

VISION BASED ACCIDENT IDENTIFICATION IN TRAFFIC VIDEOS USING DEEP LEARNING

Tehreem Qamar^{1,*}, Narmeen Zakaria Bawany¹, Jawwad Ahmed Shamsi², and Kanwal Zahoor¹

Received: April 05, 2023; Revised: October 31, 2023; Accepted: December 13, 2023

Abstract

Traffic collisions have emerged as a prominent factor contributing to a significant rise in injury-related fatalities and injuries, making them a pressing concern for public safety. Statistics show that an estimated 1.4 million people die in traffic accidents every year and around fifty million are injured. With the increasing population, ensuring road safety has become one of the critical challenges for city administration. In response to this challenge, the field of computer vision has gained prominence, focusing on the development of digital systems capable of processing, analyzing, and interpreting visual data like humans. Recently, a few studies have applied machine learning techniques for accident detection in traffic videos, offering automated solutions for city monitoring. However, these existing models are limited to binary classification, that is, they only determine whether the accident has occurred or not. This research takes a step further by proposing a deep learning-based accident classification model that not only identifies vehicle accidents but also categorizes them based on the type of collision. We employ the transfer learning technique and evaluate five pre-trained models, including DenseNet121, InceptionNetV3, ResNet50, VGG16, and Xception. Our findings demonstrate that VGG16 stands out for the vehicle accident classification task, achieving an impressive accuracy of 96.76%. This research offers a significant contribution to enhance road safety by advancing the capabilities of accident detection and classification systems. The identification of collision types presented in this research empowers safety authorities and organizations to conduct more effective safety analyses serving dual purposes. Firstly, it ensures that emergency response teams and medical personnel are adequately prepared and secondly, it allows insurance companies to make fair determinations of liability and coverage.

Keywords: Deep learning, road safety; transfer learning, vehicle accident detection

¹ Center for Computing Research, Department of Computer Science and Software Engineering, Jinnah University for Women, Karachi, Pakistan. E-mail: tehreem.qamar@juw.edu.pk; narmeen.bawany@juw.edu.pk; kanwal.zahoor@juw.edu.pk

² Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan. E-mail: jawwad.shamsi@nu.edu.pk

* Corresponding author

DOI: <https://doi.org/10.55766/sujst-2023-06-e0939>

Introduction

Urbanization is changing the way people live all around the world. Cities that were once small are developing into megacities resulting in commercial and economic development but higher traffic congestion, crime rates and waste creation. Increase in population with more vehicles sharing crowded streets with vulnerable road users escalates the risk of accidents (Sun *et al.*, 2019). According to the World Health Organization, nearly 1.4 million people die every year in traffic accidents and around 50 million people experience non-fatal injuries, many of which result in disabilities (WHO, 2022). The most apparent reason for the fatality in accidents is the failure to respond promptly to accident victims (Ghosh *et al.*, 2019). City administration has focused on improving road safety and lowering traffic accidents. The CCTV cameras are installed on roadways for continuous monitoring round the clock. However, such monitoring can be carried out under continuous human supervision so that appropriate action can be taken as an accident occurs. This constant manual monitoring is tedious and prone to inaccuracy. Thus, there is a growing need for an automated solution that is spontaneous, effective, and sustainable enough to detect and respond to accidents.

Many cities are now transforming themselves into smart cities that uses information and communication technology (ICT) to improve citizen's quality of life, optimize city operations, and foster economic growth (Papa *et al.*, 2013; Basiri *et al.*, 2017; Farman *et al.*, 2022). Transportation and mobility are the key elements of the smart city, which offers a platform for managing traffic-related issues and employs smart surveillance strategies. Different methods to identify traffic accidents have been presented by the scientific community leveraging cloud computing, wireless communication, computer vision, machine learning (ML) and deep learning (DL) (Dogru and Subasi, 2018; Ghosh *et al.*, 2019; Gupta *et al.*, 2021).

Deep learning is a subset of machine learning that has demonstrated outstanding performance in a wide range of applications, including speech recognition, image classification, object detection, image segmentation, medical image analysis, and facial recognition (Young *et al.*, 2018; Adeel *et al.*, 2020; Alzubaidi *et al.*, 2021; Koppe *et al.*, 2021). Recently, deep learning algorithms performance has surpassed humans in identifying diseases in biological images (Esteva *et al.*, 2017; Nagpal *et al.*, 2019). The most widely used deep learning

technique for image recognition and analysis is Convolutional Neural Network (CNN) that automatically detects significant features via its convolution operation (O'Shea and Nash, 2015). Various studies have seen in the literature for vehicle accident detection that detect an accident without identifying the type of collision (Dogru and Subasi, 2018; Ghosh *et al.*, 2019; Chand *et al.*, 2020; Gupta *et al.*, 2021). Though, it is important to note that every collision within an accident holds its unique characteristics while belonging to a broader category. To the best of our knowledge, existing research studies have predominantly employed binary classification for vehicle accident detection. This focus on binary classification has resulted in a lack of properly annotated data to classify accidents into distinct categories.

