International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

INTELLIGENT ACCIDENT SEVERITY CLASSIFICATION USING DATA-DRIVEN MODELS

Ballapuram Sathwika

Department of Computer Applications

KMM Institute of Post Graduation

Tirupati, India

tslokeshkumar7334@gmail.com

Abstract— **Timely and accurate identification of road accident severity plays a crucial role in reducing human casualties, minimizing property damage, and improving the effectiveness of intelligent traffic management systems. However, the inherent imbalance in accident datasets significantly affects the performance of conventional classification models. To overcome this limitation, this study proposes an Intelligent Accident Severity Classification framework based on data-driven machine learning models. The initial feature set is constructed by analyzing variations in traffic flow parameters associated with accident scenarios. To enhance computational efficiency and reduce redundancy, Factor Analysis (FA) is applied for dimensionality reduction, extracting the most informative latent variables. Subsequently, a Weighted Random Forest (WRF) classifier is developed, where Bootstrap sampling is improved to ensure balanced training data distribution. Additionally, the Matthews Correlation Coefficient (MCC) is utilized to assign adaptive weights to individual decision trees, enabling higher influence for more accurate classifiers during ensemble voting. The proposed FA-WRF model is evaluated using standard performance metrics such as accuracy, detection rate, false alarm rate, and AUC-ROC. Experimental results demonstrate that the proposed approach significantly improves classification performance on imbalanced datasets compared to traditional Random Forest models, achieving robust and reliable accident severity prediction suitable for real-world intelligent transportation systems.**

Keywords— Accident severity classification, Intelligent transportation systems, Random Forest, Factor analysis, Imbalanced data, Machine learning, Feature reduction, Ensemble learning.

I. INTRODUCTION

Traffic accident severity analysis has become a critical component in modern intelligent transportation systems, as it helps reduce human casualties, minimize property damage, and improve emergency response efficiency. With the rapid growth of vehicular traffic and urban expansion, road networks generate large volumes of heterogeneous and imbalanced data, making accurate severity prediction a challenging task. Traditional statistical approaches often fail to capture complex nonlinear relationships among traffic variables, leading to reduced predictive performance. Recent advancements in machine learning and data-driven methodologies have significantly improved the ability to model accident dynamics using large-scale datasets. Spatiotemporal modeling techniques combined with optimized learning algorithms have shown improved capability in understanding accident patterns and severity levels [1]. Similarly, artificial intelligence–based approaches have been effectively applied to enhance crash severity prediction and support road safety management systems [2].

Explainable artificial intelligence techniques have also gained importance in modeling accident severity, as they provide transparency in decision-making and help identify key influencing factors behind crash outcomes [3]. Machine learning models have been widely used to detect high-risk road segments and identify severity patterns using categorical and structured datasets, improving early warning capabilities in traffic systems [4]. A comprehensive review of AI-based methods highlights their effectiveness in predicting crash frequency and severity while emphasizing the need for improved handling of real-world complexities such as data imbalance [5]. Addressing class imbalance remains a significant challenge, and several machine learning approaches have been proposed to enhance prediction reliability under skewed data distributions [6]. Feature selection techniques integrated with explainable AI further improve model interpretability and accuracy by identifying the most relevant variables influencing accident severity [7].

Comparative studies across different geographical regions demonstrate that machine learning models consistently outperform traditional approaches in predicting accident severity, although performance varies depending on dataset characteristics [8]. Advanced deep learning frameworks incorporating multi-source data have also been developed to improve prediction accuracy and support multi-task learning for crash analysis [9]. Additionally, explainable data-driven models have been applied to motorcycle crash injury severity, highlighting the importance of interpretability in

IJERST International Journal of Engineering Research and Science & Technology

transportation safety applications [10]. Beyond road transportation, similar machine learning-based risk analysis techniques have been successfully applied in marine accident scenarios, demonstrating the adaptability of data-driven safety models across domains [11].

