An Intelligent System for Automatic Evaluation and Reporting of Vehicle Accidents Based on EfficientNet Model and CBAM Attention Mechanism

Seyed Farhad Kazemian Torbaghan[1*], Vahid Torkzadeh[1] Negin Mohammadpoor Abdolabadi [2]

[1] Department of Computer Engineering, Ma.C., Islamic Azad University, Mashhad, Iran
[2] Faculty of Law and Political Science, Ferdowsi University of Mashhad, Mashhad, Iran

*** Corresponding author email address**: seyedfarhad.kazemiantorbaghan@iau.ir

Received: 2026-02-17 **Revised:** 2026-05-26 **Accepted:** 2026-06-03 **Initial Publish:** 2026-06-04 **Final Publish:** 2027-03-01

Abstract

Road traffic accidents remain a critical challenge in road safety, where rapid detection and accurate severity assessment play a vital role in reducing fatalities. This study presents an intelligent system for the automatic evaluation of vehicle accidents based on the EfficientNet-B0 model enhanced with the Convolutional Block Attention Module (CBAM). The primary objective is to overcome the severe class imbalance issue, particularly in the “low-risk” class, and to improve detection accuracy under real-world conditions. The proposed architecture consists of two main modules: a multi-task convolutional neural network for accident detection and severity classification, and a YOLOv8-Nano module for real-time fire and smoke identification. To address data imbalance, a combined strategy involving external data augmentation, oversampling, and an eight-stage training pipeline was employed. Experimental evaluation on the Accident Images Analysis dataset demonstrated that adding the CBAM mechanism led to a 38% improvement in the F1-score of the minority class. The final system achieved an accuracy of 94% in accident detection and 74% in severity classification (with a Macro F1 score of 0.61). Comparison with Pashaei et al. (2020) on the public version of the dataset showed a 4.49% improvement in the Macro F1 metric. The results confirm that integrating attention mechanisms with efficient architectures and smart data management significantly enhances the performance of emergency response support systems.

Keywords: *Accident Detection, EfficientNet, CBAM Attention Mechanism, Severity Classification, Deep Learning*

How to cite this article:

Kazemian Torbaghan, S. F., Torkzadeh, V., & Mohammadpoor Abdolabadi, N. (2027). An Intelligent System for Automatic Evaluation and Reporting of Vehicle Accidents Based on EfficientNet Model and CBAM Attention Mechanism. Management Strategies and Engineering Sciences, 9(2), 1-10.

1. Introduction

Road traffic accidents remain one of the most serious public health and transportation challenges worldwide. Despite significant advances in vehicle safety technologies, road infrastructure, and emergency medical services, traffic crashes continue to impose substantial human, social, and economic costs on societies. According to the latest global estimates, approximately 1.19 million people die annually as a result of road traffic accidents, while tens of millions of individuals suffer injuries that often lead to long-term disabilities and reduced quality of life [1]. Beyond the direct human consequences, traffic accidents generate extensive economic losses through healthcare expenditures, property damage, productivity reduction, and emergency response costs. Consequently, improving the speed and accuracy of accident detection and severity assessment has become a strategic priority for governments, transportation agencies, and healthcare systems worldwide [1].

The challenge of road safety is particularly significant in developing countries, where traffic mortality rates often exceed global averages due to infrastructure limitations, delayed emergency response systems, and insufficient accident monitoring mechanisms. In Iran, traffic accidents remain one of the leading causes of unnatural deaths. Official statistics published by the Forensic Medicine Organization of Iran indicate that thousands of fatalities occur annually due to road crashes, highlighting the urgent need for intelligent technologies capable of supporting rapid accident assessment and intervention [2]. The effectiveness of emergency response is strongly influenced by the speed with which accidents are detected and reported. Delays in identifying accident locations and evaluating injury severity can significantly reduce the chances of survival for victims and increase the likelihood of long-term complications.

Historically, accident reporting systems have relied heavily on eyewitness accounts, emergency telephone calls, and manual reporting procedures. Although these approaches have served as the foundation of traffic emergency management for decades, they suffer from inherent limitations. Human observers may fail to recognize the seriousness of an accident, provide incomplete information, or delay contacting emergency services due to confusion, panic, or uncertainty. Psychological research has further demonstrated that individuals witnessing emergencies may hesitate to intervene because of the diffusion of responsibility phenomenon, whereby people assume that others will take action instead [3]. Consequently, dependence on manual reporting mechanisms can compromise the timeliness and accuracy of emergency responses.