Keeping in view the limitation of existing research, we have created a dataset by collecting traffic videos from YouTube and manually categorizing them based on the specific type of collision that occurred in each video. Further, our study presents a deep learning based model for vehicle accident detection and classification. The proposed model not only detects an accident, but also classifies the accidents into four common types of collisions: rear-end, head-on, side impact, and single-vehicle accidents.

Given the computational demands, large datasets, and extensive training times associated with deep learning models, we have leveraged transfer learning approach (Torrey and Shavlik, 2010). We employed five pre-trained models-DenseNet121, InceptionNetV3, ResNet50, VGG16, and Xception to develop the vehicle accident classification model. Hence, the major contributions of our study are as follows:

- We create a labeled dataset of vehicle accidents according to four types of collision.
- We develop a multi-class classification model for classifying vehicle accidents in traffic videos.
- We conduct experiments to compare the performance of pre-trained models for the detection of accidents in traffic videos.

The rest of the paper is structured as follows: Section II presents the review of the literature for the detection of accidents; Section III explains the methodology applied for this research, the dataset and the experimental design and setting used, Section IV discusses the results while Section V concludes the paper with future work.

Literature Review

To improve road safety, various accident detection techniques have been proposed and developed in the literature. These techniques can be divided into two broad categories i.e., vision-based and non-vision based. The vision-based techniques incorporate computer vision methods for processing road videos and images while non-vision based utilizes sensors, audio streams and social media data. We have divided this section into two parts; the first part discusses the non-vision based techniques applied to identify traffic accidents while the second part elaborates on the vision based techniques used for detection of accidents.

Non-Vision Based Techniques

Foggia *et al.* (2015) proposed an accident detection system by analyzing the sound streams to identify specific events like tire sliding, crashes, or other distortions caused by car collisions. Maaloul *et al.* (2017) presented the accident detection model by analyzing the speed of vehicles. However, their proposed research only performs well on main highways when traffic is moving quickly and continuously. Dogru and Subasi (2018) perform accident detection in vehicular ad-hoc networks (VANETs) and compare different machine learning algorithms for the detection task. The authors have found random forest as the superior algorithm among others with an accuracy of 91.56% and made emphasis that if location and velocity information for each car are provided, it is possible to evaluate the behavior of the vehicle.

Vision Based Techniques

Computer vision techniques such as Convolutional Neural Network and Recurrent Neural Network have been extensively applied for detecting accidents. Chand *et al.* (2020) proposes a support system for self-driving cars employing CNN. The researchers have created a custom dataset of dashcam footages and produced a vehicle accident detection model with 79.05% accuracy. Gupta *et al.* (2021) uses hierarchical recurrent neural network to detect accident in CCTV videos with an accuracy of 85.41%. The accident videos were collected from YouTube while normal videos were selected from DETRAC dataset. Ghosh *et al.* (2019) presented an accident detection system in CCTV videos by employing Raspberry Pi with CCTV camera. The researchers used a combination of CNN and LSTM and were able to achieve 92.38% accuracy of their model.

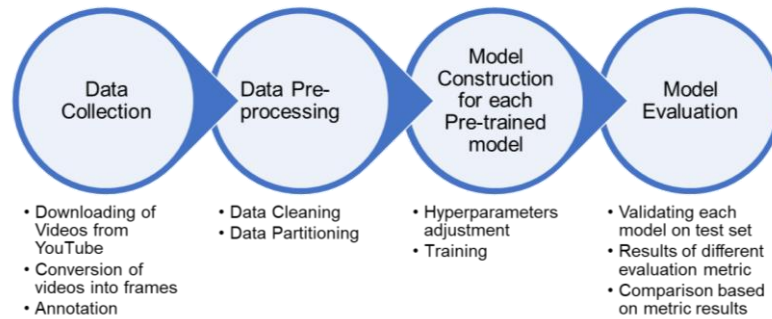
Singh and Mohan (2022) presents a model that makes use of the autoencoder to detect traffic incidents in videos. Their framework was divided

