Infact: International Journal of Computers
Vol.9, No.01, Maret 2025
p-ISSN : 2527-8363, e-ISSN: 2829-0259
DOI:

Integration of Federated Learning in Big Data Analytics for IoT-based Intelligent Transportation System

Budiman Wijaya[1*], Irma Puri Rahayu[2], Heri Wijayanto[3]
[1,2,3] Master of Information Technology, Faculty of Engineering, University of Mataram
[1,2,3]Jln. Majapahit No. 62 Mataram

E-mail: bwijaya.sgt76@gmail.com[1], irmaputri2808@gmail.com[2], heri@unram.ac.id[3]

Submitted: 27/01/2025 , Revision: 27/02/2025, Accepted : 02/03/2025

Abstract

This article examines the integration of Federated Learning (FL) into big data analytics for intelligent transportation systems based on the Internet of Things (IoT). FL enables distributed machine learning model training without transferring sensitive data to a central server, preserving privacy and reducing data breach risks.
The literature review highlights three key studies. The first demonstrates how FL improves traffic prediction accuracy using data from various sources, including vehicles and environmental sensors. The second introduces a big data architecture that integrates FL for real-time analysis and decision-making. The third emphasizes FL's role in sustainable traffic management, reducing congestion and carbon emissions through data-driven solutions.
This article identifies research gaps and offers recommendations for optimizing FL in big data analytics, aiming to enhance efficiency, safety, and sustainability in modern transportation systems.

Keywords:
Federated Learning (FL);
Big Data Analytics;
Intelligent Transportation System;
Internet of Things (IoT)
…

***Corresponding Author: Budiman Wijaya**
E-mail: bwijaya.sgt76@gmail.com

1. Introduction

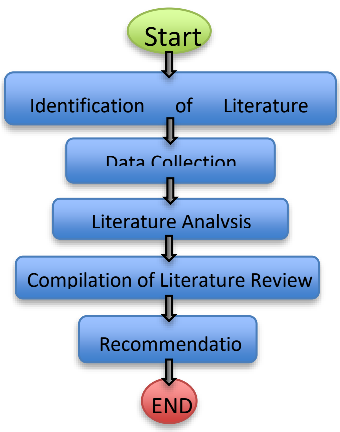
The development of Internet of Things (IoT) and big data technologies has revolutionized various sectors, including transportation systems[1]. With the increasing number of vehicles and the complexity of transportation infrastructure, challenges in traffic management, safety, and operational efficiency are increasing[2]. IoT enables real-time data collection from various sources, such as vehicles, road sensors, and mobile devices, creating opportunities for more in-depth analysis and better decision-making[3]. However, with large volumes of diverse data, innovative analytics approaches are needed to process and utilize this information effectively[4].

One promising approach is Federated Learning (FL), a machine learning method that enables distributed model training on user devices without moving data to a central server[5]. FL not only maintains data privacy, but also reduces latency and improves bandwidth usage efficiency[6]. In the context of intelligent transportation, FL can be used to predict traffic, optimize routes, and improve safety by collaboratively analyzing data generated by vehicles and infrastructure"[7]. For example, a study by Smith et al. (2021) demonstrated how FL-based models improved traffic congestion prediction in metropolitan areas, while another study by Lee et al. (2022) highlighted the role of FL in optimizing public transportation scheduling.

This article aims to explore the integration of FL in big data analytics for IoT-based intelligent transportation systems. This literature review will analyze three key papers that address FL applications in traffic prediction, big data analytics architecture development, and sustainable traffic management in smart cities. By identifying existing challenges and opportunities, this article is expected to provide valuable insights for researchers, practitioners, and policy makers in designing more efficient, safe, and sustainable transportation systems in the digital era.

2. Methods

This paper uses a literature review approach to explore the integration of Federated Learning (FL) in big data analytics for Internet of Things (IoT)-based intelligent transportation systems. The methodology consists of several detailed steps, ensuring a structured and comprehensive review process as shown in the flow chart below.



2.1 Identification of Literature Sources

The first step in the methodology was to identify and select literature sources relevant to the topic at hand[8]. This research focuses on papers that explore FL applications in the context of intelligent transportation, with the following selection criteria:

- Papers published in leading journals and international conferences.
- Studies covering aspects of IoT technology, big data, and traffic management.
- Research that demonstrates practical applications or methodological developments in FL for transportation systems.
- Literature was sourced from databases such as IEEE Xplore, Google Scholar, ACM Digital Library, SpringerLink, and ScienceDirect, using keywords like "Federated Learning in Transportation," "IoT-based Traffic Management," and "Big Data in Intelligent Transport."