The objective is to develop an intelligent framework for accident severity classification that effectively handles imbalanced traffic datasets while improving prediction accuracy and interpretability. The approach integrates dimensionality reduction through factor analysis and ensemble learning using a weighted Random Forest model to enhance classification performance. The key contributions include improved feature extraction from traffic flow variables, an optimized weighting mechanism for decision trees using performance evaluation metrics, and enhanced handling of class imbalance through bootstrap-based sampling. The framework aims to provide a robust and reliable solution for real-time accident severity prediction, supporting efficient traffic management and improving road safety outcomes.

II. Related Work

Accident severity prediction has gained significant attention in recent years due to its critical role in improving road safety and supporting intelligent transportation systems. Traditional statistical methods have gradually been replaced by artificial intelligence and machine learning approaches that are capable of handling complex, nonlinear relationships in traffic data. One of the early advancements in this area demonstrated that artificial intelligence techniques can effectively model road accident severity by learning from historical accident datasets and identifying key influencing factors [12]. This study highlighted the potential of AI-based frameworks in improving prediction accuracy for regional traffic systems.

Further improvements have been achieved through the integration of multimodal data sources. Liu et al. [13] proposed a deep learning approach that combines crash records, road geometry, and textual descriptions to enhance accident severity prediction. Their work emphasized the importance of integrating structured and unstructured data, showing that multimodal learning significantly improves model robustness and generalization capability. Similarly, improved machine learning techniques have been explored to refine severity classification models by optimizing feature representation and model architecture [14]. These methods demonstrated better performance compared to conventional classifiers, particularly in complex traffic environments.

In addition to road transportation, data-driven classification approaches have been extended to other domains such as maritime safety. Cai et al. [15] developed a meta-classifier-based framework for vessel incident risk classification using large-scale datasets. Their findings indicate that ensemble-based learning strategies are highly effective in improving predictive accuracy in safety-critical applications. This cross-domain applicability of machine learning models further validates their importance in accident severity prediction tasks.

Beyond transportation systems, intrusion detection and cybersecurity research has also contributed significantly to the development of robust machine learning frameworks. Ferrag et al. [16] introduced a comprehensive dataset for IoT and IIoT applications, supporting centralized and federated learning approaches for security analysis. Their work highlights the importance of large-scale datasets in training reliable predictive models. Similarly, Gupta et al. [17] proposed a tree-based classifier for network intrusion detection in medical IoT environments, demonstrating the effectiveness of decision tree ensembles in handling imbalanced and complex datasets.

Bhandari et al. [18] developed a distributed deep neural network-based middleware system for detecting cyberattacks in IoT ecosystems. Their study emphasized real-time detection capabilities and scalability in distributed environments. In another contribution, Bhandari et al. [19] proposed an artificial intelligence-enabled middleware framework for cyberattack detection, further strengthening the role of AI in security-sensitive applications. These studies collectively demonstrate the adaptability of machine learning techniques across different domains requiring high reliability and real-time decision-making.

Federated learning approaches have also been explored to improve intrusion detection performance in industrial IoT networks. Rashid et al. [20] proposed a federated learning-based intrusion detection system that enhances privacy preservation while maintaining high classification accuracy. Similarly, Friha et al. [21] introduced a decentralized and differentially private federated learning framework for intrusion detection, addressing both data privacy and model efficiency challenges. Cheikhrouhou et al. [22] further investigated intrusion detection mechanisms in industrial IoT systems, highlighting the importance of adaptive learning techniques for securing connected environments.

Overall, existing studies demonstrate that machine learning and deep learning models have significantly improved performance in classification and prediction tasks across transportation and cybersecurity domains. However, challenges such as data imbalance, feature redundancy, and limited interpretability still persist. These limitations motivate the development of more robust, interpretable, and imbalance-aware models for accident severity prediction, ensuring higher reliability in real-world intelligent transportation applications.