into object detection and anomaly detection and work in parallel. The first part aims to identify moving vehicles, their tracking, and determining where the objects being processed will collide. The second aims to make use of the spatial and temporal information of videos so that it may subsequently be represented by an autoencoder and used to calculate the anomaly score. Their proposed method uses the dataset from the CCTV surveillance network and achieves 77.5% accuracy for detection of accidents with an average of 22.5% false alarms on various lighting conditions. A similar model is suggested by Chan *et al.* (2017) made up of a recurrent neural network and spatial dynamic attention. The first is in charge of figuring out how much attention to give each object in each video frame. The second allows for the relationship between the signals that each item presents and the likelihood of an accident, allowing for the prediction of the event a few seconds in advance. Their approach achieves 74.35% mean average precision. Rajesh *et al.* (2020) proposed an accident detection system using CNN to detect accidents and alert the nearest control room promptly. The authors achieved an accuracy of 85% for the binary classification of accidents. Shah *et al.* (2018) offer a dataset of recordings of traffic accidents together with an accident forecasting model. The Faster RCNN architecture was modified by implementing context mining to the pooling layers and being used with LSTM architecture. With an average precision of 47.25%, they were able to attain a Time-To-Accident measure average of 1.684 sec. Yao *et al.* (2020) presented a dataset; DoTA, of different anomalous traffic videos and a new evaluation metric spatial temporal area under curve (STAUC). The dataset is annotated for when, where and what anomaly in the video and CNN based architecture is used to detect and recognize anomaly in the video. Saravanarajan *et al.* (2023) introduces a novel car crash detection system for autonomous vehicles, addressing the need to identify single-car crashes effectively. The system combines VGG16, Region Proposal Network (RPN), and CNN8L (region based detector) deep learning models, achieving an impressive Accident Detection Rate (ADR) of 86.25% and a relatively low False Alarm Rate (FAR) of 33.00%.

In the realm of vehicle accident detection within traffic videos, existing research has made notable efforts to enhance efficiency. However, accident detection using sensor-based techniques not only requires an adequate hardware setup but are likely to produce more false positive results because of the numerous events that might happen in the real-world. On the other hand, the computer vision-based models performed well on detecting

Table 1. Comparison of existing literature

Reference	Accident Classification Type	Technique Used	Evaluation Metric	Result (%)
Foggia <i>et al.</i> (2015)	Binary	Non-vision	Accuracy	78.95
Maaloul <i>et al.</i> (2017)	Binary	Non-vision	Precision, Recall, F1 Score	68, 91, 77
Dogru and Subasi (2018)	Binary	Non-vision	Accuracy	91.56
Chand <i>et al.</i> (2017)	Binary	Vision based	Accuracy	79.05
Gupta <i>et al.</i> (2021)	Binary	Vision based	Accuracy	85.41
Ghosh <i>et al.</i> (2019)	Binary	Vision based	AUC	92.38
Singh and Mohan (2022)	Binary	Vision based	AUC	77.50
Chan <i>et al.</i> (2017)	Binary	Vision based	Precision and Recall	80, 56.14
Rajesh <i>et al.</i> (2020)	Binary	Vision based	Accuracy	85
Shah <i>et al.</i> (2018)	Binary	Vision based	Average Precision	47.25
Yao <i>et al.</i> (2020)	Multiclass	Vision based	AUC	73.0
Saravanarajan <i>et al.</i> (2023)	Binary	Vision based	Accuracy	86.25
This Study	Multiclass	Vision based	Accuracy	96.76

**Figure 1. Methodology of the proposed study**

accidents in traffic videos but are limited to binary classification and considered general vehicular collisions irrespective of the type of collision. In contrast, this study employs computer vision techniques to categorize accidents into four distinct types. The accidents videos are collected from YouTube exploiting DoTA (Shah *et al.*, 2018) and converted into categorical dataset of images. Four categories are related to the type of collision that occurred while the fifth category points to the images which have No_Collision. Table 1 provides a summary of existing studies, emphasizing a comparative analysis with our study.

Methodology

The proposed methodology (depicted in Figure 1) makes use of pre-trained models of the most popular computer vision technique i.e. Convolutional Neural Network (O'Shea and Nash, 2015), for the detection and categorization of accidents in traffic videos. As there is no benchmark dataset available for traffic accidents, dataset creation is performed as the first step. The dataset is collected with the help of DoTA annotations on the videos

downloaded from YouTube. The downloaded videos are first converted into frames and then used in the development of accident classification model. Five pre-trained models are investigated namely DenseNet, InceptionV3, ResNet50, VGG16 and Xception. The pre-trained models are evaluated on the basis of accuracy, precision, recall and F1 score. The successive sections provide details about the dataset, the pre-trained models and the experimental setup utilized in the study.

Dataset Description

The dataset used in the study is prepared via the repository of dashcam video links, DoTA (Shah *et al.*, 2018). DoTA is a collection of traffic accident videos with temporal, spatial, and categorical annotations. We downloaded more than 300 videos from YouTube and converted them into images at 10 fps with 1,280×720 resolution. During preprocessing blurred, vague and duplicated images are discarded. The clear and distinctive images are categorized and annotated according to the type of collision visible in the accident video. After preprocessing, the final dataset comprises 1,800 images. The dataset was balanced as each category contains more than 300 images. The dataset is split



Figure 2. Sample images of each category in the dataset

into training, validation, and test sets with a ratio of 60:20:20, respectively.

