Preprints.org

Article

Not peer-reviewed version

Computer Vision-Based Road Accident Classification from Traffic Surveillance

Shourav Chowdhury [*] , Subrata Barua [*] , Shudipta Banik , K.M. Naimuddin , Imam Hassan Sajib [*]

Posted Date: 15 January 2025

doi: 10.20944/preprints202501.1076.v1

Keywords: Road accident; CCTV; Transfer Learning; Long Short-Term Memory (LSTM); Footage; Surveillance

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Computer Vision-Based Road Accident Classification from Traffic Surveillance

Shourav Chowdhury [1], Subrata Barua [2], Shudipta Banik [3], K.M. Naimuddin [1] and Imam Hassan Sajib [1]

[1] Department of Computer Science and Engineering, Premier University, Bangladesh
[2] Department of Electronics and Telecommunication Engineering, Chittagong University of Science and Technology, Bangladesh
[3] Department of Computer Science and Engineering, East Delta University, Bangladesh

Abstract: Traffic accidents stand as a leading cause of fatalities worldwide, significantly impacting global mortality rates. Accurate classification of road accidents through advanced technological solutions presents a crucial opportunity to revolutionize accident prevention and emergency response strategies. This paper presents an advanced deep-learning methodology customized for the classification of road accidents using CCTV surveillance footage. This real-time dataset, comprising approximately 18,000 frames, has been amassed, which is pivotal for enabling comprehensive research in this field. This substantial dataset is the foundation for these investigative efforts, providing a rich and diverse source for conducting an in-depth analysis of the features. We have achieved a remarkable accuracy of 97% on this dataset through the strategic utilization of transfer learning in conjunction with LSTM (Long Short-Term Memory) techniques. This accomplishment underscores the efficacy of our approach, combining the strengths of transfer learning and LSTM models, resulting in a highly accurate classification system for road accident events.

Keywords: Road accident; CCTV; Transfer Learning; Long Short-Term Memory (LSTM); Footage; Surveillance

1. Introduction

Due to serious vehicle accidents on roads and highways in recent decades, the general public's knowledge of traffic's detrimental impacts has increased. According to the Global Road Safety Statistics(GRSS), Millions more suffer life-threatening injuries, and over 1.35 million people pass away on the world's roadways each year [1]. Most road traffic accident analysis, decision-making, and other processes now occur during the manual processing stage, which is the primary reason for the low effectiveness and imprecise decision-making regarding the massive volume of data related to road accidents.

Many human casualties occur due to the proper monitoring of road accident events. Traditional accident detection methods often result in sluggish reaction times and limited accuracy since they primarily rely on human observation or basic sensor-based systems. Automated road monitoring surveillance systems can help a victim get life acceptance for accidental events with the proper medical assistance. The timely detection of accidents and monitoring traffic conditions are essential for implementing mitigation plans and effective reaction measures. After years of advancing safety procedures and system architecture, the technical community has concluded that accident and system loss rates have reached a tipping point beyond which more advancements in safety seem unachievable [2].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

Figure 1. Possible Collusion.

The application of computer vision to classify traffic accidents has important and broad practical ramifications. For example, strong algorithms for automatically detecting accidents in film captured by traffic surveillance might significantly increase the functionality of traffic management systems. The research attempts to develop a robust and efficient system using transfer learning in conjunction with LSTM (Long-Short-Term Memory) that can recognise and categorise different types of accidents instantly and with high performance. The findings and methods presented in this paper aim to significantly increase the steps taken to guarantee road safety and advance the development of advanced transport systems. This procedure proves that our methodology is appropriate for actual accident events on various real-time dates with environments.

