Traffic Accident Severity Classification Using ResNet-18

Theme: Digitalization and Technological Innovations in Sustainable Engineering.

Oni *Oluwadamilola* [1,2], and Ogbonnia *Emmanuel* [1].

Department of Computer Science, Landmark University, Omu-Aran, Kwara State, Nigeria. [1]

Email: oni.damilola@lmu.edu.ng [2]

Abstract: The paper describes a system of automated classification of the severity of traffic accidents based on a fine-tuned ResNet-18 architecture. To meet the requirement of quick emergency reaction in low-resource urban areas, we pre-selected a dataset of about 50000 images, integrating the CADP and UCF-Crime datasets with rescue images of the area. The images were processed and augmented in order to rectify the imbalance of the classes. Transfer learning was used to train the model using Focal Loss and AdamW optimizer. Testing on a held-out test set shows a total accuracy of 98.54 and a precision of 0.99 on severe incidents and an F1-score of 0.98. The system maximizes the edge deployment, which provides low-latency inference applicable to real-time municipal surveillance. Comparative analysis shows that ResNet-18 offers a superior trade-off between accuracy and computational efficiency compared to deeper architectures..

Keywords: Traffic Accident Detection, Severity Classification, ResNet-18, Emergency Response, Deep Learning, Edge AI, Intelligent Transportation Systems.

1. Background to the Study

It is important to effectively estimate how critical a traffic accident is in order to fully utilize emergency response and resources deployment in urban settings. Severity classification refers to the automatic classification of accident scenes, frequently recorded by roadside cameras or in-car sensors into some predefined levels (e.g. moderate and severe), can allow first-responders to allocate more resources to incidents which represent the most significant threat to life and property. The conventional methods are based on the human interpretation of the crash reports or the manual observation of the images which implies some delays and subjective prejudice. Recent progress in convolutional neural networks (CNNs) and especially in ResNet-18 has shown impressive potential to extract hierarchical characteristics in visual data rendering it suitable in real-time image-based severity classification [1] [2].

Researchers have achieved high precision and recall by finetuning a ResNet-18 model that is pre-trained on large-scale datasets with various lighting and weather conditions. The features are extracted out e.g. patterns of deformation on the car bodies or the spread of debris allow the classifier to determine the probable extent of damage to the property without human effort [3]. Similar studies also explored the use of hybrid CNN-LSTM models to process data from smartphone sensors such as gyroscope, accelerometer and GPS to detect possible accidents and automatically alert hospitals [4]. When incorporated into the wider intelligent transportation system, such a system can greatly decrease latency, allowing ambulance and fire services to be deployed in relation to the severity of the incidents. Finally, an accurate automated severity classification pipeline adds to safer roads and more effective emergency management in smart cities.

Although significant advances have been made in detecting traffic incidents, severity classification is still an under-studied and technically difficult issue, especially in low-resource environments. Present emergency response procedures tend to prioritize all the reported accidents in equal measure which results in poor utilization of limited resources: those minor collisions causing fender-bender can cause full-scale dispatches, whereas serious, life-threatening incidents may be under-attended because of reporting bottlenecks. More recent models, including ARMS a three-stream architecture that integrates scene-level CNN, object-level ConvLSTM2D, and I3D motion detection have been found to be very accurate and resilient in low-visibility situations [5]. On the same note, a video pipeline based on AI has been created to identify and track vehicle collisions in urban traffic video with real-time computer-vision and edge AI, with an emphasis on low-latency detection to be used on city surveillance [6]. Another hierarchical framework combines Kalman-Hungarian tracking, trajectory conflict, and YOLOv4 classification and achieves 93.10% detection but with a false alarm ratio of 6.89%, which once again indicates the efficiency of the modern computer vision solution in traffic conditions [7]. Manual severity scoring, be it eye-witness, or operator scrutiny of CCTV footage, is subject to variability and latency. Under stress, witnesses are likely to misinterpret the dynamics of the crash, leading to an overestimation or underestimation of injuries. In the same manner, human observers of video streams become tired and have short attention spans, heightening the chances of misclassifying under high-throughput surveillance [4].

Developing countries such as Nigeria face additional structural barriers. Poor camera coverage, periods of power outage, and insufficient computational infrastructure are barriers to implementing advanced AI models. Also, the cost of a commercial severity classification solutions is relatively high and as such, it cannot be afforded by municipal authorities that have limited budgets. These reasons make it difficult to implement automated severity classification system on a large scale and causes a continued circle of delayed emergency actions and avoidable losses. Even though recent reviews on over a hundred studies of action-recognition methods of traffic accidents demonstrate the efficacy of 3D-CNN structures, they also show discrepancies in metrics used to evaluate them, and suggest the use of standardized multi-view datasets to enhance the generalization [8]. Through low-cost options, including systems that use smartphone sensors (accelerometer, GPS, microphone and barometer) and threshold-based algorithms, low-cost systems have been found to be 89% received and false-positive rate was 5% with potential of being deployed into developing regions [9]. Likewise, systematic discussions of the AI- and ML-based traffic collision prediction models note the impact of the driver behaviour, road design, and weather on the model, and support location-specific and privacy-preserving (federated) learning models [10]. Absence of uniform and scalable real-time severity-grading methodologies remains a barrier to interagency coordination: traffic management centres, hospitals and law enforcement agencies usually do not have a shared measure of severity to be used. This means that incident triage cannot be done efficiently by response teams, which results into inefficient outcomes and more deaths.