The images are classified into following five categories and a sample of each category is presented in Figure 2:

- **Rear_end_Collision:** Collision with other vehicles that is moving ahead or in waiting condition. They usually happen at red lights, stop signs, and in congested traffic (Kim *et al.*, 1994).
- **Head_on_Collision:** Collision with another oncoming vehicle, these collisions mostly occur when people drive in wrong direction (Kim *et al.*, 1994).
- **Side_impact_Collision:** Collision with the side of another vehicle, either head-on or at an angle. They typically occur when a driver fails to stop at red light or making a turn (Kim *et al.*, 1994).
- **Single vehicle - Collision:** The vehicle becomes out of control and leaving the roadway to the left or right. This might happen when drivers swerve to avoid hitting road debris, an animal, or another vehicle (Kelley *et al.*, 2019).
- **No_Collision:** The vehicle runs smoothly i.e. No_Collision occurred.

Transfer Learning

Building, training, and optimizing production-quality models from scratch is expensive, requiring numerous iterations, domain expertise, and countless hours of computation. Therefore, the notion of transfer learning emerged that allows to fine tune a solution to a particular problem by using pre-trained models (Torrey and Shavlik, 2010). Pre-trained models are built on massive datasets like ImageNet (Deng *et al.*, 2009) and can be re-used to solve a similar task. Thus, save time and obtain high performing results in comparatively lower cost. This

study assesses five pre-trained models on accuracy, precision, recall and F1-score which are described below:

DenseNet: A DenseNet (Huang *et al.*, 2017) is a type of convolutional neural network that utilizes dense connections between layers, through Dense Blocks, where all layers are connected directly with each other. The input of a layer inside DenseNet is the concatenation of feature maps from previous layers. DenseNet121 is the most common type that consists of 121 convolutional layers, four average pooling and one fully connected (FC) layer with 8.1 million parameters. In DenseNet121, each layer is directly connected to every other layer in a feed-forward fashion. This dense connectivity leads to feature reuse and gradient flow, resulting in efficient training and better performance, particularly on datasets with limited training data.

Inception: An inception network is a deep neural network with an architectural design that consists of repeating components referred to as Inception module (Szegedy *et al.*, 2015). An Inception module uses 1×1, 3×3, and 5×5 convolutions in parallel to capture both fine and coarse-grained features, resulting in efficient and accurate image recognition. It was introduced in 2014 by Google, often referred to as GoogLeNet, and up till now is available in four versions. The inception V3 is the superior among all with 42 layers and 23.9 million parameters.

Residual Net (ResNet): ResNet (He *et al.*, 2016) is developed by the winner of ILSVRC (Russakovsky *et al.*, 2015). ResNet is an Artificial Neural Network (ANN) of a kind that stacks

Table 2. Performance of each pre-trained model in terms of accuracy

Pre-trained models	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
DenseNet121	99.92	95.94	96.22
InceptionNetV3	100	93.44	94.05
ResNet50	76.9	71.25	67.03
VGG16	100	95.6	96.76
Xception	99.8	91.2	92.43

residual blocks on top of each other to form a network. In a residual block, there are skip connections that enable the gradient to flow more easily during training. Several types of ResNet were developed based on the number of layers (starting with 34 layers and going up to 1,202 layers). The most common type is ResNet50, which comprised 49 convolutional layers with the kernel size of 1×1 and 25.6 million trainable parameters. Additionally, ResNet50's architecture utilizes global average pooling, and a final fully connected layer to perform tasks such as image classification, making it a popular choice for a wide range of computer vision applications.

Visual Geometry Group (VGG): VGG (Simonyan and Zisserman, 2014) is a multilayer deep CNN architecture that simulates the relations of the network representational capacity in depth. It is one of the most used image-recognition architectures and available in two variation VGG16 and VGG19 consisting of 16 and 19 convolutional layers respectively. It is characterized by its deep architecture with 16 weight layers, including 13 convolutional layers and 3 fully connected layers. VGG16 has a total of 138 million parameters. Notably, all convolutional layers have a small receptive field (3×3) with a small stride and use the same padding which helps in learning consistent and hierarchical features. This uniformity is often beneficial for recognizing intricate patterns in the data. It has a total of 138.4 million trainable parameters which allows it to generalize well to various tasks, especially when the target task has a relatively small dataset.

Xception: Xception stands for “extreme inception”, it takes the principles of Inception to an extreme (Chollet 2017). It replace the standard Inception modules with depthwise separable convolutions, which significantly reduce the number of parameters in the network. It is 71 layers deep with 36 convolutional layers and 22.9 million parameters.

Experimental Setup

We have conducted five experiments, each of which examines one pre-trained model. The weights

of pre-trained models are frozen, whereas two dense, one flattened, and one dropout layers are added on top of each. The input images are resized to 224×224 pixels and used in a batch size of 64. RMSProp optimizer is used with categorical cross entropy as the loss function. The number of epochs is set to 30. However, the early stopping regularization technique is implemented to prevent overfitting. All experiments are carried out on Google Colab.