The paper is organized into key sections as follows. Section 1 outlines the problem statement with the proposed solution. The pertinent background has been discussed throughout the Section 2. Section 3 presents the solution architecture and methodology of an innovative approach that integrates Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs), outlines a systematic process encompassing data collection, preprocessing, feature extraction, and the application of learning algorithms to establish a robust framework for identifying road accidents. The experimental result and analysis of the spatial and temporal features in video sequences, enhancing the accuracy of accident detection and classification, have been presented in the next Section 4. Lastly, Section 5 concludes and synthesizes the findings, highlighting the importance of advanced technological solutions in transforming accident prevention and emergency response strategies with the direction for future work.

2. Related Work

Automated systems can swiftly detect accidents as they occur, allowing for quicker emergency response times. This can be crucial in reducing the severity of injuries and preventing further accidents caused by the initial collision. The fundamental objective of automated systems is to determine how to extract many parameters at low computing cost to anticipate or identify risky scenarios, as vision-based processing involves a lot of computation. Earlier systems relied on basic sensor technologies like pressure plates, induction loops, and simple traffic cameras. These systems had limited capabilities to detect accidents and relied more on human intervention for verification. Systems now leverage advanced sensor technologies such as AI-powered cameras, lidar, radar, and IoT-connected devices. Now, Researchers are focusing on developing a system that can work on many parameters to anticipate potentially harmful circumstances. Earnest Paul Ijjina [3] et al. highlight an architecture that makes

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

use of Mask R-CNN for precise object recognition and then an effective object tracking algorithm based on centroid for surveillance footage with a high Detection Rate and a low False Alarm Rate. A vehicle's speed and trajectory abnormalities following an overlap with other cars are used to calculate the likelihood of an accident in various scenarios, including bright sunshine, poor visibility, rain, hail, and snow. Similarly, Yanbin [4] et al. propose a forecasting model of traffic accident occurrences and give a traffic accident index system that considers weather, wind speed, month, and week. Likewise, S. Ghosh [5] et al. develop a system that can identify an accident by analysing a live video feed from a CCTV camera mounted on a roadway. The concept involves utilising a deep learning convolutional neural network model, trained to identify whether a frame in a video is an accident to process each frame of the footage. Compared to other image classification algorithms, CNN-based classifiers require less preprocessing and have achieved accuracy levels of over 95% On the other hand, A.K. and Prabhakar [6] et al. introduce a unique method for identifying and classifying abnormal vehicles on urban highways. Model-based tracking data is used to compute speed information and identify slow-moving and long-time stopped vehicles. Moreover, Ola et al. present a model that can analyse accident events using the Accident Evolution and Barrier Function (AEB) method [7]. The strategy compels professionals in human factors and other fields to collaborate in a cooperative process culminating in the analysis. However, Yan Fu Wang et al. have an accident analysis model that combines the Bayesian network (BN) with the Human Factors Analysis and Classification System (HFACS) [8]. The model also employs Evidential Reasoning (ER) and the best-fit technique to rate the suggested preventive actions according to their cost-effectiveness. The recommended approach can give accident investigators a resource to create safety intervention plans that are both economical and effective. Mirza M. Lutfe Elahi et al. present an approach that tracks and determines several kinetic properties of cars using vision-based methods and learns traffic patterns from roadside video data [9]. Ultimately, the suggested approach finds anomalies (potential for collision) and roadside incidents. Regarding identifying exceptional circumstances, the suggested approach has an accuracy level of about 85%. H. Sharma et al. propose an automated smartphone-based accident detection system [10]. Despite the sensitivity and range limits of the different smartphone sensors, the system can correctly predict an accident based on the collected data. Recent studies have demonstrated that complex characteristics may be automatically extracted from photos using Deep Learning techniques. Several researchers have focused on extracting temporal and spatial information from footage using deep learning techniques. Sen et al. propose a model for classifying soccer actions based on knowledge extraction from football films, including highlight extraction, match summary, content-based video retrieval, and context-based advertising [11]. Different football action durations are classified using a combination of convolutional and recurrent neural networks. Transfer learning first identifies distinctive spatial properties from a pre-trained VGG network. Similarly, Moumita Sen Sarma et al. created a scratch model that combines the two well-known deep learning methods, long short-term memory (LSTM) and convolutional neural network (CNN) [12]. Utilising both spatial and temporal feature extraction, the Deep Learning technique is applied to traditional Bangladeshi sports films to identify them.