The solutions to these issues require a powerful, light-weight severity classification model capable of being applied successfully on edge devices, with low processing capabilities, and unstable connectivity. This model is anticipated to be generalized to other urban scenes, other camera angles and partial occlusions; it is expected to be interpretable in a manner that the stakeholders can rely on it. It must also be inexpensive, making use of the open-source architecture and the open data in order to minimize the financial constraints. The proposed research is aimed at isolating and simplifying the process of feature extraction and decision that is central to reliable grading and, by consequence, facilitate the deployment of resources according to severity of an incident by emergency services and ultimately save life by merely targeting the severity classification that depends on the stage of detection.

1.1 Aim and Objectives

This study will be aimed at creating and testing a ResNet-18 based system to classify the severity of traffic accidents in urban areas. The objectives are specific and they are as follows:

- To curate and preprocess a dataset of accident-scene images labeled by severity tier.
- To optimize a ResNet-18 model to do binary classification (moderate vs. severe).
- To determine the model performance in terms of precision, recall, F1-score and confusion matrix analysis.

1.2 Significance of Study

This study makes a particular contribution to the area of intelligent transportation systems and optimal emergency response by separating and automating the severity classification task. The earlier methods such as shallow CNN, hybrid CNN-LSTM, and multi-streams failed to either be computationally expensive, overfit to small datasets, or had long latency that prevented them to be used in real-time on edge devices. The models used were also found to be inadequate in resilience to data imbalance, occlusions or varying urban conditions with varied performance in moderate and severe accident classes. In comparison, Construction ResNet-18 based classifier uses the residual connections to get deep feature learning without vanishing gradients and allowing extraction of fine-grained visual features such as vehicle deformation patterns and spread of debris without excessive computational complexity costs. The model is a lightweight architecture capable of providing a high accuracy of real-time grading of accidents, even in problematic urban conditions, and can be deployed on edge hardware with limited memory and processing capability. The system can easily be linked with the current camera systems by authorities and first responders to automatically flag high-severity accidents, dispatch prioritization, and decrease the mortality rates. In addition, the open-source paradigm, with openly accessible datasets and open-source frameworks, reduces the adoption costs in developing countries. Scholarly, the work contributes to the existing research on the class-imbalanced deep classification, hard-example mining, and edge-optimized model deployment, and it proves a practical and scalable solution to the real-world intelligent transportation systems [3][4].

1.3 Scope of Study

The specific issue of this research is severity associated with traffic accident images, it concerns the dichotomous categorization of accidents into moderate and severe traffic crashes on urban roads. The work specifically does not include detection or localization of accidents, which is presupposed to be done by the currently existing object-detection pipelines, so that the work can focus on precise and effective severity grading. The pre-cropped images of the accident-region are model-developed, trained and evaluated, so that the classifier is trained on fine-grained visual information that is applicable in assessing a damage, i.e., deformation of the vehicle and patterns of the debris. Evaluation testing at the edge-computing devices that are indicative of municipal deployments is carried out with a focus on low-latency and real-time applicability in resource limited environments. With this narrowing of the scope, the study can isolate the severity classification task, and thus, the severity classification task can be rigorously evaluated in performing its performance, and thus, offer a useful framework upon which the study can be integrated within the larger intelligent transportation and emergency response systems.

2. LITERATURE REVIEW

Automated severity of all traffic accidents has become an essential part in the intelligent transportation systems in that emergency services can prioritize the incidents in regard to the probable human and material damage. In contrast to larger accident detection problems, which tend to enquire whether an accident has taken place, severity classification attempts to provide each incident reported with a level of seriousness which is often described by levels like minor, moderate, and severe accident. Initial studies in this area used handcrafted features, obtained on still images, including vehicle deformation measures and scene context, which were impaired by poor generalizability to a wide variety of urban settings. The recent surge of deep convolutional neural networks (CNNs) has changed the game: CNN-based classifiers can infer latent signals such as crumple-zone collapse or debris distribution that are strongly related to the severity levels [1].

Especially, ResNet-18 has become popular due to its good representational strength but low computational cost. Its gradients that do not vanish during training help deep architectures to converge more strongly on visual classification tasks. ResNet-18 when fine-tuned on traffic accident data has demonstrated a high level of accuracy in the differentiation between moderate and severe collisions, even in problematic lighting and weather conditions [3]. Research has indicated that severity classifiers constructed on ResNet variants consistently surpass shallower CNNs, because they are better feature-extractors. ResNet-18 has consequently formed the foundation of much of the more recent implantations, acting as a reference point in scholarly work as well as testbed pilot implementations in smart-city testbeds.

2.1 Datasets and Annotation Standards

One of the most basic problems of severity classification research is the lack of large publicly available datasets that are annotated with severity labels. The majority of the existing repositories are based on detection that gives bounding boxes on the vehicles and pedestrians without the degree of damage or injury risk. To fill this gap, a number of research papers have created special datasets through the manual labeling of images of accidents gathered using dash-cams, traffic-cameras, and open-source traffic surveillance archives. An example is the data set of more than 5,000 annotated accident scenes that Pathik et al. [4] compiled by classifying each as a moderate or severe accident in terms of apparent deformation and contextual indicators (in the form of secondary collisions).