Results and Discussion

The vehicle accident classification model is developed by evaluating five pre-trained models over accuracy, precision, recall and F1 score. Table 2 presents the summary of all experiments in terms of accuracy. ResNet50 exhibits poor performance in all stages while VGG16 and DenseNet121 have achieved more than 95% accuracy in each phase. Detailed results of each experiment are discussed in successive subsections.

DenseNet121

DenseNet is one of the efficient pre-trained models that encourage feature reuse with reduced number of parameters. Figure 3 shows the accuracy and loss curves developed during training that clearly marks its efficiency. It achieves 96.22% test accuracy after getting trained for 30 epochs. The model has shown greater performance for identifying true positives and false positives as shown in Table 3.

InceptionNetV3

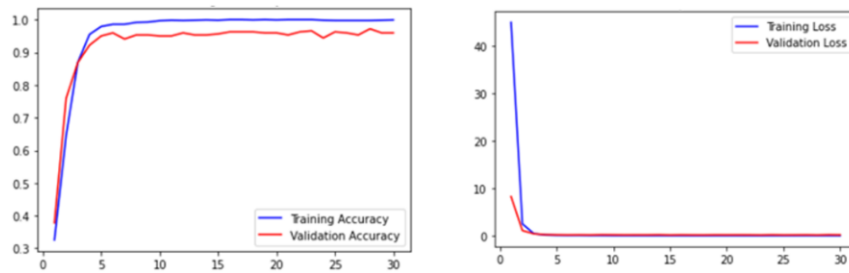
The InceptionNetV3 model achieves 94.05% test accuracy in 30 epochs with 100% training accuracy. It is observed that this model would be improved if the training had continued for more epochs. However, to evaluate each model on same parameters, the number of epochs was not increased. Figure 4 shows its learning curves while Table 4 indicates its precision, recall and F1 score.

ResNet50

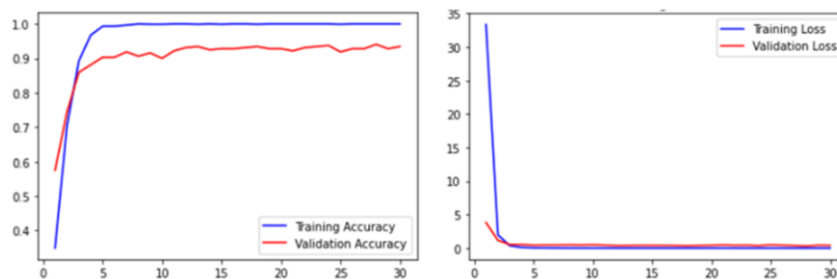
ResNet50 is developed to overcome the optimization problem of deep neural networks.

Table 3. DenseNet121 classification report

Class	Precision (%)	Recall (%)	F1-Score (%)
Head_On_Collision	100	86	92
No_Collision	97	97	97
Rear_end_Collision	100	100	100
Side_impact_Collision	93	100	96
Single_vehicle_Collision	91	97	94

**Figure 3. DenseNet121 learning curves****Table 4. InceptionNetV3 classification report**

Class	Precision (%)	Recall (%)	F1-Score (%)
Head_On_Collision	94	86	90
No_Collision	95	97	96
Rear_end_Collision	89	100	94
Side_impact_Collision	97	95	96
Single_vehicle_Collision	97	91	94

**Figure 4. InceptionNetV3 learning curves**

It has shown significant results in numerous studies. However, in this study's experimental setting this pre-trained model has shown the worst performance among others. The instability of the model is clearly evident from the learning curves in Figure 5.

Moreover, its precision, recall and F1 score have also not displayed satisfactory results as presented in Table 5. The No_Collision category got the highest precision value but lowest recall.

VGG16

This pre-trained model is the best performer among all models in the classification of accidents.

It achieved 100% training and 96.76% accuracy on 28th epoch as depicted in Figure 6.

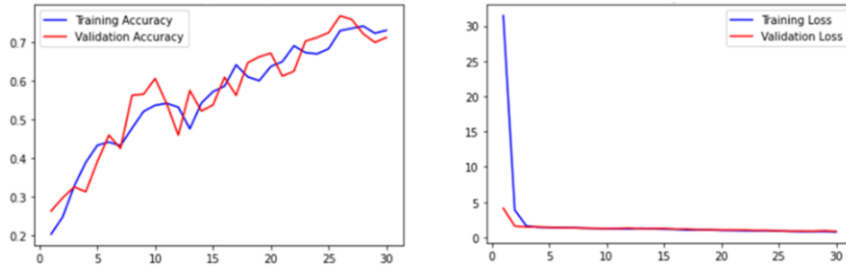
VGG16 has also shown best results in comparatively lesser time for precision, recall and F1 score as presented in Table 6.