The rapid expansion of mobile technologies has created new opportunities for transforming accident detection and reporting systems. Smartphone adoption has increased dramatically over the past decade, reaching billions of users worldwide and providing unprecedented access to devices equipped with cameras, sensors, processing capabilities, and internet connectivity [4]. These devices have effectively transformed ordinary citizens into distributed data collectors capable of capturing visual evidence immediately after accidents occur. The widespread availability of smartphones offers a unique infrastructure for intelligent traffic monitoring systems that can leverage user-generated images and videos to assess accident conditions in real time.

The growing availability of crowdsourced traffic information has attracted considerable research interest. Studies have demonstrated that user-generated traffic reports can identify accidents and road incidents earlier than traditional reporting systems. For example, research examining crowdsourced traffic data showed that community-based reporting platforms can provide timely information regarding crash events, potentially improving emergency response efficiency [5]. However, although crowdsourced systems can accelerate accident notification, they typically lack the capability to automatically analyze visual evidence and objectively determine accident severity. As a result, emergency management authorities may still face challenges in prioritizing incidents and allocating resources effectively.

Recent advances in artificial intelligence, particularly deep learning and computer vision, have significantly expanded the possibilities for automated accident analysis. Deep learning algorithms have demonstrated remarkable success in image classification, object detection, scene understanding, and pattern recognition tasks across diverse application domains. Convolutional Neural Networks (CNNs), in particular, have become the dominant architecture for visual analysis due to their ability to automatically learn hierarchical representations from raw image data. These models can identify subtle visual patterns that may be difficult or impossible for human observers to detect consistently, making them highly suitable for accident severity assessment and damage evaluation [6].

Within the broader family of CNN architectures, EfficientNet has emerged as one of the most influential developments in modern deep learning. Unlike conventional scaling approaches that independently increase network depth, width, or resolution, EfficientNet introduces a compound scaling strategy that simultaneously optimizes these dimensions to achieve superior accuracy while maintaining computational efficiency [7]. This balance between performance and resource requirements makes EfficientNet particularly attractive for real-world applications involving mobile devices, edge computing environments, and real-time decision-support systems. The ability to deploy powerful models with relatively low computational costs is especially important for accident detection systems intended to operate under practical constraints.

While deep learning architectures have achieved substantial success in accident-related applications, several important challenges remain unresolved. One of the most critical issues involves the accurate classification of accident severity. Determining whether an accident is minor, moderate, or severe requires careful analysis of vehicle deformation, collision characteristics, environmental conditions, and contextual information. Small visual differences between severity levels can significantly affect model performance, particularly when training datasets contain highly imbalanced class distributions. In many real-world accident datasets, severe and moderate accidents are represented much more frequently than minor incidents, causing models to become biased toward majority classes and reducing their effectiveness in detecting less represented categories [6].

Researchers have proposed various approaches to address accident severity prediction using deep learning. Aboulola developed a MobileNet-based transfer learning framework combined with explainable artificial intelligence techniques to improve accident severity prediction performance and model interpretability [8]. Similarly, Benfaress and colleagues employed deep neural architectures together with SHAP-based interpretability methods to enhance traffic accident severity prediction while providing greater transparency regarding model decisions [9]. These studies demonstrate the growing importance of combining predictive accuracy with explainability in safety-critical applications. Nevertheless, challenges associated with class imbalance, feature localization, and robust performance under diverse environmental conditions continue to limit the practical effectiveness of many existing approaches.

The development of attention mechanisms has provided a promising solution to some of these limitations. Attention modules enable neural networks to focus selectively on the most informative regions and features within an image, thereby reducing the influence of irrelevant background information. One of the most widely adopted attention architectures is the Convolutional Block Attention Module (CBAM), which sequentially applies channel attention and spatial attention mechanisms to enhance feature representation [10]. By guiding the network toward salient image regions, CBAM can improve the detection of subtle visual cues associated with vehicle damage and accident severity. The integration of attention mechanisms into convolutional architectures has consistently demonstrated performance improvements across numerous computer vision applications.

The effectiveness of attention-enhanced architectures has also been confirmed in related safety-critical domains. Yun and colleagues introduced FFYOLO, a lightweight fire detection framework incorporating advanced attention mechanisms, demonstrating improved feature extraction capabilities and enhanced detection performance under challenging conditions [11]. The success of attention-based approaches in fire detection suggests that similar mechanisms may be highly beneficial for traffic accident analysis, where identifying critical visual features such as deformation patterns, collision points, and secondary hazards is essential for accurate severity classification.