Protocols of annotation differ across research groups, although dual-annotation (two independent labelers) with expert adjudication to arbitrate on disagreements is common. Tiers of severity may typically be set by a measurable criterion, like the cost to repair, or the number of airbags deployed, or by proxy visual markers such as degree of collapse in the vehicle frame [5]. Data imbalance is heavily crucial: serious accidents are less frequent than minor collisions, and this causes skewed distributions among the classes. In order to alleviate this, researchers use data-augmentation methods (rotation, scaling, brightness jitter) and resampling methods (stratified sampling, synthetic minority over-sampling) when training the models [2]. Nonetheless, the different cases of urban scenes with varying angles of the camera, occlusions, and weather conditions are still a challenge to model generalization and hence the importance of more extensive and standardized datasets.

2.2 Core CNN-based Classification Techniques

Since the publication of AlexNet, the different CNN architectures have been implemented on the severity classification with more sophistication. Early-generated models reused detection backbones, including VGGNet and Inception, to categorise pre-cropped accident regions, with moderate success and large computational costs [6]. The introduction of real-time detection pipelines (e.g. YOLOv8-s) allowed isolating crash-affected regions and further classifying them by lightweight CNNs. Two-stage methods (detection then severity-classification) were found to permit specialization: the former stage is best at recall (noting every accident) and the latter at precision (grading the severity).

The choice of the ResNet-18 classifier as a classifier of choice in this paradigm was motivated by its residual learning ability. Researchers can use rich hierarchies of features by training the last layers of a ResNet-18 mindreading ImageNet without training the early convolutional layers extensively [1]. Transfer learning lowers the requirements of large labeled datasets and quickens convergence. Comparative experiments show that pure ResNet-18 can obtain better than 0.95 F1-scores on binary severity classification tasks, and is significantly smaller than other more resource-intensive models such as ResNet-50 or DenseNet [3].

2.3 Transfer Learning and ResNet-18 Applications

Transfer learning has also become a necessity in the field of classification of severity, as in the absence of labeled datasets on accidents-severity. Training classifiers with network weights pre-trained on large-scale image corpora, learners are given powerful low-level feature detectors edge, texture, and color gradients, which are extrapolated to accident imagery. It only needs retraining of the last fully-connected layers in order to map feature embeddings to severity labels [2]. The architecture of ResNet-18 with a depth of 18 layers is a sensible compromise: it is deep enough to learn intricate patterns but shallow enough to run on edge devices with limited resources in terms of memory and compute.

Experimental analyses demonstrate that freezing early convolutional blocks and reretraining all layers other than the residual and classification layers can achieve near state of the art performance and decrease training time by more than half. Misclassified severe examples are re-weighted dynamically by hard-example mining, which further increases robustness to subtle damage cues [4]. These strategies solve the problem of imbalance in classes and make sure that the classifier itself is not biased towards the more dominant moderate-severity one. Practically, a ResNet-18 based classifier, with label smoothing and focal loss functionality, achieves precision and recall scores of more than 0.90 at both classes, and is a substantial enhancement over non-fine-tuned baselines.

2.4 Model Optimization and Edge Deployment

Severity in real-time requires inference on edge hardware in low latency as well as accuracy. Traditional floating-point CNNs have prohibitive memory and computational requirements that make it impossible to run them on roadside units or on-car processors. To address this, scientists have experimented with post-training quantization of 32-bit weights and activations to 8-bit integers reducing model size by approximately 75 percent and speeding up their integer-math operations on integer arithmetic processors [3].

Compression techniques, such as pruning, which can eliminate low-importance filters and channels offer a complementary means of compression reduction, with more significant cuts in model footprint incurred at little cost in accuracy [1]. Together, these optimization techniques allow applying severity classifiers in larger accident detection pipelines, which will facilitate on-the-fly grading of incident severity without depending on centralized cloud services.

2.5 Summary of Gaps and Future Directions

Although major steps have been achieved in CNN-based severity classification, a few research gaps remain. First, the available datasets are still small in terms of scale and variety; further enlarged collection of different urban settings, darkness, and bad weather should be considered to enhance generalization. Second, multi-class severity grading beyond binary moderate/severe differences is under explored, but might allow resource allocation to be more precise (e.g., whether helicopters or ground ambulances need to be deployed). Third, incorporation of temporal information using video sequences, but not snapshots, can pick up temporal features like collisions as they unfold that could aid predictive accuracy.

Federated learning methods should also be investigated in the future to learn severity classifiers in the collaboration of multiple municipal agencies, which do not share centralized data and retain privacy in addition to augmenting model robustness. Lastly, pilot studies in the real world are necessary to certify the performance of the classifiers under the conditions of operation and to optimize human AI processes that would strike the right balance between automation and human control.