3. ARCHITECTURE & METHODOLOGY

This study introduces an innovative approach integrating advanced deep learning techniques to accurately discern between accident and non-accident events depicted in video sequences. The fundamental methodology centers on fusing Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs) to extract and analyze spatial and temporal features embedded within video frames. By systematically progressing through data collection, preprocessing, feature extraction, and learning algorithms, the objective is to craft a robust framework adept at identifying accidents. This proposed methodology holds significant promise in augmenting safety protocols and fortifying surveillance systems. Figure 2. depicts the workflow of the manuscript, illustrating the systematic approach taken throughout the research process, from data collection to implementing the proposed methodology.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

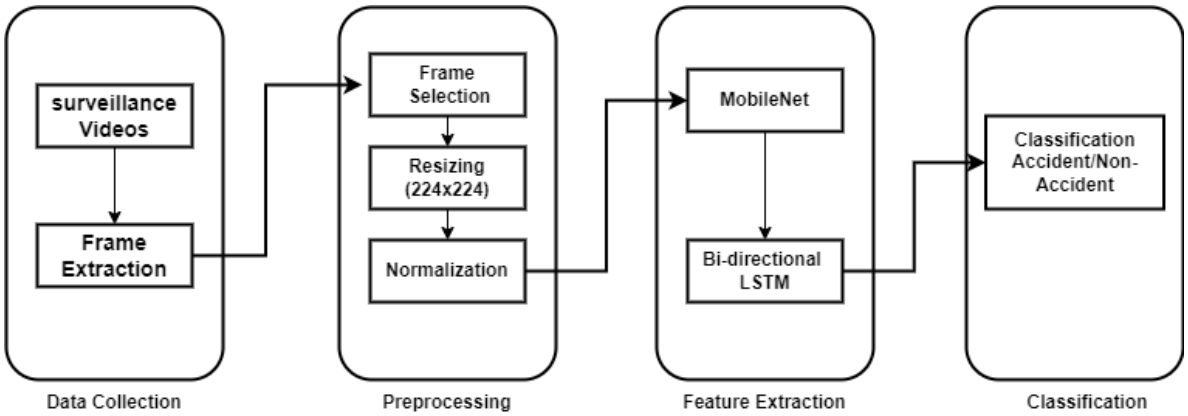


Figure 2. Work-flow Diagram

3.1. Data Collection

The dataset under examination is meticulously compiled, encompassing a rich repository of real-life scenarios capturing diverse accident and non-accident events. We have collected our dataset from multiple sources such as YouTube videos, Instagram videos, and Facebook, and some non-accidental videos have been collected from Kaggle, as shown in Figure 3. Our custom dataset consists of 120 training videos, with 65 videos of accidents and 55 videos labelled as non-accident. Videos range from a minimum of 4 s to 15 s. The approximate number of frames of the whole dataset is 1.8k. This expansive dataset enhances the model's learning capacity and enables it to discern accidents across varying real-worlds. The dataset comprises various frames, as exemplified in Figures 3 and 4, illustrating various scenarios relevant to accident detection and classification.

Figure 3. Sequential frames of Non-Accident

Figure 4. Sequential frames of Accident

3.2. Preprocessing

The dataset acquisition process involved careful extraction of frames from entire video sequences, ensuring adequate representation while maintaining temporal consistency. The method employed for frame selection aimed at capturing key temporal moments within the video, ensuring a balanced representation of various scenes. Figure 5. visually represents the frame selection process. The frame extraction process from a video of length L involves determining the skip interval S to uniformly sample frames across the entire video duration. The number of frames, N, to be extracted for a sequence of length SEQUENCE_LENGTH is calculated as follows:

$$N = L/SEQUENCE\_LENGTH \quad (1)$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

The skip interval S is then computed as:

$$S = L/N \tag{2}$$

This skip interval ensures that frames are uniformly sampled throughout the video sequence, with a representative selection that accounts for different temporal instances. The extracted frames F = $\{f_1, f_2, f_3, ...f_n\}$ where f_i represents a frame, undergoing subsequent preprocessing steps to ensure uniformity, quality, and diversity within the dataset. The meticulous curation of this dataset, capturing various accident scenarios under diverse environmental conditions and camera perspectives, fortifies the model's capacity to generalize across real-world scenarios.