2.2 Data Collection

Once the literature sources were identified, the next step was to collect data from each paper[9]. The data collected included:

- Research objectives and proposed hypotheses.
- Methodology used in the research.
- Key results and findings obtained from each study.
- Recommendations and implications of the research.

To improve reproducibility, data was extracted and organized using the reference management tool Mendeley.

2.3 Literature Analysis

Once the data was collected, the next step was to analyze the information obtained from each paper[10]. This process includes:

- Categorize the findings based on key themes, such as traffic prediction, big data analytics architecture, and sustainable traffic management.
- Identify similarities and differences between the approaches taken in each study.
- Assess the contribution of each study to the understanding of FL in the context of intelligent transportation systems.

Comparative analysis was applied to assess the advantages and limitations of different approaches, highlighting gaps and areas for further research.

2.4 Compilation of Literature Review

Once the analysis is complete, the information will be organized into a structured literature review[11]. This review will include:

- Summary of each paper reviewed.
- The relationship of the papers to each other to show the development of thinking in this area.
- Emphasis on challenges faced in the application of FL and opportunities for further research.

2.5 Recommendation

Based on the findings from the literature review, this study can provide recommendations for future research, which include suggestions for the development of more adaptive FL algorithms, as well as the need for collaboration between various parties in the transportation ecosystem to optimize the use of this technology.

The literature review methodology applied in this paper aims to provide a comprehensive understanding of the integration of Federated Learning in big data analytics for IoT-based intelligent transportation systems. With this approach, it is expected to identify relevant challenges, opportunities, and future research directions in this context.

2.6 Algorithm

The following algorithm illustrates the integration process of Federated Learning (FL) in big data analytics for IoT-based intelligent transportation systems.

2.6.1 Initialize

- Input:
  - Data from IoT devices (such as vehicles, road sensors, etc.)
  - Initial model parameters (global_model), which will be used as the starting point for training.
  - num_clients: The number of devices or clients that will participate in FL training.
  - local_epochs: The number of epochs for local training in each client.
- Output: Trained FL model.

2.6.2 Distribution Initial Model

- Objective: Deliver the initial model to each client.
- Process: For each client, a copy of the global_model is sent. This allows each client to start training their own model based on the same model.

2.6.3 Local Training

- Objective: Each client trains their own model with local data.
- Process:
  - Clients retrieve local data from their IoT devices.
  - Initialize the local model as a copy of the global_model.
  - In each epoch, the client performs model training by following these steps:
    - Retrieving a subset of local data for training.
    - Calculate loss using local data.
    - Update the local_model with an optimization algorithm, such as Stochastic Gradient Descent (SGD), based on the gradient of the loss.

2.6.4 Local Model Collection

- Objective: Collect training results from all clients to update the global model.
- Process:
  - Collects parameters from each local_model that has been trained by the client.
  - Calculating the average of parameters from all clients to form a new global_model. This is done by summing all parameters and dividing by the number of clients, resulting in a model that is more representative of the overall available data.

2.6.5 Model Evaluation

- Objective: Assess the performance of the updated global model.
- Process:
  - Using validation data to evaluate the performance of the global_model.
  - Calculate performance metrics, such as accuracy or F1-score, to determine whether the model is good enough.
  - If the stopping criteria have not been met (e.g., accuracy is below a threshold or the maximum number of iterations has not been reached), return to step 3 to continue training.

2.6.6 Iterations

- Objective: Repeats the training process until the termination criteria are met.
- Process: The steps from local training to model evaluation are repeated. This process allows the model to be continuously improved based on more and varied data.

2.6.7 Model Implementation

- Objective: Using the trained model in real applications.
- Process:
  - Implement the trained global_model into the intelligent transportation system.
  - The model is used for traffic prediction and optimizing traffic arrangements and vehicle routing, helping in better decision-making.

2.6.8 Monitoring and Maintenance

- Objective: Ensure the model remains effective over time.
- Process:
  - Periodically monitor the performance of the model to ensure that it remains accurate and relevant.
  - If necessary, update the model by repeating the FL process to accommodate new data or changes in traffic patterns.

Each part of the algorithm contributes to the overall goal of building accurate and safe models for intelligent transportation systems. With this distributed approach, FL not only improves efficiency in model training, but also maintains the privacy of user data while enabling more responsive learning to dynamic traffic conditions.