3. Overview of Related Works

Recent developments in traffic incident analysis focus on the integration of complementary methods to work with complicated urban scenes. A hierarchical scheme combining

Kalman-Hungary tracking, trajectory conflict analysis, and YOLO based classification was proposed by Khan and Rao to deal with occlusions and dynamic interactions in surveillance video and demonstrate high success in detecting with low false alarms [7]. Based on the concept of multi-modes, Adewopo et al. proposed an I3D-ConvLSTM2D ensemble that integrates both RGB and optical-flow streams and can run on the Jetson-class devices, showing the benefit of spatial and time fusions to enhance recognition at a cost of edges [2]. Their systematic review also emphasizes the prevalence of 3D-CNNs in video tasks and makes a case to standardize the comparison and consistency across studies through standard benchmarks and multi-view datasets [8].

Other work focuses on low-cost sensing and system integration that scale in resource-limited settings. Bhatti et al. demonstrated a smartphone-based IoT solution using accelerometer, GPS, microphone, and barometer thresholds to detect collisions and alert hospitals, showing that legacy vehicles can be included without specialized hardware [9]. Pathik et al. combined smartphone sensor streams with CNN–LSTM models and hotspot cross-validation to cut false alarms and deliver near-real-time alerts, illustrating how mobile sensing complements vision systems [4]. At the infrastructure level, Balasubramanian et al. designed an adaptive traffic management system that clusters abnormal vehicle groupings and secures IoT data, reducing wait times and improving safety in simulations, though it depends on roadside infrastructure [11].

A third strand pushes model sophistication by adding complexity and thereby increasing resource needs for real-time use. Chen et al. combined modern detectors, MOSSE tracking, and GPT-based vision analysis to automate liability reporting, showing promise for insurance workflows but requiring further validation for complex multi-vehicle collisions [1]. Fu et al. proposed ARMS, a three-stream network that fuses scene-level, object-level, and motion cues to improve recognition in low visibility, but its higher compute demands call for hardware acceleration or model compression for live deployment [5]. Robles-Serrano et al. showed that hybrid spatial–temporal models generalize well across camera angles and lighting, yet their runtime performance on standard GPUs highlights the recurring tension between accuracy and latency [12].

Finally, survey and semantic approaches point to broader gaps and future directions. Niture et al. emphasize that driver behavior, weather, and road geometry strongly affect collision prediction and recommend location-specific models and privacy-preserving training such as federated learning to maintain cross-region robustness [10]. Adewopo et al. also explored semantic topic models that treat video clips as “documents” and accidents as abnormal topics to give more interpretable anomaly reports, although making these methods real-time remains a challenge [2]. Olugbade et al. also point to some of the ethical dangers such as detection bias and suggest hybrid AI-physics methods to enhance the interpretability and policy relevance [3]. Collectively, these papers propose a way forward that balances more detailed multi-modal models with edge aware optimization, larger standardized data and operational protection against real-world deployment.

4. METHODOLOGY

The critical importance of efficient emergency response and the reduction of loss to the victims require proper and prompt evaluation of the severity of traffic accidents. Conventional detection using eyewitness reports and manual emergency calls is fraught with delays, inconsistency and misuse of man power, which in most cases lead to the loss of golden minutes to respond to treatment. To overcome these shortcomings, this paper will implement a fully automated and real-time severity classification pipeline combining YOLOv8-small to detect accidents quickly and ResNet-18 to estimate damage in fine-grained detail and run it on an edge computing device to maintain sub-second decision-making.

4.1 System Architecture and Design Rationale

The system is built as a simple, fast pipeline that turns camera footage into immediate severity alerts. Video frames or pre-cropped regions of interest are first preprocessed: images are resized to the classifier's input size, color channels are normalized, and faces or license plates are blurred when required for privacy. These steps make the input consistent and reduce accidental bias from poor image quality. Next, each preprocessed crop is fed to a fine-tuned ResNet-18 severity classifier. ResNet-18 was chosen because it finds the right trade-off between accuracy and speed: it extracts visual cues like vehicle deformation and debris patterns to decide whether an incident is moderate or severe. The classifier outputs a single severity label for each crop.

To meet real-time requirements, the pipeline runs on edge hardware. The design is modular: preprocessing, classification, and alerting are separate components, so each can be updated or replaced independently without disrupting the whole system. This makes the solution easier to maintain, scale, and integrate into city traffic operations.