Figure 5. Sequence of Frames

These actions for frame selection ensure a balanced representation of temporal moments within the video while maintaining temporal coherence and diversity in the dataset. The resultant dataset comprises systematically sampled frames from various video sequences, forms the foundational corpus for subsequent preprocessing and model training stages. After the extraction of frames, the images were uniformly resized to dimensions of 224x224 pixels. Moreover, the original extracted frames consist of RGB coefficients in the 0-255 range that will be too high to be processed. Therefore, we converted the pixel values to the [0,1] interval by scaling with a 1/255.0 factor. Standardizing raw input pixels through normalization involves re-scaling them to possess a mean of 0 and a standard deviation of 1, ensuring uniformity and aiding in data preprocessing.

3.3. Spatial Feature Extraction

The feature extraction procedure involves taking the preprocessed video frames and then employing the MobileNetV2 pre-trained CNN architecture to extract high-level spatial information. They are precisely reshaped to ensure the extracted features are compatible with the LSTM architecture's input requirements. This allows for smooth temporal analysis in the future. This transformation aims to efficiently capture temporal dynamics so that the model can interpret complex temporal dependencies necessary for precise event classification. Decoding geographic information while maintaining temporal coherence requires the inclusion of MobileNet as a feature extraction technique, which paves the way for robust event identification.

3.3.1. MobileNet & ImageNet

MobileNets are a family of TensorFlow computer vision models focused on mobile devices and are intended to optimize accuracy while considering the limited resources of embedded or on-device applications [13]. MobileNets are designed to fulfil the resource requirements of devices and are compact, low-latency, and low-power. The models find application in classification, detection, embeddings, and segmentation tasks. Researchers at Google created a type of convolutional neural network called MobileNets. They are resource-friendly and instantly functional on mobile devices since they are "mobile-first" [14]. MobileNets have width and resolution multiplier settings that can be adjusted to balance the resource-accuracy tradeoff. While the resolution multiplier can alter the dimensions of the input image, the width multiplied can narrow the network. These modifications may weaken the internal organization of each layer. An image database called ImageNet is set up according to the WordNet hierarchy [15]. Thousands upon thousands of photos represent each hierarchical node. Currently, each node has an average of over 500 photos.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

3.4. Learning Algorithm

The architectural design involving TimeDistributed layers and Bidirectional LSTM units is central to the proposed methodology. This architecture enables the model to capture intricate temporal dependencies present within sequences of video frames. The Lstm layer's pivotal role involves processing the encoded features, effectively encoding and decoding temporal dynamics, essential for accurate event classification. By incorporating sophistical learning algorithms and architectures, the model gains the ability to discern nuanced patterns and temporal correlations with video sequences. The systematic approach ensures a comprehensive understanding of spatial and temporal characteristics, empowering the model to proficiently differentiate between accident and non-accident events. Figure 6. illustrates the model architecture described above, visually representing the integration of TimeDistributed layers and Bidirectional LSTM units.