Xception

This pre-trained model uses depth-wise separable convolutions and is considered as the improved version of inceptionNet. In this study, the model has achieved 92.43% test accuracy that is smaller than the test accuracy achieved by inceptionNetV3.

Table 5. ResNet50 classification results

Class	Precision (%)	Recall (%)	F1-Score (%)
Head_On_Collision	77	77	77
No_Collision	100	38	55
Rear_end_Collision	48	77	59
Side_impact_Collision	89	76	82
Single_vehicle_Collision	62	73	67

**Figure 5. ResNet50 learning curves****Table 6. VGG16 classification report**

Class	Precision (%)	Recall (%)	F1-Score (%)
Head_On_Collision	92	94	93
No_Collision	97	97	97
Rear_end_Collision	100	100	100
Side_impact_Collision	95	98	96
Single_vehicle_Collision	100	94	97

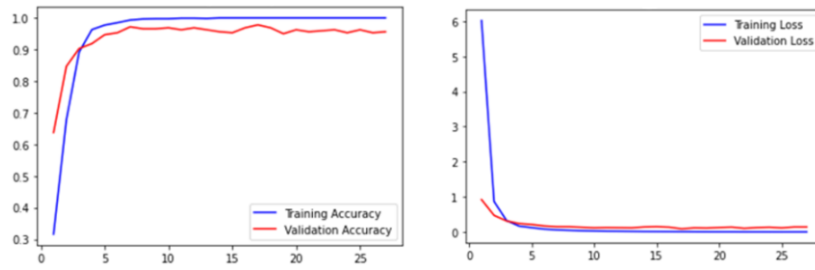
**Figure 6. VGG16 learning curves**

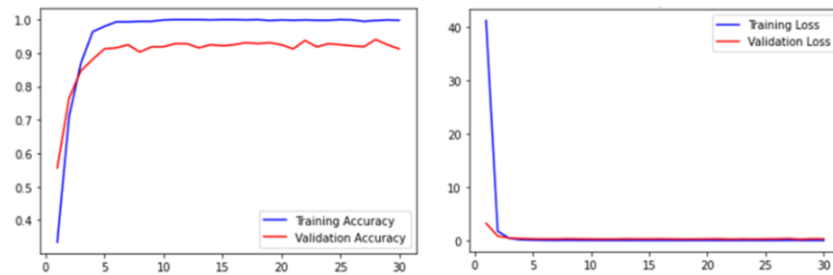
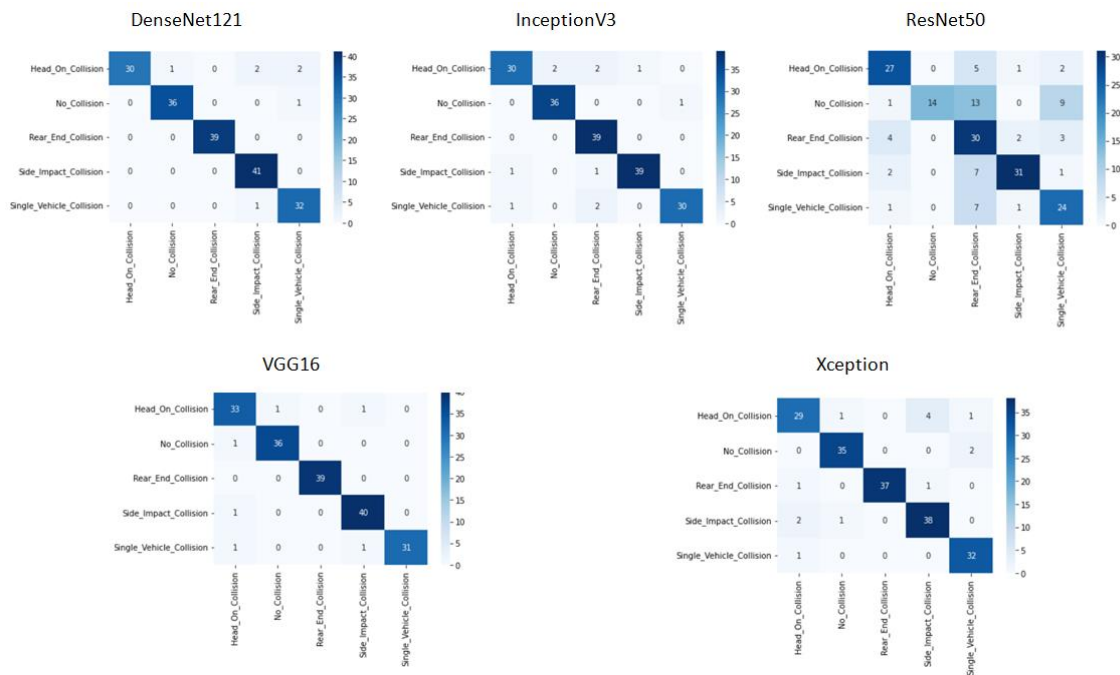
Figure 7 displays its learning graphs while Table 7 represents its precision, recall and F1 scores.