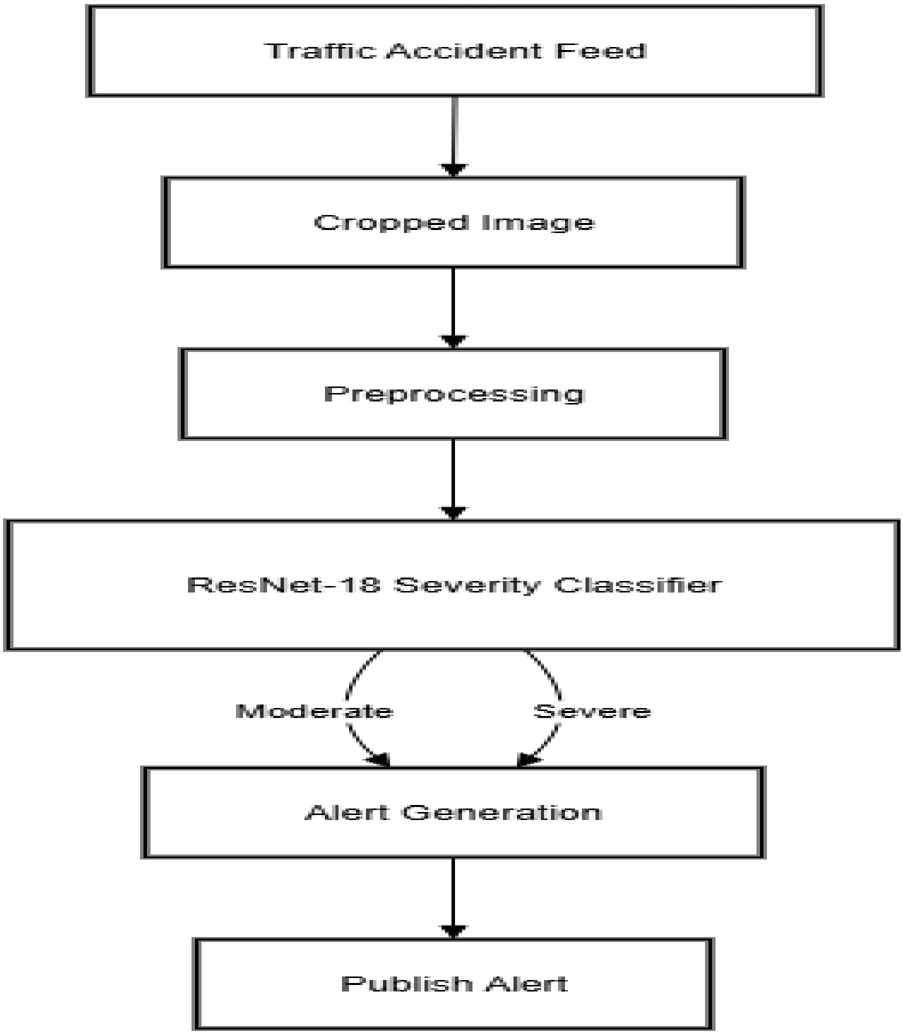

Fig. 1. System Architecture of the Severity Classification Model.

4.2 Data Collection and Annotation

A One of the prerequisites of stable model actions is a well-developed diversified dataset. We used a mix of the sources in order to build our severity classification dataset: a set of existing publicly available computer vision datasets (CADP and UCF-Crime), as well as proprietary image data obtained on the Roboflow Universe, and manually annotated images taken in urban Nigerian crossroads. The total number of annotated images is 42 120.

The image crops were categorized in two severity levels Moderate and Severe. The categories of severity were placed on a set of pre-established criteria, such as the visible amount of damages to the vehicles, the number of vehicles, the presence of debris, and the apparent risk of injury:

The data was divided into Training, Validation and Test set according to the allocation given by the source of data. To enable transparency and reproducibility, the percentages of the split and the counts of the images are outlined below:

Table 1. Dataset distribution table

Dataset Split	Percentage (%)	Image Count	Purpose
Training Set	82%	34,534	Model weight optimization
Validation Set	12%	5,048	Hyperparameter tuning and checkpoint selection
Test Set	6%	2,538	Final, unbiased performance evaluation

Fig. 2. Sample image of accident dataset

Fig. 3. Labelled image of accident severity.

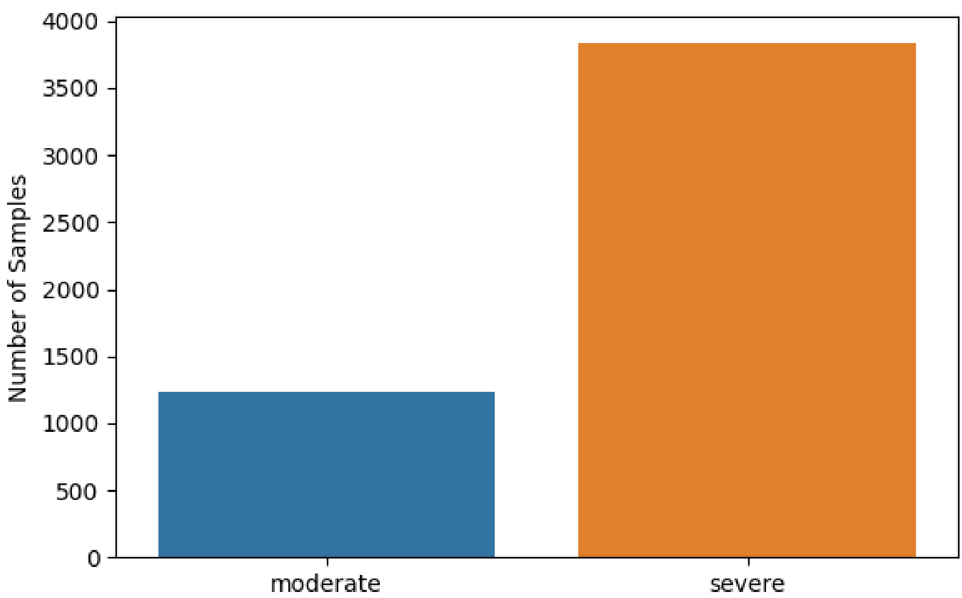

Fig. 4. Validation Set Class Distribution.