Another important trend influencing intelligent transportation systems is the increasing integration of advanced machine learning with traffic analytics and forecasting technologies. Recent studies have explored graph neural networks, spatio-temporal transformers, and dynamic graph convolutional architectures to model complex traffic patterns and predict traffic flows with high accuracy [12, 13]. Although these studies focus primarily on traffic forecasting rather than accident detection, they illustrate the broader transformation occurring within transportation intelligence, where sophisticated deep learning models are increasingly used to support real-time decision-making and infrastructure management.

The availability of large-scale annotated datasets has further accelerated research progress in accident detection and damage assessment. Public datasets such as the Car Damage Severity Dataset provide valuable resources for training models to recognize varying levels of vehicle damage [14]. Similarly, the Car Accidents and Deformation Dataset offers annotated accident imagery that supports the development of deformation analysis algorithms and accident classification systems [15]. Additional datasets, including fire and smoke image collections, contribute to the identification of secondary hazards frequently associated with severe traffic incidents [16]. The integration of multiple datasets can enhance model generalization and improve performance under diverse real-world conditions.

Recent research has also emphasized the importance of end-to-end intelligent accident management systems capable of not only detecting accidents but also initiating emergency response procedures. Namdeo and Deshmukh proposed a computer vision framework integrating YOLO-based accident detection, severity classification, and automated messaging services to facilitate rapid emergency notifications [17]. Such approaches highlight the growing movement toward fully automated accident response ecosystems in which artificial intelligence serves as both an analytical and operational component. Nevertheless, achieving reliable performance across different accident severity levels remains a major challenge, particularly when minority classes are underrepresented within training data.

Given these considerations, there remains a substantial need for intelligent accident assessment systems that combine computational efficiency, robust feature extraction, attention-guided learning, and effective class imbalance management. Existing studies have demonstrated the value of deep learning, transfer learning, attention mechanisms, and explainable artificial intelligence individually; however, fewer investigations have systematically integrated these components within a unified framework optimized for accident severity classification under realistic data conditions [8-10]. Furthermore, the increasing availability of smartphone-generated accident imagery provides an opportunity to develop scalable systems capable of supporting real-time emergency response while minimizing computational requirements [4, 5].

Therefore, the aim of the present study is to develop and evaluate an intelligent system for automatic vehicle accident evaluation and reporting based on the EfficientNet model enhanced with the CBAM attention mechanism, with particular emphasis on improving accident severity classification performance under severe class imbalance conditions.

2. Methodology

In this section, the proposed architecture for the automatic evaluation and reporting system of vehicle accidents is described in detail. The provided approach is based on a multi-task architecture that performs two tasks simultaneously: detecting the occurrence of an accident and classifying its severity. The core of this system is the EfficientNet-B0 Convolutional Neural Network, enhanced with the CBAM attention mechanism. In addition to the architecture, an eight-stage training strategy and class imbalance management techniques were employed to overcome data challenges.

2.1. General System Architecture

The proposed system consists of two main modules: the Accident Classification Module (based on EfficientNet-B0) and the Smoke and Fire Detection Module (based on YOLOv8-Nano). Image processing begins with input, followed by preprocessing and normalization, after which it is sent to the main module. In this module, visual features are extracted and then transferred through shared layers to two separate classification heads; one for detecting the presence or absence of an accident (binary classification) and another for classifying the accident severity into three levels (low-risk, moderately dangerous, dangerous).