After analyzing each model, it is observed that VGG16 has outperformed all pre-trained models in classifying accidents. Figure 8 displays the confusion matrix generated on the test sets by each pre-trained model. The Rear_end_Collision class seems to be the most confusing for all models except VGG16 as it has obtained 100% test accuracy for this class Overall, VGG16 has attained 96.76% test accuracy. It can be due to the large parameter count empowering the model to discern intricate visual patterns within accident images. Moreover, the consistent use of 3×3 kernel sizes across

all convolutional layers simplifies the learning process, enabling VGG16 to efficiently capture a wide range of visual features, from basic edges to complex object arrangements. DenseNet121 has also shown high performing results i.e., 96.22% test accuracy but requires more time to converge than VGG16 as it has comparatively low number of parameters. InceptionNetV3 and Xception have shown average but stable results as compared to VGG16 and achieved 94.05% and 92.43% test accuracies respectively. On the other hand, ResNet50 has demonstrated poor performance during training and only achieves 67.03% test accuracy. Thus, for classifying accidents VGG16 is

Table 7. Xception classification report

Class	Precision (%)	Recall (%)	F1-Score (%)
Head_On_Collision	88	83	85
No_Collision	95	95	95
Rear_end_Collision	100	95	97
Side_impact_Collision	88	93	90
Single_vehicle_Collision	91	97	94

**Figure 7. Xception learning curves****Figure 8. Confusion matrix of each pre-trained model**

found to be the most stable and fastest pre-trained model.

Conclusions

Traffic accidents are increasing rapidly with the growing population. The surveillance cameras installed across the cities lack an automated

mechanism for identifying emergency conditions and typically need manual monitoring. This study presents an accident detection deep learning model that analyzes traffic videos and classifies vehicle accidents into four categories according to the nature of collision. Experiments were conducted with five pre-trained models DenseNet121, InceptionNetV3, ResNet50, VGG16 and Xception. Among these five models, VGG16 has achieved

the highest performance, having 96.76% accuracy. To this end, the model can classify images according to the type of collision. However, this study has considered four common types of vehicle collisions. Therefore, to further advance the field of vehicle accident detection a threefold strategy can be adopted in future. First, the dataset diversity can be enhanced to include more types of accidents for better generalization. Second, to explore the model's robustness more extensive experimentation on large scale could be conducted. Third, extending research to analyze collision intensity and identify various road anomalies offers promising directions for reducing traffic accidents and enhancing accident detection systems. Nevertheless, understanding collision types plays a pivotal role in improving safety, resource allocation, insurance, law enforcement, vehicle safety, and traffic engineering.

Acknowledgement

This research is supported by the Higher Education Commission (HEC), Pakistan, NRP grant 15873.