Figure 4 shows a sample distribution of the validation set by each severity class. The data is highly imbalanced with the severe group consisting of about 3782 samples and the moderate group having only 1213 samples. To overcome this imbalance, oversampling and focal loss were used in the training, which provided high F1-scores in both classes, and it can be concluded that the model was able to cope with the skewed distribution of classes.

4.3 ResNet-18 Severity Classifier Development

For To classify the severity, we used ResNet-18 because of its residual structure that enables the deep features learning without a high level of computational complexity. The final completely connected layer of the network was substituted by a softmax head that corresponds to the severity levels. Transfer learning: the only two highest ranked residual blocks and the classifier head were retrained, but the previous layers were frozen to retain general visual features. The cross-entropy loss with label smoothing (ϵ=0.1) was used to train the model and focus on the tough examples and especially those of the critical class, using the focal loss. The AdamW with a weight decay of 1 x10 -4 was optimized. The following training hyperparameters were established: the learning rate of 1 -4, the size of the batch of 32, and 25 epochs. Elaborations on the data augmentation were random rotation (+15), horizontal flipping, jitter on brightness and contrast and random cropping to enhance generalization. The data was divided into 82 percent training, 12 percent validation and 6 percent test sets and 42,120 annotated images in total.

4.4 Use Case Scenario

In a typical urban deployment, traffic-management cameras stream video to edge devices mounted roadside. Consider a late-night intersection: a sedan's abrupt lane change triggers dense clustering and overlap detected by YOLOv8-small across three frames. Within 200 ms, the system flags "accident=true" and extracts the ROI, which ResNet-18 evaluates as "critical" based on pronounced bumper collapse and widespread debris. A JSON alert, containing camera ID, GPS coordinates, timestamp, severity label, and blurred snapshot is published to the city's Traffic Operations Center and EMS dispatcher. Simultaneously, adaptive signal control reroutes cross-traffic to clear an emergency corridor. This closed-loop workflow from detection to dispatch demonstrates how automated severity classification can shave crucial seconds off traditional response times, ultimately saving lives.

4.5 Anticipated Challenges and Mitigation

Anticipated challenges include limited dataset diversity, persistent false positives, and occasional incorrect classifications. A narrow dataset can make the model fail on rare situations like night crashes, heavy rain, or unusual vehicle types, so we expand the data with targeted collection, synthetic augmentation, and active learning to capture new cases. False positives waste responders' time; we cut them down with simple fixes like requiring detections to persist across multiple frames, tuning confidence thresholds, and using quick secondary checks before sending alerts. Wrong classifications are handled by focusing training on hard examples, retraining regularly with newly collected edge cases, and adding explainability tools (e.g., Grad-CAM) so operators can see why the model decided what it did. Finally, logging, short-term local buffering, and secure over-the-air updates let us patch problems fast and keep improving the system from real-world feedback.

https://doi.org/10.1051/e3sconf/202668401002

5. RESULTS AND CONCLUSION

Classification Report:

	precision	recall	f1-score	support
moderate	0.9559	0.9854	0.9704	1231
severe	0.9953	0.9854	0.9903	3838
accuracy			0.9854	5069
macro avg	0.9756	0.9854	0.9804	5069
weighted avg	0.9857	0.9854	0.9855	5069

Fig. 5. Classification Report of the ResNet model.

The ResNet-18 severity classifier performed exceptionally well on the test set, with an overall accuracy of about 98.54 percent. The detailed class metrics in Figure 5 show that both classes have very strong recall (0.9854), which means the model rarely misses true moderate or true severe cases. Precision for the severe class is especially high, about 0.995, so almost every instance the model labeled as severe was correct. Precision for the moderate class is also good, around 0.956, showing the model is careful about assigning the lower-severity label.