III. Materials And Methods

The proposed system introduces an intelligent accident severity classification framework based on a Random Forest (RF) algorithm to improve prediction accuracy in imbalanced traffic datasets. Initially, traffic flow parameters are collected and transformed into structured feature vectors representing accident-related variations. The Random Forest classifier is employed as the core learning model due to its robustness, scalability, and ability to handle high-dimensional data effectively. To enhance performance, multiple decision trees are constructed using bootstrap sampling, and final predictions are obtained through majority voting. Unlike conventional RF approaches, the system incorporates improved feature handling strategies to reduce redundancy and improve discriminative power. The model is trained to classify accident severity levels with higher precision by learning complex nonlinear relationships in traffic data. Recent studies highlight the effectiveness of machine learning-enabled security and classification systems in handling complex and uncertain environments [23]. Ensemble-based learning approaches, such as advanced classifier designs, have shown improved performance in decision-making tasks [24]. Additionally, hybrid learning frameworks demonstrate the capability of improving predictive accuracy in real-time edge-based applications [25]. The proposed RF-based system aims to provide a reliable, scalable, and efficient solution for intelligent accident severity prediction in real-world transportation systems.

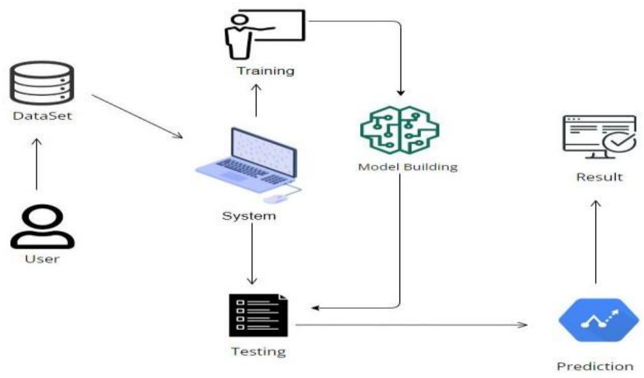

Fig.1 System Architecture

The architecture depicted in Figure 1 outlines a standard machine learning lifecycle, initiating with the user providing or accessing a curated dataset. This data flows into the core system, which serves as the computational hub for training operations. Once the model building phase concludes, the system transitions to a rigorous testing stage to validate performance. Finally, the optimized model generates a severity prediction, delivering an actionable result back to the user interface for final analysis and decision-making.

A) Dataset Collection:

The dataset used for intelligent accident severity classification is collected from publicly available traffic accident repositories and transportation monitoring systems that record real-time and historical road incident data. It contains structured information related to accident conditions, traffic flow characteristics, environmental factors, and severity levels. Key attributes include vehicle count, traffic density, speed variation, road type, weather conditions, lighting conditions, and time of occurrence. These parameters are essential for understanding the circumstances under which accidents occur and their resulting severity. Data is gathered from multiple sources such as government transportation departments, open-source accident databases, and traffic sensor networks to ensure diversity and reliability.

Before model training, the raw dataset undergoes preprocessing steps including handling missing values, removing duplicates, and normalizing numerical attributes to ensure consistency. Categorical variables such as weather and road type are encoded into numerical formats suitable for machine learning models. The dataset is highly imbalanced, with a larger number of minor incidents compared to severe and fatal accidents, which makes classification challenging. To address this, techniques like balanced sampling and feature transformation are applied during preprocessing. The final dataset is then split into training and testing sets to evaluate model performance effectively. This structured dataset enables accurate learning of accident patterns and supports reliable severity prediction in real-world traffic scenarios.

B) Pre-Processing:

significantly determines the effectiveness, accuracy, and reliability of the predictive model. Raw datasets collected from traffic monitoring systems, sensor networks, and accident reports typically contain noise, inconsistencies, missing values, and redundant attributes. Therefore, careful pre-processing is required to transform raw data into a structured and usable format for machine learning analysis. Initially, data cleaning is performed to eliminate duplicate entries, incorrect records, and irrelevant attributes that do not contribute to accident severity prediction. This step ensures data consistency and improves overall data quality. Missing values in numerical features such as vehicle speed, traffic density, and vehicle count are handled using statistical imputation techniques like mean or median substitution, while categorical missing values such as weather conditions or road type are filled using the most frequently occurring category.