References

- Adel, A., Gogate, M., and Hussain, A. (2020). Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments. *Information Fusion*, 59:163-170. <https://doi.org/10.1016/j.inffus.2019.08.008>
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria J., Fadhel, M.A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53. <https://doi.org/10.1186/s40537-021-00444-8>
- Basiri, M., Azim, A.Z., and Farrokhi, M. (2017). Smart city solution for sustainable urban development. *European Journal of Sustainable Development*, 6(1):71-84. <https://doi.org/10.14207/ejsd.2017.v6n1p71>
- Chan, F.H., Chen, Y.T., Xiang, Y., and Sun, M. (2017). Anticipating accidents in dashcam videos. In: *Computer Vision - ACCV 2016*. Lai, S.H., Lepetit, V., Nishino, K., and Sato, Y. (eds). Springer, Cham, 10114:136-153. https://doi.org/10.1007/978-3-319-54190-7_9
- Chand, D., Gupta, S., and Kavati, I. (2020). Computer vision based accident detection for autonomous vehicles. In: *2020 IEEE 17th India Council International Conference (INDICON)*, New Delhi, India, p. 1-6. <https://doi.org/10.1109/INDICON49873.2020.9342226>
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, p. 1800-1807. <https://doi.org/10.1109/CVPR.2017.195>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, p. 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Dogru, N. and Subasi, A. (2018). Traffic accident detection using random forest classifier. In: *2018 15th Learning and Technology Conference (L&T)*, Jeddah, Saudi Arabia, 2018, p. 40-45. <https://doi.org/10.1109/LT.2018.8368509>
- Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115-118. <https://doi.org/10.1038/nature21056>
- Farman, H., Khan, Z., Jan, B., Boulila, W., Habib, S., and Koubaa, A. (2022). Smart transportation in developing countries: an internet-of-things-based conceptual framework for traffic control. *Wireless Communications and Mobile Computing*, 2022:Article ID 8219377. <https://doi.org/10.1155/2022/8219377>
- Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., and Vento, M. (2015). Audio surveillance of roads: A system for detecting anomalous sounds. *IEEE Transactions on Intelligent Transportation Systems*, 17(1):279-288. <https://doi.org/10.1109/TITS.2015.2470216>
- Ghosh, S., Sunny, S.J., and Roney, R. (2019). Accident detection using convolutional neural networks. In: *2019 International Conference on Data Science and Communication, (IconDSC)*, Bangalore, India, 2019, p. 1-6. <https://doi.org/10.1109/IconDSC.2019.8816881>
- Gupta, G., Singh, R., Singh Patel, A., Ojha, M. (2021). Accident detection using time-distributed model in videos. In: Yang, X.S., Sherratt, S., Dey, N., and Joshi, A. (eds). *Proceedings of Fifth International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing*, Springer, Singapore, 1184:214-223. https://doi.org/10.1007/978-981-15-5859-7_21
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, p. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, p. 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>
- Kelley, M.E., Talton, J.W., Weaver, A.A., Usoro, A.O., Barnard, E.R., and Miller, A.N. (2019). Associations between upper extremity injury patterns in side impact motor vehicle collisions with occupant and crash characteristics. *Accident Analysis & Prevention*, 122:1-7. <https://doi.org/10.1016/j.aap.2018.09.029>
- Kim, K., Nitz, L., Richardson, J., and Li, L. (1994). Analyzing the relationship between crash types and injuries in motor vehicle collisions in Hawaii. *Transportation Research Record*, 1467:9-13.
- Koppe, G., Meyer-Lindenberg, A., and Durstewitz, D. (2021). Deep learning for small and big data in psychiatry. *Neuropsychopharmacology*, 46(1):176-190. <https://doi.org/10.1038/s41386-020-0767-z>
- Maaloul, B., Taleb-Ahmed, A., Niar, S., Harb, N., and Valderrama, C. (2017). Adaptive video-based algorithm for accident detection on highways. In: *2017 12th IEEE International Symposium on Industrial Embedded Systems (SIES)*, Toulouse, France, 2017, p. 1-6. <https://doi.org/10.1109/SIES.2017.7993382>
- Nagpal, K., Foote, D., Liu, Y., Chen, P.-H.C., Wulczyn, E., Tan, F., Olson, N., Smith, J.L., Mohtashamian, A., Wren, J.H., Corrado, G.S., MacDonald, R., Peng, L.H., Amin, M.B., Evans, A.J., Sangoi, A.R., Mermel, C.H., Hipp, J.D., and Stumpe, M.C. (2019). Development and validation of a deep learning algorithm for improving Gleason scoring of

- prostate cancer. *NPJ Digital Medicine*, 2(1):48. <https://doi.org/10.1038/s41746-019-0112-2>
- O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv:1511.08458*. <https://doi.org/10.48550/arXiv.1511.08458>
- Papa, R., Gargiulo, C., and Galderisi, A. (2013). Towards an urban planners' perspective on Smart City. *TeMA Journal of Land Use, Mobility and Environment*, 6(1):5-17. <https://doi.org/10.6092/1970-9870/1536>
- Rajesh, G., Benny, A.R., Harikrishnan, A., Jacob Abraham, J., and John, N.P. (2020). A deep learning based accident detection system. In: 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, July 2020, p. 1322-1325. <https://doi.org/10.1109/ICCSP48568.2020.9182224>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., and Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- Saravananarajan, V.S., Chen, R.-C., Dewi, C., Chen, L.-S., and Ganesan, L. (2023). Car crash detection using ensemble deep learning. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-023-15906-9>
- Shah, A.P., Lamare, J.B., Nguyen-Anh, T., and Hauptmann, A. (2018). CADP: A novel dataset for cctv traffic camera based accident analysis. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, p. 1-9. <https://doi.org/10.1109/AVSS.2018.8639160>
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- Singh, D. and Mohan, C.K. (2022). Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):879-887. <https://doi.org/10.1109/TITS.2018.2835308>
- Sun, L.-L., Liu, D., Chen, T., and He, M.-T., (2019). Road traffic safety: An analysis of the cross-effects of economic, road and population factors. *Chinese Journal of Traumatology*, 22(5):290-295. <https://doi.org/10.1016/j.cjtee.2019.07.004>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, p. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Torrey, L. and Shavlik, J. (2010). Transfer learning. In: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global, p. 242-264. <https://doi.org/10.4018/978-1-60566-766-9.ch011>
- WHO. (2022). Road Traffic Injuries. [Online]. Available from: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>. Accessed date: Aug. 29, 2022.
- Yao, Y., Wang, X., Xu, M., Pu, Z., Atkins, E., and Crandall, D. (2020). When, Where, and What? A New Dataset for Anomaly Detection in Driving Videos. [Online]. Available: <http://arxiv.org/abs/2004.03044> Accessed: Aug 02, 2022.
- Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3):55-75. <https://doi.org/10.1109/MCI.2018.2840738>