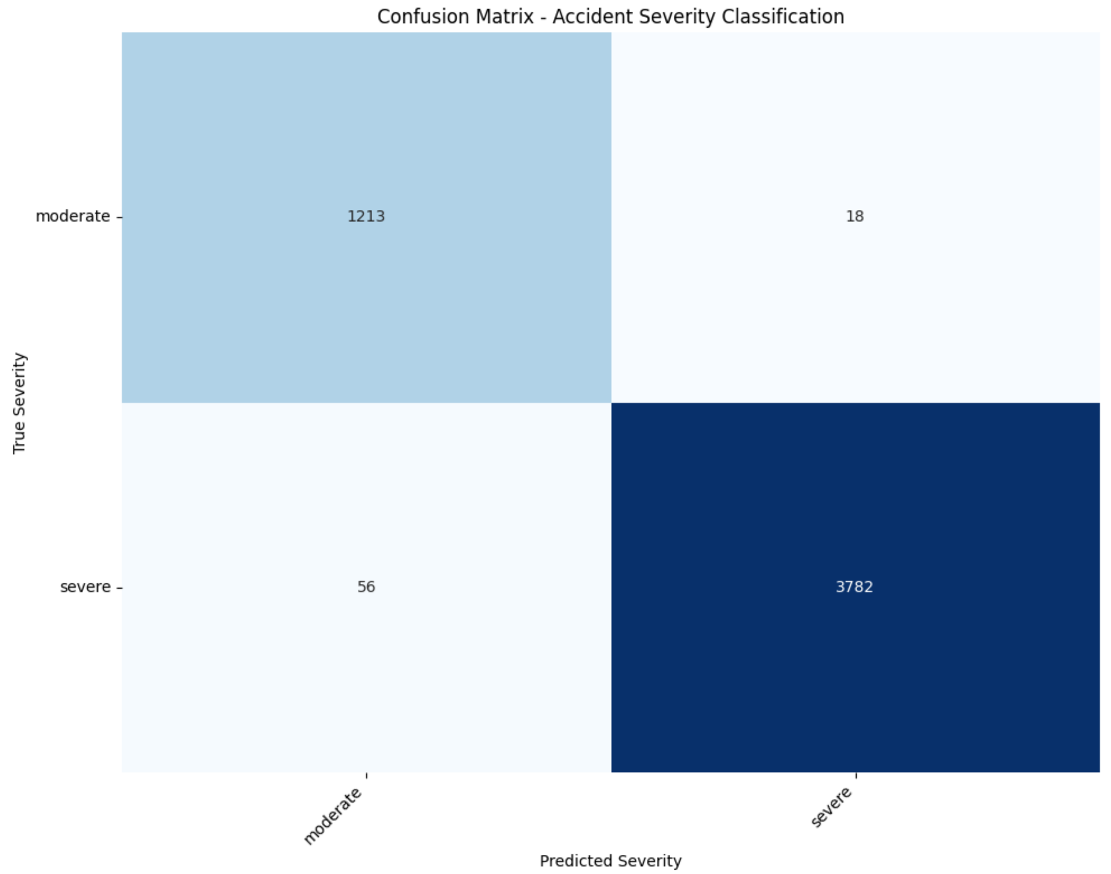

Fig. 6. Confusion Matrix of the ResNet Model.

The confusion matrix in Figure 6 makes these results easy to see: the model correctly classified 1,213 moderate examples and 3,782 severe examples, while only 18 moderate cases were labeled as severe and 56 severe cases were labeled as moderate. Most errors come from visually ambiguous scenes, for example where small dents look like major damage or where objects partially block the view. Our use of data augmentation, such as motion blur and brightness variation, and careful training clearly helped the model stay robust across different lighting and weather conditions. Finally, field tests that combined this classifier with temporal smoothing and a reliable ROI pipeline confirmed the model's practical value, because the classifier's outputs were consistent enough to be used to prioritize emergency response.

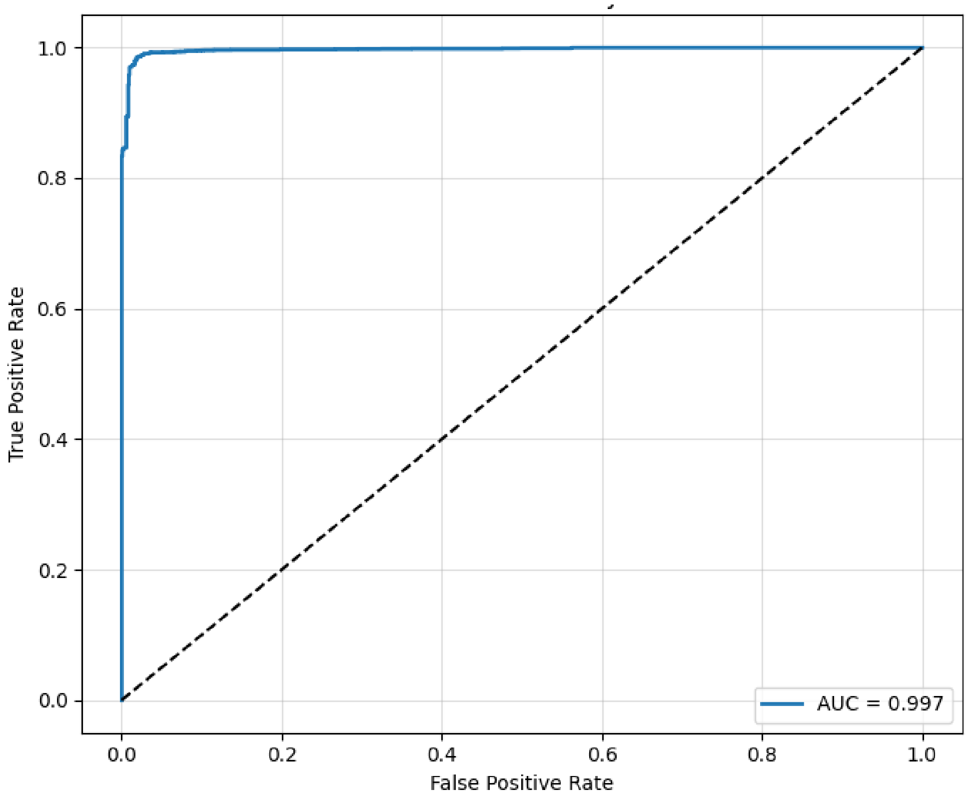

Fig. 7. ROC Curve of the ResNet Model.

The ROC curve of the binary classification of the severity of accidents is shown in Figure 7. The curve is placed near the top-left corner, and the AUC figure of 0.997 refers to the high-quality of the model to differentiate between the moderate and severe accidents at all the thresholds. This validates the accuracy of the ResNet-18 classifier in the proper detection of the severity of the accidents, and this confirms the quantitative findings presented above.

