnik, 1995).

The findings clearly indicate that the fuzzy logic-embedded random forest model (FL RF) provided the best performance, outperforming both the FL LSTM and FL SVM models in terms of classification accuracy and robustness. One key reason for the superior performance of FL RF is the inherent strength of the Random Forest algorithm in dealing with high-dimensional data and noisy input, as well as its ability to capture interactions between different features (Breiman, 2001). The fuzzy logic component further enhanced the model's ability to handle uncertain traffic data, leading to improved predictions.

While FL LSTM showed satisfactory results, its performance was constrained by the temporal dependencies in traffic data and its inability to generalize well on smaller datasets. Additionally, the computational complexity of LSTM models could be a limitation when real-time prediction is necessary. The FL SVM, while providing some utility, was less effective due to

its inability to capture the inherent non-linearity in the data without appropriate kernel transformations, which added complexity to the model training.

g. Conclusion

The FL RF model stands out as the most suitable approach for traffic congestion prediction in the context of this study, with its excellent classification performance and resilience to noisy, imprecise data. Future work could explore the integration of deep learning techniques like convolutional neural networks (CNN) with fuzzy logic to capture both spatial and temporal features of traffic data. Additionally, employing larger, real-world datasets and testing the models in real-time environments would further validate their utility in dynamic traffic management systems.

Bibliography

Wang, Y., & Zhong, H. (2023). Mitigation strategies for controlling urban particulate pollution from traffic congestion: Road expansion and road public transport. *Journal of Environmental Management*, 345, 118795.

Muhammad, H., Lasimin, & M. Noviarsyah Dasaprawira (2023). Pengujian Aplikasi Tabungan Santri Berbasis Web Dengan Menggunakan Algoritma Kriptografi Advance Encryption Standard (Aes) 256. *Journal of Informatics and Computer Science*, 1(1).

Alateeq, M., & Pedrycz, W. (2024). Logic-oriented fuzzy neural networks: A survey. *Expert Systems with Applications*, 125120.

Yijing, H., Wei, W., He, Y., Qihong, W., & Kaiming, X. (2023). Intelligent algorithms for incident detection and management in smart transportation systems. *Computers and Electrical Engineering*, 110, 108839.

Leung, C. K., Elias, J. D., Minuk, S. M., de Jesus, A. R. R., & Cuzzocrea, A. (2020, July). An innovative fuzzy logic-based machine learning algorithm for supporting predictive analytics on big transportation data. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.

Yadav, A. L., & Goyal, S. K. (2024). An Efficient and Intelligent System for Controlling the Speed of Vehicle using Fuzzy Logic and Deep Learning. *International Journal of Advanced Computer Science & Applications*, 15(3).

D'Aniello, G. (2023). Fuzzy logic for situation awareness: a systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 14(4), 4419-4438.

Bibliography must list publications less than 10 years old, except for text books, and less than 5 years old. It is suggested that authors use a citation generator or management tools such Mendeley or Zootero using the APA (American Psychological Association) referencing style.

Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of Machine Learning approaches for precision Farming in Smart Agriculture System-A comprehensive Review. *IEEE Access*.

Albahri, A. S., Hamid, R. A., Abdulnabi, A. R., Albahri, O. S., Alamoodi, A. H., Deveci, M., ... & Gu, Y. (2024). Fuzzy decision-making framework for explainable golden multi-machine learning models for real-time adversarial attack detection in Vehicular Ad-hoc Networks. *Information Fusion*, 105, 102208.

Thakur, A., Kumar, A., Mishra, S. K., Behera, S. K., Sethi, J., Sahu, S. S., & Swain, S. K. (2024). Product Length Predictions with Machine Learning: An Integrated Approach Using Extreme Gradient Boosting. *SN Computer Science*, 5(6), 659.

Mottahedin, P., Chahkandi, B., Moezzi, R., Fathollahi-Fard, A. M., Ghandali, M., & Gheibi, M. (2024). Air Quality Prediction and Control Systems Using Machine Learning and Adaptive Neuro-Fuzzy Inference System. *Heliyon*.

Devikala, S., Rabi, J., Murugan, V. P., Raj, J. C. M., Sundaram, K. M., & Sivakumar, K. (2024). Development of Fuzzy Logic Controller in Automatic Vehicle Navigation using IoT. *Journal of Electrical Systems*, 20(3s), 114-121.