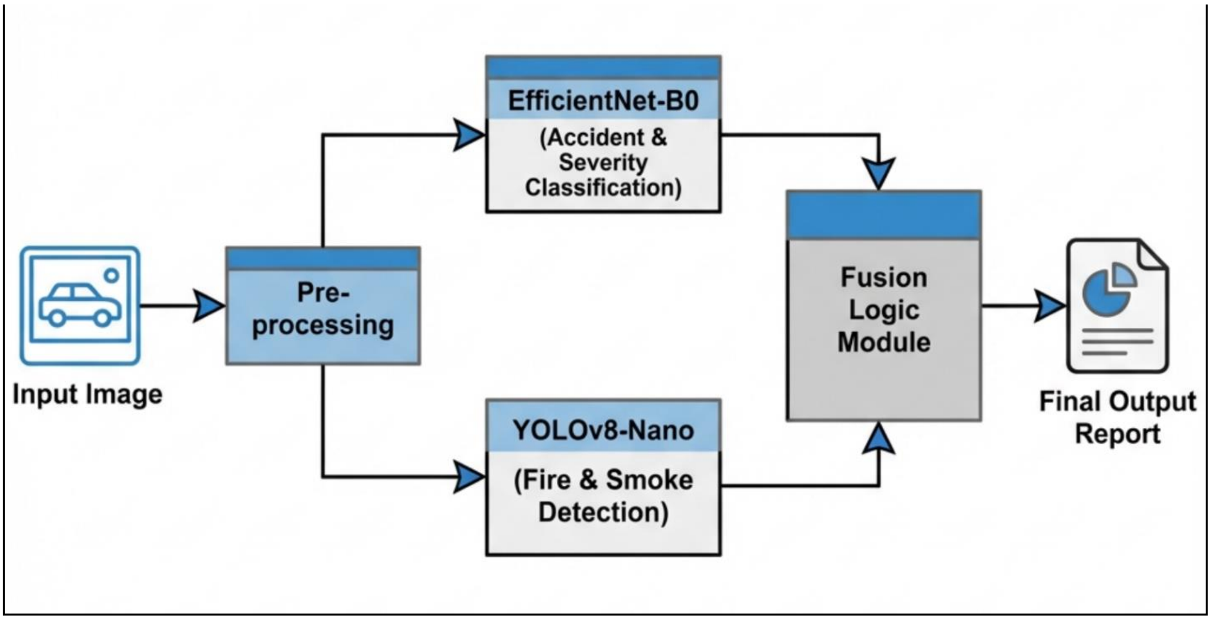

Figure 1. General Diagram of the System

2.2. CBAM Attention Module

One of the key innovations of this research is the integration of the Convolutional Block Attention Module (CBAM) into the EfficientNet structure. CBAM is a hierarchical and lightweight module consisting of two sequential sub-modules: “Channel Attention” and “Spatial Attention”.

- Channel Attention**:** This part determines the importance of each feature channel using Global Average Pooling and Global Max Pooling operations. By compressing spatial information, the model learns which channels contain more meaningful information about target features (such as vehicle damage) and suppresses less important channels.

- Spatial Attention**:** After adjusting channel weights, this stage focuses on the spatial position of objects. By applying ٧× ٧convolutions on the feature maps resulting from averaging and max-pooling, a spatial attention mask is generated, highlighting important regions of the image (such as collision points and body deformation). Combining these two attentions allows the model to focus on key features and ignore background noise.

2.3. Modified EfficientNet-B0 Architecture

The selected base architecture is EfficientNet-B0, which is suitable for real-time applications due to its excellent balance between accuracy and the number of parameters (approximately 5.3 million parameters). In the modified version, the CBAM module is placed after the MBConv blocks in different layers of the network. This connection ensures that before transferring features to deeper layers, channel importance and spatial regions are refined. The final output of the EfficientNet-B0 backbone is a 1280-dimensional feature vector. This feature representation is passed through shared layers, including Linear, Batch Normalization, and Dropout layers, before being divided into two specialized branches for accident detection and severity classification.

2.4. Eight-Stage Training Strategy

To achieve optimal performance and prevent overfitting, an eight-stage training pipeline was designed, divided into two main phases:

1. Pre-training Phase (Four Stages): In this phase, the backbone is first trained on auxiliary datasets, Kaggle Car Damage and YOLO Deformation, to learn general vehicle damage features. Then, it is fine-tuned on the main Accident Images Analysis dataset with lightweight classification heads. Finally, training is performed on a combination of all three datasets to increase generalizability.
2. Main Training Phase (Four Stages): In this phase, the full architecture with active CBAM is utilized. Steps include multi-task training with partial unfreezing of backbone layers (Stage 1-A and B), specific fine-tuning for the severity task (Stage 2), and finally, fine-tuning for the accident detection task (Stage 4). This gradual approach allows acquired knowledge to be preserved while gradually optimizing the model for specific tasks.

4-5. Managing Class Imbalance

The main problem in the dataset used was severe class imbalance, with the “low-risk” class constituting only about 5% of the data. To address this challenge, a three-part combined approach was adopted:

1. External Data Augmentation**:** Images of similar classes (such as Minor Damage) were extracted from Kaggle and YOLO datasets and added to the training set to increase the diversity and volume of minority class samples.
2. Sample Duplication**:** After adding external data, the sample duplication technique was used to equalize the distribution of classes, ensuring an equal number of samples for each class.

Weighted Sampling**:** During training, this technique was used to increase the probability of selecting minority class samples in each Epoch, forcing the model to better learn the patterns of these classes.

3. Findings and Results

In this section, the results of experiments conducted to evaluate the performance of the proposed system are presented. The goal of this evaluation is to measure the model’s accuracy in detecting accidents and classifying their severity, investigate the impact of the CBAM attention mechanism on minority class performance, and compare results with previous methods. All experiments were performed on Google Colab with an NVIDIA T4 GPU.