3. Results and Discussion

3.1 Result

In this paper, a literature review on the integration of Federated Learning (FL) in big data analytics for IoT-based intelligent transportation systems has been conducted by analyzing three major papers: Zhang et al. (2022), Wang et al. (2023), and Musa et al. (2023). The table below shows the key findings of each of the analyzed papers:

Study	Methodology	Key Findings	Contribution
Zhang et al. (2022)	Integration of FL with Temporal Convolutional Networks (TCN)	Improved short-term traffic flow prediction accuracy, better responsiveness to traffic conditions, enhanced data privacy	Demonstrated that FL enables real-time adaptation while preserving privacy
Wang et al. (2023)	Development of a big data analytics architecture integrating FL	Enhanced real-time data processing, reduced latency, improved traffic management decision-making	Proposed a scalable big data architecture for FL-based analytics
Musa et al. (2023)	Application of FL in sustainable traffic management	Reduced congestion and carbon emissions, improved decision-making for urban transport systems	Highlighted FL's role in sustainability and smart city traffic optimization

- Zhang et al. (2022) showed that the integration of FL with Temporal Convolutional Networks (TCN) can improve the accuracy of short-term traffic flow predictions. The results showed that the use of local data from multiple sources allows the model to be more responsive to changing traffic conditions, while maintaining user privacy.(Liu et al., 2024)
- Wang et al. (2023) developed a big data analytics architecture that integrates FL to improve data processing efficiency in intelligent transportation systems. The study noted that real-time data processing at the source location reduces latency and improves efficiency, providing better support for decision-making in traffic management.
- Musa et al. (2023) emphasized the importance of FL in supporting sustainable traffic management in smart cities. This research shows how FL can optimize decision-making to reduce congestion and carbon emissions, which are crucial in the context of sustainability.

3.2 Discussion

The results of this literature review show that the integration of Federated Learning in big data analytics offers many benefits for IoT-based intelligent transportation systems. Some key points to discuss include:

- Data Privacy and Security
  One of the key advantages of FL is its ability to train models without moving sensitive data to a central server. This is especially important in the context of transportation, where user data can include personal information[13]. With FL, the data remains on the local device, minimizing the risk of data leakage[14].
- Improved Model Accuracy
  As pointed out by Zhang et al. (2022), the use of local data in model training can improve prediction accuracy. This suggests that more responsive models can be built by utilizing up-to-date information from various sources, which is crucial for proper and quick decision-making in transportation systems.
- Efficiency in Data Processing
  Wang et al. (2023) highlighted that an analytics architecture that integrates FL can improve data processing efficiency. By reducing latency and enabling real-time analysis, FL can support better decision-making in emergency situations or during congestion.
- Sustainability and Management
  Musa et al. (2023) showed that FL can contribute to sustainable traffic management. In the context of smart cities, where population and vehicle growth are constantly increasing, the application of FL can help in achieving sustainability goals by reducing emissions and improving the efficiency of the transportation system.
- Challenges and Future Research Directions
  While the results are promising, there are challenges to overcome, including the development of more adaptive and collaborative FL algorithms, as well as the need for interoperability standards between various IoT systems. Further research is needed to explore the practical implementation of FL in real-world scenarios and to identify how best to integrate this technology into existing infrastructure.

These results and discussions confirm that the integration of Federated Learning in big data analytics for IoT-based intelligent transportation systems has great potential to improve efficiency, safety, and sustainability. By understanding the challenges and opportunities, it is hoped that future research can maximize the use of FL in achieving better and smarter transportation systems.

4. Conclusions

This paper provides a comprehensive review of the integration of Federated Learning (FL) in big data analytics for Internet of Things (IoT)-based intelligent transportation systems. By analyzing three key studies Zhang et al. (2022), Wang et al. (2023), and Musa et al. (2023), this research highlights the significant contributions of FL in addressing major challenges in traffic management, data privacy, and computational efficiency.

Findings indicate that FL enhances the accuracy of traffic flow prediction, ensures secure and efficient data processing, and supports sustainable traffic management. FL's ability to maintain user data privacy while improving real-time decision-making makes it highly relevant for modern transportation systems. Additionally, FL's role in reducing congestion and carbon emissions contributes to the sustainability goals of smart cities.