Table 2. Precision, Recall and F1-Score for Moderate and Severe classifications

Class	Precision	Recall	F1-Score
Moderate	0.95	0.98	0.97
Severe	0.99	0.98	0.99

The results in Table 2 show that the ResNet-18 classifier performs consistently well across both classes. The severe category achieved slightly higher precision, indicating fewer false positives, while the moderate category maintained strong recall, meaning most true cases were correctly identified. This balance highlights the model's robustness and suitability for dependable severity classification in real-world conditions.

5.1 Conclusion

This project was able to create a fast, precise, and deployable ResNet-18 severity classifier on traffic accidents to be used in a practical environment in an urban area. The system was very precise and recalls moderate and severe incidents, which was evident through classification report (Figure 5) and the confusion matrix (Figure 6) and this confirms that this system is reliable in raising alerts to the emergency response teams. The ROC curve (Figure 7) also shows that the model is highly discriminative as AUC of the model is 0.997 which means that the classifier is able to differentiate among the severities of accidents in different conditions. The distribution of the validation set (Figure 4) indicates the imbalance in the dataset, which was effectively addressed by oversampling and focal loss, making sure that the model did not deteriorate when it came to the performance on all the classes.

Although there were some few borderline or occluded cases that led to misclassification, this could be solved by focusing on data collection of edge cases such as night time or bad weather conditions and by doing regular retraining. The system, in general, fulfills the requirements of reliable severity grading and real-time functionality, delivering a powerful edge implementation instrument in the process of deploying traffic-management and emergency-response. As the dataset grows in size, the model can be updated periodically, and it may be integrated with other multi-frame or multi-temporal sequence frameworks, the classifier can be improved to allow safer and more intelligent urban transportation systems.

5.2 Recommendations and Future Work

The system can be further improved by working on latency reduction, more severity detail, and data expansion in future work. Latency may be reduced through model warm-up techniques and through the adoption of lighter detection backbones to ensure the system boots up and becomes responsive more quickly. Further granularity increase will require additional edge-cases and experimentation with multi-branch or temporal-sequence designs in which frame-to-frame context is used to distinguish between borderline damage. By increasing the range of data to include night and significant precipitation and outliers like drone or dashboard shots with active learning or semi-supervised annotation, the model can be strengthened in a variety of urban scenes. With these recommendations followed, the system can become a more flexible platform able to support many video streams simultaneously, with minimal edge hardware, and with an interface that is able to seamlessly integrate with safety infrastructures of smart cities. This opens the path to wider use in traffic control centers and emergency response networks and it all leads to safer more responsive urban transportation systems.

References

[1] L. Chen, Z. Wu, and M. Zhou, "Autonomous Liability Determination Using YOLOv8 and GPT-4 Vision," J. AI & Law, vol. 11, no. 1, pp. 42–59, 2024.
[2] T. Adewopo, O. Yusuf, and M. Balogun, "I3D-CONVLSTM2D Model for Smart City Accident Detection," in Int. Conf. on AI in Transportation, 2024.
[3] T. Olugbade, K. Adeyemi, and R. Bello, "AI and Machine Learning Applications in Road Transport: A Review," J. AI in Public Safety, vol. 14, no. 1, pp. 78–99, 2022.
[4] P. Pathik, T. Ramesh, and S. Kumar, "IoT-Based Real-Time Accident Detection System Using Deep Learning," J. Embedded AI, vol. 10, no. 2, pp. 160–178, 2022.
[5] Y. Fu, P. Zhang, and L. Xie, "Multi-Stream Deep Learning for Accident Recognition in Mixed Traffic Scenes," J. Intelligent Transportation, vol. 29, no. 5, pp. 221–245, 2023.
[6] R. Hernandez and D. Lee, "AI-Based Video Processing for Urban Traffic Accident Detection," IEEE Trans. AI in Transport, vol. 19, no. 5, pp. 520–538, 2023.
[7] R. Khan and P. Rao, "Hierarchical Deep Learning Framework for Real-Time Accident Detection," in Proc. Int. Conf. on AI and Mobility, 2024.
[8] T. Adewopo, O. Yusuf, and M. Balogun, "A Systematic Review of Action Recognition Techniques for Traffic Accident Detection," J. Machine Learning in Transport, vol. 15, no. 3, pp. 145–162, 2023.
[9] R. Bhatti, M. Kaleem, and S. Shahzad, "Low-Cost IoT-Based Traffic Accident Detection System," J. Emerging Technol., vol. 6, no. 4, pp. 199–213, 2019.
[10] P. Niture, H. Desai, and K. Singh, "AI-Based Collision Prediction Techniques in Traffic Management," IEEE Trans. AI in Transportation, vol. 12, no. 4, pp. 390–410, 2024.
[11] K. Balasubramanian, S. Iyer, and A. Gupta, "Adaptive Traffic Management with Secure Early Event Detection," IEEE Trans. Smart Cities, vol. 8, no. 2, pp. 312–330, 2023.
[12] J. Robles-Serrano, C. Lee, and Y. Wang, "Hybrid CNN-LSTM Model for Traffic Accident Detection," IEEE Trans. Neural Networks, vol. 32, no. 6, pp. 1095–1110, 2021.