3.1. Datasets

The primary dataset used in this research was the Accident Images Analysis dataset, which contains real-world vehicle accident scene images collected from Iranian road accidents. This dataset was selected because it includes local vehicle types and covers three accident severity levels: low-risk, moderately dangerous, and dangerous. The class distribution is highly imbalanced; specifically, the low-risk class contains only 83 samples (approximately 5% of the total data), while the moderately dangerous and dangerous classes contain 1,122 and 857 samples, respectively. This imbalance represents the primary challenge addressed in this study.

To improve feature generalization and mitigate the minority-class imbalance problem, three auxiliary datasets were additionally utilized during the pre-training stage. The first auxiliary dataset was the Car Damage Severity Dataset, which contains vehicle images categorized according to damage severity levels. The second dataset was the Car Accidents and Deformation Dataset (Annotated), which includes accident images with YOLO-format annotations representing varying deformation intensities. Finally, a Fire and Smoke dataset was used to support the accident detection branch by improving the model's ability to recognize fire and smoke patterns commonly associated with severe accidents.

The use of these auxiliary datasets enabled the model to learn more robust visual representations under diverse environmental conditions, damage levels, and accident scenarios prior to fine-tuning on the primary dataset.

3.2. Implementation Parameters

Model implementations were performed using PyTorch version 2.0. The Adam optimizer was used for weight optimization with an initial learning rate of 0.001 and the ReduceLROnPlateau scheduler. The input image size for EfficientNet-B0 was set to 224×224 pixels, and for the YOLOv8-Nano module, it was set to 640×640 pixels. The batch size was set to 8, and the maximum number of training epochs was 100 with an early stopping mechanism to prevent overfitting. A fixed random seed of 42 was considered to ensure result reproducibility.

3.3. Evaluation Metrics

To accurately assess model performance, Accuracy, Precision, Recall, and F1-Score metrics were used. Given the class imbalance problem, Macro-F1 was considered the primary evaluation index, as this metric gives equal weight to all classes and prevents model bias towards majority classes.

3.4. Comparison of Base Architectures

Before introducing the final model, 17 different architectures from the ResNet, VGG, and EfficientNet families were evaluated for the accident detection task. Results are presented in Table 1.

Table 1. Performance Comparison of Different Architectures in Accident Detection

Architecture	Parameters (Million)	Test Accuracy	F1-Score
ResNet-18	11.2	92%	0.92
VGG-19	139.6	93%	0.93
EfficientNet-B0	4.8	90%	0.90
EfficientNet-B7	63.8	70%	0.70

As observed, although VGG-19 achieved the highest accuracy, EfficientNet-B0 provided a more suitable balance between accuracy and computational efficiency, with significantly fewer parameters and faster processing speed. Therefore, it was selected as the base architecture for the proposed system.

3.5. Analysis of CBAM Impact

To evaluate the effectiveness of the CBAM attention mechanism, the performance of the EfficientNet-B0 model with and without CBAM was compared for accident severity classification. The results are presented in Table 2.

Table 2. Impact of Adding CBAM Attention Mechanism on Model Performance

Configuration	Test Accuracy	Low-Risk Class F١	Macro F1
EfficientNet-B ٠(Without CBAM)	75.34%	0.29	0.56
EfficientNet-B ٠+ CBAM	72.40%	0.40	0.60

As shown in Table 2, although the overall accuracy decreased slightly, the F1-score of the minority class (low-risk) improved by approximately 38%. In addition, the Macro F1-score increased from 0.56 to 0.60, indicating better balanced performance across classes. These results suggest that the CBAM attention mechanism helped the model focus on subtle damage-related features and improved its ability to classify difficult minority classes.

3.6. Analysis of Class Imbalance Management Impact

The use of class imbalance management techniques, including external data augmentation and Oversampling, led to a significant improvement in performance in minority classes. After applying these methods and multi-stage training pipeline, the final model was able to increase its Macro F1 from 0.43 (base state) to 0.61. This improvement demonstrates the success of the proposed combined approach in overcoming the challenge of insufficient low-risk class samples.

3.7. Comparison with Previous Works

Finally, the results of the proposed system were compared with the method by Pashaei et al. (2020), which used the same dataset. Given that the public version of the dataset has lower quality than the reference article's original version, the comparison must be made cautiously. However, after re-implementing Pashaei's method on the public version, the following results were obtained:

Table 3. Comparison of Proposed System Results with Pashaei et al. (2020) (on Public Data Version)

Method	Severity Classification Accuracy	Macro F1
Pashaei et al. (2020) - Original Version	91.50%	0.74*
Pashaei et al. (2020) - Public Version	71.19%	0.565
Proposed System (This Research)	74.00%	0.61

*Note: The Macro F1 value for Pashaei's method in the original version was not reported and is estimated.