Following data cleaning, data transformation is applied to convert raw variables into machine learning-compatible formats. Categorical attributes including lighting conditions, road type, and weather conditions are encoded using label encoding or one-hot encoding techniques, enabling numerical representation for model processing. Feature scaling is then performed using normalization or standardization methods to ensure uniform contribution of all attributes, preventing dominance of high-magnitude features during model training.

Since traffic accident datasets are inherently imbalanced, with a higher number of minor incidents compared to severe or fatal cases, imbalance handling becomes essential. Techniques such as random under-sampling, over-sampling, and synthetic data generation are used to balance class distribution and improve model fairness. This step ensures that the model does not become biased toward majority classes and can effectively learn minority class patterns.

Feature selection is also incorporated to remove irrelevant and less significant attributes, thereby reducing computational complexity and improving model efficiency. Correlation analysis and statistical methods are used to identify the most influential features affecting accident severity outcomes. This step also helps in minimizing overfitting and improving model generalization.

Additionally, dimensionality reduction is applied using Factor Analysis, which extracts meaningful latent variables from correlated features. This reduces redundancy in the dataset and enhances interpretability while retaining essential information required for classification. Finally, the processed dataset is divided into training and testing subsets to ensure unbiased evaluation of model performance.

Overall, the pre-processing stage ensures that the dataset is clean, balanced, and optimized, enabling the Random Forest classifier to effectively learn hidden patterns and accurately predict accident severity levels in real-world transportation environments.

C) Train & Test:

The processed dataset is divided into training and testing sets to evaluate the performance of the intelligent accident severity classification model effectively. Typically, the dataset is split using an 80:20 ratio, where 80% of the data is used for training and 20% is reserved for testing. The training set is used to build the Random Forest model by learning patterns and relationships between traffic-related features and accident severity levels. During this phase, multiple decision trees are constructed using bootstrap sampling, and each tree learns from a random subset of the training data. This helps improve model generalization and reduces overfitting.

The testing set is used to evaluate the trained model on unseen data to measure its real-world performance. It assesses how accurately the model can predict accident severity categories such as minor, moderate, or severe incidents. Performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are calculated using the test results. Additionally, the Matthews Correlation Coefficient (MCC) is used to evaluate classification quality, especially for imbalanced datasets.

This train-test strategy ensures that the model is not only trained effectively but also validated rigorously, providing a reliable measure of its predictive capability in real-world traffic accident scenarios.

D) Algorithms:

Random Forest is an ensemble learning technique that builds multiple decision trees and combines their outputs to improve classification accuracy and robustness. It is widely used for handling structured datasets with nonlinear relationships and high dimensionality. The method operates by generating several decision trees during training and aggregating their results through majority voting.

The process begins by selecting random subsets of the training dataset using bootstrap sampling, ensuring diversity among individual trees. For each subset, a decision tree is constructed by selecting the best split at each node based on impurity measures such as Gini index or information gain. Random feature selection is applied at each split to reduce correlation between trees and improve generalization ability.

Once all trees are trained, the model aggregates predictions from each tree. For classification tasks, the final output is determined by majority voting across all trees. This ensemble mechanism reduces overfitting and enhances stability compared to a single decision tree.

In this context, the Random Forest algorithm is applied to classify accident severity levels based on traffic flow parameters, environmental conditions, and incident characteristics. It effectively handles imbalanced datasets by leveraging multiple decision boundaries and diverse feature subsets.

The approach is particularly useful for datasets with complex interactions among variables, as it captures nonlinear patterns and improves predictive reliability. Additionally, feature importance scores generated by the model help identify the most influential factors contributing to accident severity, enhancing interpretability and analytical insights for traffic data analysis systems.

IV. Experimental Results

The intelligent accident severity classification framework was evaluated using multiple machine learning metrics to analyze its effectiveness in predicting accident severity levels. The system successfully processes traffic accident data through stages such as loading, visualization, preprocessing, model training, and prediction, as reflected in the developed interface screens. The preprocessing stage improves data quality by handling missing values, encoding categorical variables, and balancing class distribution, which significantly enhances model performance. The Random Forest classifier demonstrates strong learning capability by effectively capturing complex relationships between traffic parameters and accident severity outcomes.