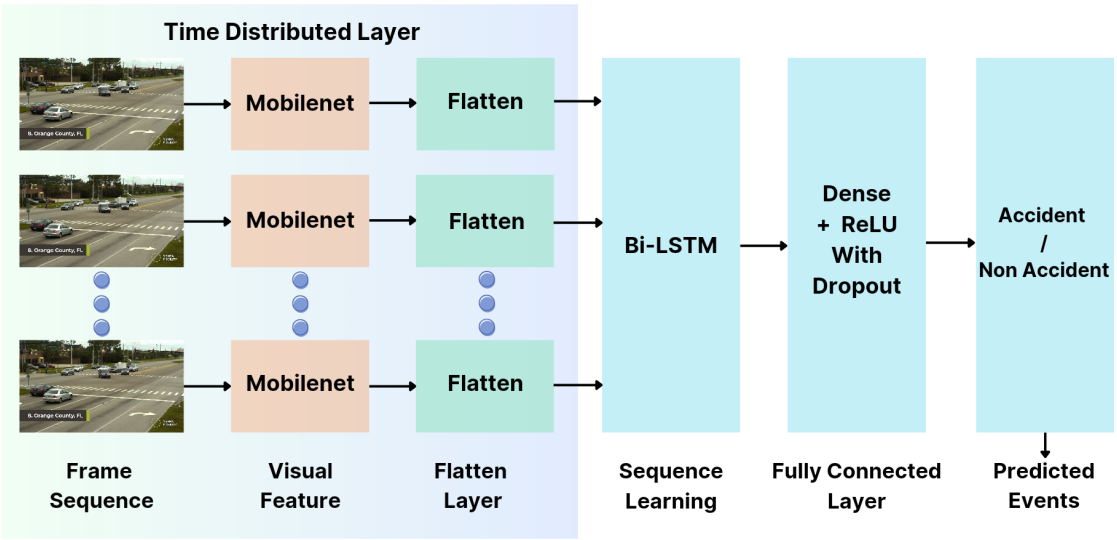


Figure 6. Mobi-LSTM network architecture

3.4.1. Model Description

The proposed model architecture orchestrates a comprehensive framework for sequential image data processing, integrating convolutional and recurrent operations for binary classification tasks. Comprising an input layer tailored to accommodate sequential image data of dimensions (SEQUENCE_LENGTH, IMAGE_HEIGHT, IMAGE_WIDTH, 3) representative of RGB images, the model strategically incorporates a TimeDistributed layer employing MobileNet independently on each frame within the sequence.

Table 1. Model Architecture Summary

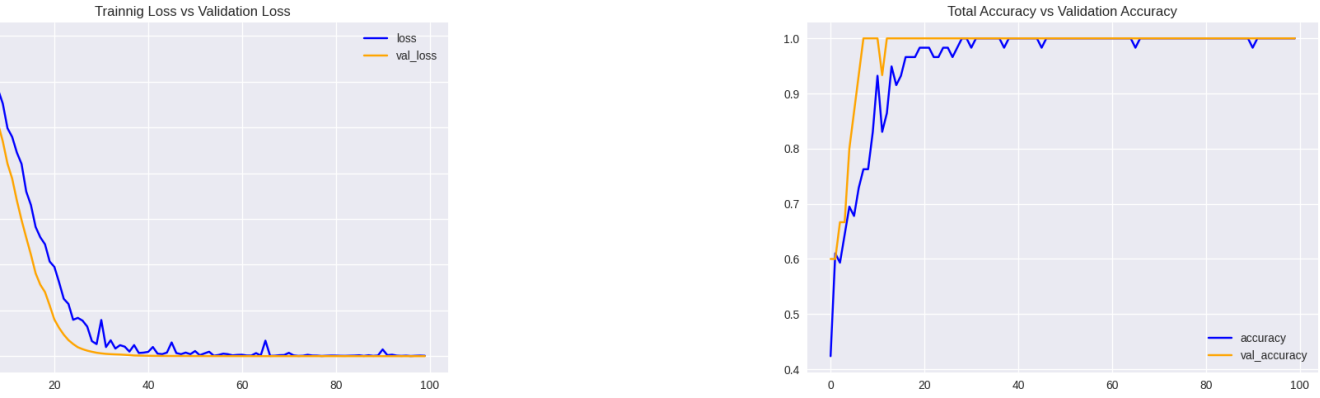
Layer	Neurons	Activation	Dropout
Input	2257920	N/A	N/A
TimeDistributed	N/A	N/A	0.25
TimeDistributed	N/A	N/A	N/A
Bidirectional LSTM	64	N/A	0.25
Dense	256	ReLU	0.25
Dense	128	ReLU	0.25
Dense	64	ReLU	0.25
Dense	32	ReLU	N/A
Output	2	Sigmoid	N/A