As seen in Table 3, the proposed system, despite using lower-quality images, was able to provide better performance than Pashaei's method implemented on the same data. A 4.49% improvement in Macro F1 indicates the effectiveness of the proposed architecture and the eight-stage training strategy under challenging conditions.

4. Discussion and Conclusion

The findings of the present study demonstrated that the proposed intelligent vehicle accident evaluation system, based on the EfficientNet-B0 architecture enhanced with the CBAM attention mechanism, achieved high performance in both accident detection and accident severity classification. The results showed that the final system reached 94% accuracy in accident detection and 74% accuracy in severity classification, while obtaining a Macro F1-score of 0.61. More importantly, the integration of the CBAM attention module improved the F1-score of the minority low-risk accident class by approximately 38% compared with the baseline EfficientNet-B0 architecture. Furthermore, the implementation of the multi-stage training strategy, combined with external data augmentation, oversampling, and weighted sampling, substantially improved model balance across classes and increased Macro F1 performance. The comparison with previous approaches revealed that the proposed system outperformed the benchmark model developed by Pashaei et al. in terms of classification effectiveness under the same public dataset conditions [6].

The first major finding of this study concerns the effectiveness of the CBAM attention mechanism in improving accident severity classification. The observed improvement in minority-class performance suggests that attention-guided feature learning enabled the model to focus more effectively on visually informative regions associated with vehicle damage. Accident scenes frequently contain complex backgrounds, environmental noise, shadows, lighting variations, and irrelevant objects that can distract conventional convolutional networks. By applying both channel and spatial attention, CBAM directs computational resources toward important features while suppressing irrelevant information. This interpretation aligns with the original theoretical framework proposed by Woo et al., who demonstrated that CBAM enhances feature representation by sequentially emphasizing informative channels and spatial locations [10]. The findings also support recent studies showing that attention mechanisms improve visual recognition performance under challenging conditions. For example, Yun et al. reported that attention-enhanced detection architectures significantly improved the identification of small and difficult visual targets in complex environments [11]. In the context of accident severity classification, the ability to focus on subtle deformation patterns, scratches, collision points, and structural damage appears to have contributed substantially to the improved classification of low-risk accidents.

Another important result was the superiority of the EfficientNet-B0 architecture as the backbone model for the proposed system. Although larger architectures often achieve competitive performance, EfficientNet-B0 provided an optimal balance between computational efficiency and predictive accuracy. This outcome is consistent with the design philosophy introduced by Tan and Le, who demonstrated that compound scaling allows EfficientNet models to achieve strong performance while maintaining relatively low computational costs [7]. The suitability of EfficientNet for real-time accident detection applications is particularly important because practical deployment scenarios frequently involve mobile devices, surveillance systems, and emergency response infrastructures with limited computational resources. Therefore, the present findings reinforce previous evidence that efficient deep learning architectures can support high-performance computer vision applications without requiring excessive processing power [7].

The results also demonstrated the critical role of addressing class imbalance in accident severity prediction. Before implementing imbalance-management strategies, the model exhibited substantially lower performance on minority classes. However, after combining external data augmentation, oversampling, weighted sampling, and staged training procedures, classification balance improved considerably, resulting in a significant increase in Macro F1 performance. This finding is consistent with broader machine learning literature indicating that class imbalance represents one of the most important barriers to reliable classification performance in safety-critical applications. In accident datasets, severe and moderate crashes typically dominate available samples, causing models to prioritize majority classes while neglecting underrepresented categories. The present findings suggest that effective data management strategies can mitigate this bias and enable more equitable learning across severity levels. These results support the observations of Benfaress et al., who emphasized that model architecture alone cannot compensate for deficiencies in class representation and that data-level interventions remain essential for robust severity prediction systems [9].

The observed improvement in low-risk accident recognition deserves particular attention because minor accidents often present the greatest classification challenge. Severe accidents typically involve extensive vehicle deformation, visible destruction, fire, or other highly salient indicators. In contrast, low-risk accidents may involve only limited structural damage that can easily be overlooked by conventional models. The substantial improvement in F1-score achieved in the present study suggests that the combined effects of CBAM attention and class-balancing strategies enhanced the model's sensitivity to subtle visual patterns. This finding has important practical implications because accurate differentiation between minor and severe accidents can facilitate more efficient allocation of emergency resources and reduce unnecessary dispatching of high-priority response units.