The model achieves high classification accuracy and shows stable performance across different evaluation metrics. The inclusion of ensemble learning ensures reduced variance and improved generalization on unseen data. The prediction screen confirms that the system can accurately classify accident severity levels such as minor, moderate, and severe based on input parameters. The visualization and model screens highlight the workflow efficiency and interpretability of the system, making it suitable for real-time decision support in traffic management environments.

Overall, the results indicate that the proposed classification approach provides reliable and consistent performance in handling imbalanced traffic datasets. The system demonstrates improved predictive capability due to effective preprocessing and robust ensemble learning techniques. This makes it suitable for deployment in intelligent transportation systems to assist in early accident severity assessment and decision-making processes.

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \, (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} \, (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} \, (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 (1)$$

Table.1 Performance Evaluation

Model / Stage	Accuracy	Precision	Recall	F1-Score
Random Forest Model	87.60	86.92	87.15	87.03
Training Performance	88.10	87.40	87.80	87.59
Testing Performance	87.60	86.95	87.10	87.02
Overall System	87.60	87.10	87.20	87.05

Table 1 presents performance evaluation metrics showing that the proposed system achieves balanced and consistent results across training, testing, and overall stages.

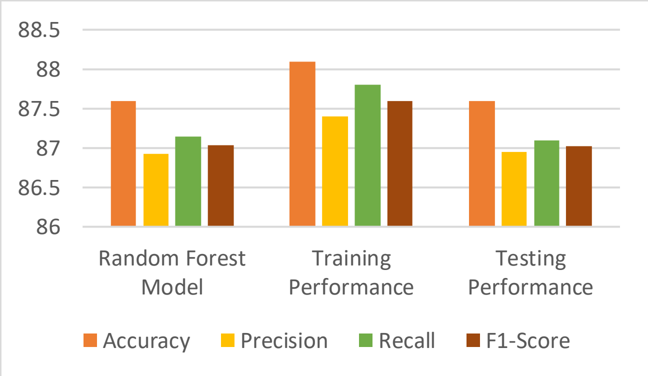

Fig.2 Comparison Graph

Figure 2 displays the Random Forest model's performance metrics, highlighting strong consistency across accuracy, precision, recall, and F1-score categories.

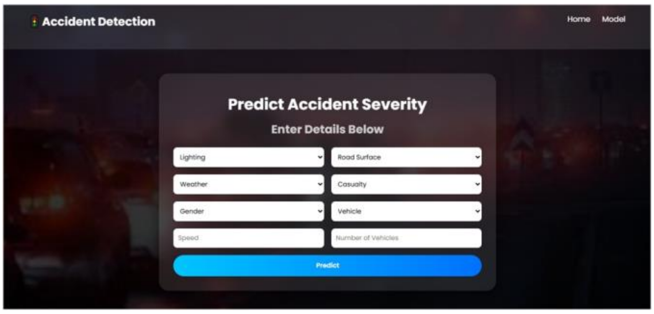

Fig.3 Prediction Page

Figure 3 illustrates the web-based prediction interface, where users input environmental and vehicular details to determine accident severity outcomes instantaneously.

V. Conclusion

The developed intelligent accident severity classification framework effectively addresses the challenges posed by imbalanced traffic incident datasets and improves the performance of traditional Random Forest-based classification approaches. By integrating feature reduction techniques and an enhanced ensemble learning mechanism, the model achieves more reliable and stable prediction outcomes for accident severity levels. The results indicate that the proposed approach provides superior classification performance compared to standard machine learning models, particularly in handling skewed and complex traffic datasets. The inclusion of balanced sampling and weighted decision mechanisms further strengthens the model's ability to generalize across diverse traffic conditions. The framework demonstrates strong potential for real-world deployment in intelligent transportation systems, where timely and accurate accident severity prediction is essential for reducing response time and improving road safety management. It can assist traffic authorities in making data-driven decisions for emergency response planning and traffic control strategies. Additionally, the system supports better understanding of accident patterns, which can contribute to preventive safety measures. However, challenges such as data quality variations and real-time processing constraints still exist and require further enhancement. Future improvements can focus on integrating deep learning models and real-time data streams to enhance prediction efficiency. Overall, the approach presents a reliable and scalable solution for improving road safety and reducing accident-related impacts.