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

Post-MobileNet application, a dropout layer with a 25% dropout rate regulates overfitting, followed by a TimeDistributed Flatten layer to prepare the data for subsequent processing. Subsequently, Bidirectional LSTM layers, each equipped with 32 units in both forward and backward directions, capture temporal dependencies within the sequential data. Additional dropout layers (25% rate) post-bidirectional LSTM aid in regularization. The model further incorporates dense layers (256, 128, 64, and 32 units) with ReLU activation to distil high-level features from the data. Culminating in an output layer with a Sigmoid activation function, the architecture facilitates binary classification predictions aligned with the predefined classes. Compiled with a Binary Cross-Entropy loss function, Adam optimizer, and accuracy-based metrics, this model amalgamates convolutional insights from MobileNet with temporal understanding via Bidirectional LSTMs, presenting a robust and versatile solution for sequential binary classification tasks.

4. Result & Analysis

The combined MobileNet and LSTM model used to classify accident and non-accident events in surveillance video data yielded promising results. The dataset was divided into train, test, and validation data at a ratio of 50:30:20 to use the scratch model and the refined pre-trained models in classifying the accident videos. The categorization process requires selecting a certain number of frames from each video. To find the ideal frame in this regard, we experimented with 120 video samples in the whole dataset with all sequence lengths between 5 and 15. Additionally, with a test accuracy of 97%, the model performs optimally for sequences of length 15. Following thorough examination and experimentation with several hyper-parameter combinations, we developed our scratch model, which performed admirably in identifying traffic accidents. Transfer learning and LSTM were used to create this scratch model, which was then assembled using the Adam optimizer with a 0.0005 learning rate and a binary cross-entropy loss function. A kind of loss function called binary cross-entropy uses a one-hot vector to represent the goal values for binary classification. The categorical cross-entropy loss function uses the probability distribution that the sigmoid activation function in the output layer provides to calculate the prediction error or loss of the model. The suggested model's loss curves and training and validation accuracy are shown in Figure 7. This figure shows that the model's maximum training accuracy has reached during the 100th epoch, where validation accuracy is maximum.



(**a**) (**b**)

Figure 7. (a) **Loss Curve** - The graph depicts the training and validation loss over epochs. A consistent decrease in loss indicates effective learning by the model, while a significant divergence between training and validation loss may suggest overfitting. (b) **Accuracy Curve** - The plot shows the training and validation accuracy over epochs. An upward trend in accuracy signifies improved predictive performance, with convergence reflecting alignment between training and validation results.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

One of the most effective methods for evaluating a model's performance in classification tasks is the confusion matrix, distinguished among comparison tools. The confusion matrix shows the percentages of each class's accurate and inaccurate predictions generated by the classifier. As depicted in Figure 8, it facilitates a comprehensive assessment of the model's performance across individual classes.