Despite its numerous advantages, several challenges remain to be addressed. The development of more adaptive and scalable FL algorithms is required to handle dynamic and heterogeneous IoT data. Interoperability between various transportation systems and IoT platforms must also be considered to ensure seamless integration. Furthermore, security issues such as adversarial attacks and model poisoning require robust mitigation strategies to ensure reliable implementation.

Future research should focus on developing a more efficient FL framework that balances computational costs and prediction accuracy while ensuring privacy and security. Studies on real-world FL implementation across various urban environments are also crucial to assessing its impact and scalability. Moreover, collaboration among stakeholders, including policymakers, researchers, and industry practitioners, is essential to promoting the widespread adoption of FL.

In conclusion, the integration of Federated Learning in big data analytics offers promising opportunities to create more efficient, secure, and sustainable intelligent transportation systems. By addressing existing challenges through technological advancements and strategic research initiatives, FL can play a transformative role in shaping the future of smart mobility and urban transportation management.

Acknowledgments

The author sincerely thanks the Faculty of Engineering, University of Mataram, for the support and facilities provided during this research. Special appreciation is extended to the research supervisor for their guidance and valuable insights. Gratitude is also expressed to family, colleagues, and all parties who contributed directly or indirectly to the completion of this study.

References:

[1] J. Rane, Ö. Kaya,S. K. Mallick, and N. L. Rane, "Enhancing customer satisfaction and loyalty in service quality through artificial intelligence, machine learning, internet of things, blockchain, big data, and ChatGPT," in Generative Artificial Intelligence in Agriculture, Education, and Business, Deep Science Publishing, 2024. doi: 10.70593/978-81-981271-7-4_3.

[2] M. Elassy, M. Al-Hattab, M. Takruri, and S. Badawi, "Intelligent transportation systems for sustainable smart cities," Jun. 01, 2024, Elsevier Ltd. doi: 10.1016/j.treng.2024.100252.

[3] A. Addas, M. N. Khan, and F. Naseer, "Waste management 2.0 leveraging internet of things for an efficient and eco-friendly smart city solution," PLoS One, vol. 19, no. 7 July, Jul. 2024, doi: 10.1371/journal.pone.0307608.

[4] C. Ann Udeh, O. Henry Orieno, O. Donald Daraojimba, N. Leonard Ndubuisi, O. Itohan Oriekhoe, and C. Author, "BIG DATA ANALYTICS: A REVIEW OF ITS TRANSFORMATIVE ROLE IN MODERN BUSINESS INTELLIGENCE," Computer Science & IT Research Journal, vol. 5, no. 1, pp. 219–236, 2024, doi: 10.51594/csitrj.v5i.718.

[5] J. Liu et al., "From distributed machine learning to federated learning: A survey," Knowl Inf Syst, vol. 64, no. 4, pp. 885–917, 2022.

[6] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," IEEE Transactions on Communications, vol. 68, no. 2, pp. 1146–1159, 2019.

[7] S. Kaleem, A. Sohail, M. U. Tariq, and M. Asim, "An Improved Big Data Analytics Architecture Using Federated Learning for IoT-Enabled Urban Intelligent Transportation Systems," Sustainability (Switzerland), vol. 15, no. 21, Nov. 2023, doi: 10.3390/su152115333.

[8] G. Paré and S. Kitsiou, "Methods for literature reviews," in Handbook of eHealth evaluation: An evidence-based approach [Internet], University of Victoria, 2017.

[9] H. Taherdoost, "Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects," International Journal of Academic Research in Management (IJARM), vol. 10, no. 1, pp. 10–38, 2021.

[10] A. Moser and I. Korstjens, "Series: Practical guidance to qualitative research. Part 3: Sampling,

data collection and analysis," European journal of general practice, vol. 24, no. 1, pp. 9–18, 2018.

[11] I. Karunarathna, P. Gunasena, K. De Alvis, and A. Jayawardana, "Structured reviews: Organizing, synthesizing, and analyzing scientific literature," Retrieved from ResearchGate, 2024.

[12] T. Liu, Y. Wang, H. Zhou, J. Luo, and F. Deng, "Distributed Short-Term Traffic Flow Prediction Based on Integrating Federated Learning and TCN," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3474300.

[13] D. Milne and D. Watling, "Big data and understanding change in the context of planning transport systems," J Transp Geogr, vol. 76, pp. 235–244, 2019.

[14] T. Wang et al., "Federated Learning-based Information Leakage Risk Detection for Secure Medical Internet of Things," ACM Trans Internet Technol, 2024.