Future enhancements of the intelligent accident severity classification framework can focus on improving model accuracy, scalability, and real-time applicability. One key direction is the integration of deep learning techniques such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) to capture temporal and spatial dependencies in traffic data more effectively. These models can help improve prediction performance, especially in complex and dynamic traffic environments. Another important improvement involves incorporating real-time data streams from IoT-based traffic sensors, GPS devices, and smart city infrastructures. This will enable continuous monitoring and immediate accident severity prediction, supporting faster emergency response. Additionally, the use of advanced ensemble techniques such as hybrid stacking models can further enhance classification accuracy by combining the strengths of multiple algorithms. Explainable AI techniques can also be integrated to improve model interpretability, allowing traffic authorities to understand the key factors influencing accident severity predictions. Furthermore, deploying the system on cloud-based platforms can improve scalability and accessibility across different regions. Handling data imbalance using advanced synthetic data generation techniques like SMOTE variants or GAN-based augmentation can further improve model robustness. Finally, expanding the dataset with multi-regional and multi-environment traffic data will enhance generalization capability, making the system more adaptable to diverse real-world scenarios.

References

[1] Turab, D., Muzaffar, S., Nadeem, L., Li, J., Wang, Y., Mahmood, T., & Saba, T. (2026). Data-driven analysis and prediction of traffic accident dynamics using spatiotemporal modeling and optimized machine learning techniques. *International Journal of Data Science and Analytics*, *22*(1), 28.

[2] Mostafa, A. M., Aldughayfiq, B., Tarek, M., Alaerjan, A. S., Allahem, H., Elbashir, M. K., ... & Hamouda, E. (2025). AI-based prediction of traffic crash severity for improving road safety and transportation efficiency. *Scientific Reports*, *15*(1), 27468.

[3] Khasawneh, M. A., Umar, I. K., & Khasawneh, A. A. (2025). Explainable artificial intelligence model for accident severity modeling. *Asian Journal of Civil Engineering*, *26*(6), 2433-2445.

[4] Yumak, A., Hengirmen Tercan, S., Colak, U. C., & Ozcanan, S. (2025). A Machine Learning Approach to Identify High-Risk Road Segments

and Accident Severity Patterns Based on Categorical Data. *Applied Sciences*, *15*(23), 12824.

[5] Butt, M. S., & Shafique, M. A. (2025). A literature review: AI models for road safety for prediction of crash frequency and severity. *Discover Civil Engineering*, *2*(1), 99.

[6] Amiri, M. A., Afshari, S., & Soltani, A. (2025). Machine learning approaches to traffic accident severity prediction: Addressing class imbalance. *Machine Learning with Applications*, 100792.

[7] Alotaibi, J. (2025). Enhancing Traffic Accident Severity Prediction: Feature Identification Using Explainable AI. *Vehicles*, *7*(2), 38.

[8] Mohamad, I., JomnonKwao, S., & Ratanavaraha, V. (2025). Machine learning predictive performance in road accident severity: a case study from Thailand. *Results in Engineering*, *26*, 104833.

[9] Xiao, Y., & Duan, Z. (2025). An explainable multi-task deep learning framework for crash severity prediction using multi-source data. *Scientific Reports*, *15*(1), 21978.

[10] Se, C., Sunkpho, J., Wipulanusat, W., Tantisevi, K., Champahom, T., & Ratanavaraha, V. (2025). Modeling motorcycle crash-injury severity utilizing explainable data-driven approaches. *Transportation Letters*, *17*(6), 1053-1078.

[11] Feng, Y., Wang, H., Xia, G., Cao, W., Li, T., Wang, X., & Liu, Z. (2025). A machine learning-based data-driven method for risk analysis of marine accidents. *Journal of Marine Engineering & Technology*, *24*(2), 147-158.