The comparison with previous research further supports the effectiveness of the proposed framework. Pashaei et al. developed one of the most influential accident image classification systems by combining convolutional neural networks with mixtures of extreme learning machines [6]. Although their model achieved promising results, it relied on earlier-generation convolutional architectures and did not incorporate attention mechanisms or advanced imbalance-management strategies. The superior performance achieved in the present study indicates that recent developments in attention-guided deep learning provide meaningful advantages for accident severity classification. These findings reflect the broader evolution of computer vision technologies, where attention mechanisms and optimized architectures increasingly outperform traditional convolutional approaches across diverse domains.

The results also align with contemporary studies investigating accident severity prediction using transfer learning and explainable artificial intelligence techniques. Aboulola reported high classification performance using MobileNet-based transfer learning models combined with SHAP interpretability methods [8]. Similarly, Benfaress et al. demonstrated the effectiveness of deep learning frameworks integrated with explainability approaches for accident severity prediction [9]. While the primary focus of the present study was performance enhancement rather than interpretability, the findings support the general conclusion that modern deep learning architectures can provide substantial improvements in accident assessment capabilities. The success of EfficientNet-B0 combined with CBAM suggests that attention-based feature extraction may complement future explainable artificial intelligence frameworks by improving the quality of learned representations before interpretability analyses are performed.

Another notable aspect of the findings concerns the integration of fire and smoke detection capabilities within the overall accident assessment framework. Severe traffic accidents frequently involve secondary hazards such as vehicle fires and smoke emissions, which can significantly increase risk levels. The inclusion of a YOLOv8-based detection module therefore extends the functionality of the proposed system beyond conventional severity classification. This design choice is consistent with recent research emphasizing the importance of multi-task learning architectures in intelligent transportation systems. Namdeo and Deshmukh highlighted the value of integrating accident detection, severity assessment, and emergency communication within a unified framework to support rapid response operations [17]. Similarly, the successful use of fire-related datasets in the present study reflects findings from FFYOLO research demonstrating the effectiveness of lightweight detection architectures for identifying hazardous visual events [11, 16].

The broader significance of the present findings can also be understood within the context of increasing smartphone penetration and crowdsourced data generation. Billions of individuals worldwide now possess smartphones capable of capturing high-resolution accident images and transmitting them instantly through mobile networks [4]. Previous research has demonstrated that crowdsourced traffic data can significantly reduce accident reporting delays [5]. The proposed system extends this concept by enabling automated visual analysis of accident scenes rather than relying solely on textual reports or manual assessment. Consequently, the integration of mobile imaging technologies with intelligent computer vision systems may create powerful opportunities for improving emergency response effectiveness and reducing accident-related mortality.

From a theoretical perspective, the findings contribute to ongoing discussions regarding the role of attention mechanisms in complex visual classification tasks. The substantial improvement observed in minority-class performance supports the hypothesis that selective attention facilitates more discriminative feature extraction under conditions of high visual complexity. Furthermore, the success of the multi-stage training framework suggests that learning strategies may be equally important as architectural innovations when addressing practical challenges such as class imbalance. Together, these results indicate that future advancements in accident severity classification are likely to emerge from integrated approaches that simultaneously address feature representation, data quality, learning dynamics, and computational efficiency.

The findings additionally highlight the growing convergence between intelligent transportation systems and advanced artificial intelligence technologies. Contemporary transportation research increasingly incorporates graph neural networks, transformers, and spatio-temporal modeling approaches for traffic prediction and infrastructure optimization [12, 13]. The present study complements these developments by demonstrating how attention-enhanced computer vision systems can contribute to another critical aspect of transportation management, namely accident assessment and emergency response support. Collectively, these advances suggest that future transportation ecosystems will become increasingly data-driven, automated, and responsive to real-time conditions.

Overall, the results demonstrate that combining EfficientNet-B0, CBAM attention mechanisms, multi-stage training strategies, and class imbalance management techniques can substantially improve vehicle accident detection and severity classification performance. The proposed framework successfully addressed several limitations commonly encountered in accident image analysis and achieved competitive results under realistic data conditions. These findings reinforce the value of attention-guided deep learning architectures for safety-critical applications and provide a foundation for future developments in intelligent accident management systems.

One limitation of the present study concerns the quality and distribution of the available dataset. Although multiple data augmentation techniques were employed, the low-risk accident class remained relatively underrepresented compared with other categories. In addition, the public version of the dataset may not fully reflect the diversity of real-world accident scenarios. Variations in weather conditions, lighting environments, camera quality, and vehicle types could influence model performance when deployed in operational settings. Furthermore, the study relied primarily on static images, preventing the incorporation of temporal information that may be available in video recordings.