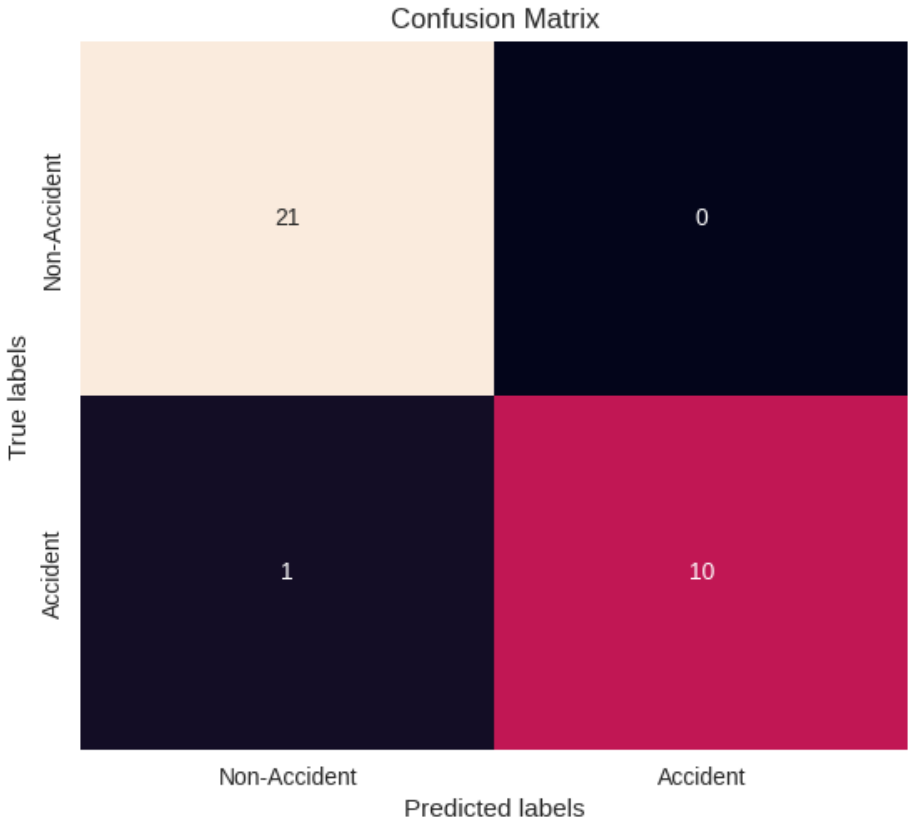


Figure 8. confusion Matrix

It represents the confusion matrix of the proposed model. Certain classes have been misclassified to a small degree due to semantic feature similarities with the incorrectly categorized classes. As detailed in Table II of the classification report, this matrix is a foundational element in computing various performance metrics, including accuracy, precision, recall (sensitivity), and the F1 score.

Table 2. Classification Report

Class	Precision	Recall	F1-Score	Support
Non-Accident	0.95	1.00	0.98	21
Accident	1.00	0.91	0.95	11
Macro Avg	0.98	0.95	0.96	32
Weighted Avg	0.97	0.97	0.97	32

We got a precision of 100% for accidents and 95% for non-accidents; the model exhibited a strong ability to correctly identify these instances.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

Despite a slightly lower recall of 91% for accidents, indicating some missed identifications. It calculates the model's capability to correctly identify positive instances by using Equation (4).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4}$$

The overall high F1 scores of 0.95 for accidents and 0.98 for non-accidents showcased a balanced performance in capturing true positives and minimizing false negatives.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

The Results section reports experiments with a real-time dataset of approximately 18,000 frames. By applying transfer learning in conjunction with LSTM techniques, the experiments achieved a notable accuracy of 97%. This underscores the methodology's effectiveness in classifying various types of road accidents and its potential to significantly improve traffic management systems.

This study underscores the efficacy of the hybrid model in distinguishing between event classes, albeit with the potential for further refinement to enhance sensitivity in detecting accident events. It emphasizes the developed system's potential to improve road safety and traffic surveillance efficiency, paving the way for future research in this critical domain.

5. Conclusions & Future Works

This paper presents a hybrid Mobilenet-Bi-LSTM architecture for detecting vehicle collisions. The architecture learns from real video data from various public repositories. The created dataset, comprising 120 videos encompassing both accidental and non-accidental scenarios, serves as the foundation for model training and evaluation. The data is studied, and the results are compared to the proposed architecture. The proposed architecture can detect accidents currently with a 97% detection rate on the accident videos obtained under various ambient conditions. The experimental results show the prowess of the proposed architecture.