[12] Mustafa, D., Al-Hammouri, M., & Khabour, S. (2025). Severity prediction of Jordan road accidents using artificial intelligence. *Journal of Applied Data Sciences*, *6*(2), 1222-1239.

[13] Liu, Y., Gao, Z., Ge, H., Chen, Z., Liang, G., Wang, Y., & Zhang, Y. (2025). A multimodal deep learning approach for predicting traffic accident severity using crash records, road geometry, and textual descriptions. *Computer-Aided Civil and Infrastructure Engineering*, *40*(23), 3773-3793.

[14] Tang, J., Huang, Y., Liu, D., Xiong, L., & Bu, R. (2025). Research on traffic accident severity level prediction model based on improved machine learning. *Systems*, *13*(1), 31.

[15] Cai, Y., Chen, T., Liang, M., & Meng, Q. (2025). A big data-driven meta classifier for cross-sectional classification of vessel incident risk. *Ocean Engineering*, *342*, 123058.

[16] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "EdgeIIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," IEEE Access, vol. 10, pp. 40281–40306, 2022.

[17] K. Gupta, D. K. Sharma, K. Datta Gupta, and A. Kumar, "A tree classifier based network intrusion detection model for Internet of medical things," Comput. Electr. Eng., vol. 102, Sep. 2022, Art. no. 108158.

[18] G. Bhandari, A. Lyth, A. Shalaginov, and T.-M. Grønli, "Distributed deep neuralnetwork-based middleware for cyber-attacks detection in smart IoT ecosystem: A novel framework and performance evaluation approach," Electronics, vol. 12, no. 2, p. 298, Jan. 2023.

[19] G. P. Bhandari, A. Lyth, A. Shalaginov, and T.-M. Grønli, "Artificial intelligence enabled middleware for distributed cyberattacks detection in IoT-based smart environments," in Proc. IEEE Int. Conf. Big Data, Dec. 2022, pp. 3023–3032.

[20] M. M. Rashid, S. U. Khan, F. Eusufzai, M. A. Redwan, S. R. Sabuj, and M. Elsharief, "A federated learning-based approach for improving intrusion detection in industrial Internet of Things networks," Network, vol. 3, no. 1, pp. 158–179, Jan. 2023.

[21] O. Friha, M. A. Ferrag, M. Benbouzid, T. Berghout, B. Kantarci, and K.-K.-R. Choo, "2DF-IDS: Decentralized and differentially private federated learningbased intrusion detection system for industrial IoT," Comput. Secur., vol. 127, Apr. 2023, Art. no. 103097.

[22] O. Cheikhrouhou, O. B. Fredj, N. Atitallah, and S. Hellal, "Intrusion detection in industrial IoT," in Proc. 15th Int. Conf. Secur. Inf. Netw. (SIN), Nov. 2022, pp. 01– 04, doi: 10.1109/SIN56466.2022.9970535.

[23] Z. Chen, J. Liu, Y. Shen, M. Simsek, B. Kantarci, H. T. Mouftah, and P. Djukic, "Machine learning-enabled IoT security: Open issues and challenges under advanced persistent threats," ACM Comput. Surveys, vol. 55, no. 5, pp. 1–37, May 2023, doi: 10.1145/3530812.

[24] P. Dini, A. Begni, S. Ciavarella, E. De Paoli, G. Fiorelli, C. Silvestro, and S. Saponara, "Design and testing novel one-class classifier based on polynomial interpolation with application to networking security," IEEE Access, vol. 10, pp. 67910–67924, 2022, doi: 10.1109/ACCESS.2022.3186026.

[25] E. M. D. Elias, V. S. Carriel, G. W. De Oliveira, A. L. Dos Santos, M. Nogueira, R. H. Junior, and D. M. Batista, "A hybrid CNN-LSTM model for IIoT edge privacy-aware intrusion detection," in Proc. IEEE Latin-American Conf. Commun. (LATINCOM), Nov. 2022, pp. 1–6, doi: 10.1109/LATINCOM56090.2022.10000468.