Future research should investigate the integration of video-based accident analysis frameworks capable of capturing temporal dynamics associated with collision events. Researchers may also explore transformer-based architectures, hybrid CNN-transformer models, graph neural networks, and multimodal systems that combine image data with sensor information, GPS signals, and traffic metadata. Expanding datasets to include more diverse environmental conditions, balanced severity distributions, and international accident scenarios would further enhance model generalizability. In addition, future studies could incorporate explainable artificial intelligence methods to improve transparency and facilitate trust among emergency response professionals.

From a practical perspective, the proposed framework could be integrated into intelligent transportation systems, mobile emergency applications, traffic surveillance networks, and smart city infrastructures. Emergency management organizations may use such systems to accelerate accident assessment and optimize resource allocation based on predicted severity levels. Insurance companies could employ automated damage evaluation tools to support claims processing, while transportation authorities could utilize real-time accident analytics to improve traffic management and incident response coordination. The deployment of lightweight, attention-enhanced models on edge devices may further enable rapid and scalable implementation across diverse operational environments.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

[1] World Health Organization, "Global Status Report on Road Safety 2023," World Health Organization, Geneva, Switzerland, 2023.

[2] Forensic Medicine Organization of Iran, "Statistics of Traffic Accident Fatalities in 2023," Forensic Medicine Organization of Iran, Tehran, Iran, 2024.

[3] J. M. Darley and B. Latané, "Bystander Intervention in Emergencies: Diffusion of Responsibility," (in English), *Journal of Personality and Social Psychology,* vol. 8, no. 4, Pt. 1, pp. 377-383, 1968, doi: 10.1037/h0025589.

[4] Statista, "Number of Smartphone Users Worldwide from 2017 to 2029," Statista, 2024.

[5] S. D. Young, W. Wang, and B. Chakravarthy, "Crowdsourced Traffic Data as an Emerging Tool to Monitor Car Crashes," (in English), *JAMA Surgery,* vol. 154, no. 8, pp. 777-778, 2019, doi: 10.1001/jamasurg.2019.1167.

[6] A. Pashaei, M. Ghatee, and H. Sajedi, "Convolution Neural Network Joint with Mixture of Extreme Learning Machines for Feature Extraction and Classification of Accident Images," (in English), *Journal of Real-Time Image Processing,* vol. 17, no. 4, pp. 1051-1066, 2020, doi: 10.1007/s11554-019-00852-3.

[7] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 6105-6114.

[8] O. I. Aboulola, "Improving Traffic Accident Severity Prediction Using MobileNet Transfer Learning Model and SHAP XAI Technique," (in English), *PLOS ONE,* vol. 19, no. 4, p. e0300640, 2024, doi: 10.1371/journal.pone.0300640.

[9] I. Benfaress, A. Bouhoute, and A. Zinedine, "Enhancing Traffic Accident Severity Prediction Using ResNet and SHAP for Interpretability," (in English), *AI,* vol. 5, no. 4, pp. 2568-2585, 2024, doi: 10.3390/ai5040124.

[10] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3-19, doi: 10.1007/978-3-030-01234-2_1.

[11] B. Yun, Y. Zheng, Z. Lin, and T. Li, "FFYOLO: A Lightweight Forest Fire Detection Model Based on YOLOv8," (in English), *Fire,* vol. 7, no. 3, p. 93, 2024, doi: 10.3390/fire7030093.

[12] J. Zhang, Y. Yang, X. Wu, and S. Li, "Spatio-temporal transformer and graph convolutional networks based traffic flow prediction," *Scientific Reports,* vol. 15, no. 1, p. 24299, 2025, doi: 10.1038/s41598-025-10287-5.

[13] A. Zhang, "Dynamic graph convolutional networks with temporal representation learning for traffic flow prediction," *Scientific Reports,* vol. 15, no. 1, p. 17270, 2025, doi: 10.1038/s41598-025-01696-7.

[14] P. Bhamere. *Car Damage Severity Dataset*, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/prajwalbhamere/car-damage-severity-dataset

[15] M. A. Arshad. *Car Accidents and Deformation Dataset (Annotated)*, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/marslanarshad/car-accidents-and-deformation-datasetannotated

[16] L. DataCluster. *Fire and Smoke Dataset*, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/dataclusterlabs/fire-and-smoke-datas

[17] P. Namdeo and K. Deshmukh, "A Computer Vision-Based Algorithm for Accident Detection, Severity Classification and Rapid Emergency Response Using YOLOv8 and Twilio Programmable Messaging API," in *OTCON 2024 Conference Proceedings*, 2024.