However, the created surveillance video dataset predominantly comprises daytime videos, potentially posing challenges for the model in detecting nighttime scenarios. We commit to addressing this limitation in future iterations, enhancing the model's performance across a broader range of lighting conditions. One of the limitations of the proposed architecture is its ineffectiveness for high-density traffic in vehicle detection, which will be addressed in future work. Additionally, large obstacles obstructing camera views are recognized as a potential impediment to the vehicle detection. Future research could concentrate on augmenting the model through expanded datasets or fine-tuning parameters to optimize its accuracy and reliability in real-time accident detection scenarios.

Acknowledgments: We thank Mohammad Hasan for his valuable review and feedback on the paper. His contributions were crucial to the success of this research.

References

1. Brake. "Global road safety statistics, [Online]. Available: https://www.brake.org.uk/get-involved/take-action/mybrake/knowledge-centre/global-road-safety. [Accessed: 19-Dec-2023].
2. Nivolianitou, Z.; Leopoulos, V.; Konstantinidou, M. Comparison of techniques for accident scenario analysis in hazardous systems. *Journal of Loss Prevention in the Process Industries* **2004**, *17*, 467–475. https://doi.org/https://doi.org/10.1016/j.jlp.2004.08.001.
3. Ijjina, E.P.; Chand, D.; Gupta, S.; Goutham, K. Computer vision-based accident detection in traffic surveillance. In Proceedings of the 2019 10th International conference on computing, communication and networking technologies (ICCCNT). IEEE, 2019, pp. 1–6.
4. Yanbin, Y.; Lijuan, Z.; Mengjun, L.; Ling, S. Early warning of traffic accident in Shanghai based on large data set mining. In Proceedings of the 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, 2016, pp. 18–21.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2025 doi:10.20944/preprints202501.1076.v1

5. Ghosh, S.; Sunny, S.J.; Roney, R.M. Accident Detection Using Convolutional Neural Networks. *2019 International Conference on Data Science and Communication (IconDSC)* **2019**, pp. 1–6.
6. HD, A.K.; Prabhakar, C. Vehicle abnormality detection and classification using model based tracking. *International Journal of Advanced Research in Computer Science* **2017**, *8*.
7. Svenson, O. Accident and incident analysis based on the accident evolution and barrier function (AEB) model. *Cognition, Technology & Work* **2001**, *3*, 42–52.
8. Wang, Y.F.; Xie, M.; Chin, K.S.; Fu, X.J. Accident analysis model based on Bayesian Network and Evidential Reasoning approach. *Journal of Loss Prevention in the Process Industries* **2013**, *26*, 10–21.
9. Elahi, M.M.L.; Yasir, R.; Syrus, M.A.; Nine, M.S.Z.; Hossain, I.; Ahmed, N. Computer vision based road traffic accident and anomaly detection in the context of Bangladesh. In Proceedings of the 2014 International Conference on Informatics, Electronics & Vision (ICIEV). IEEE, 2014, pp. 1–6.
10. Sharma, H.; Reddy, R.K.; Karthik, A. S-CarCrash: Real-time crash detection analysis and emergency alert using smartphone. In Proceedings of the 2016 International Conference on Connected Vehicles and Expo (ICCVE). IEEE, 2016, pp. 36–42.
11. Sen, A.; Deb, K. Categorization of actions in soccer videos using a combination of transfer learning and Gated Recurrent Unit. *ICT Express* **2022**, *8*, 65–71.
12. Sarma, M.S.; Deb, K.; Dhar, P.K.; Koshiba, T. Traditional Bangladeshi sports video classification using deep learning method. *Applied Sciences* **2021**, *11*, 2149.
13. PACKT Books. What is transfer learning? | PACKT Books. [Online], 2023. [Accessed: 08-Dec-2023].
14. Harvey, M. Creating Insanely Fast Image Classifiers with MobileNet in TensorFlow, 2017. Accessed: 08-Dec-2023.
15. Website. Accessed: 08-Dec-2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.