

Visual Classification of Vehicle Types Using Transfer Learning

Misbah Jamil ¹, Anjum Rehman ^{2,*} and Rauf Ahmad ³

¹ Department of Computer Science, Swedish College of Engineering & Technology, Rahim Yar Khan, 64200, Pakistan

² Department of Information Security, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

³ Department of Electrical Engineering, Swedish College of Engineering & Technology, Rahim Yar Khan, 64200, Pakistan

*Corresponding author: Anjum Rehman (Email: anjum.rehman@kc.au.edu.pk).

Received: 02/07/2026, Revised: 20/09/2026, Accepted: 30/09/2026

Abstract— Road traffic crashes remain a major cause of death among children and young people, and reliable visual recognition of the vehicles involved can support later accident documentation. Vehicle type classification is well studied on ordinary traffic imagery, whereas accident-related use requires models that remain informative when vehicles are damaged, occluded, or poorly illuminated. This paper reports a four-class vehicle type classification experiment that uses transfer learning with the Visual Geometry Group 16-layer network (VGG16). A curated corpus of 1,750 images was assembled from publicly accessible web sources through a manually supervised image-retrieval process and was manually annotated before being divided into training, validation, and testing partitions (1,260 / 140 / 350). Augmentation was applied only during training. On the held-out test set of 350 images, the classifier attained 89% overall accuracy and a macro-averaged F1-score of 0.89. Class-wise F1-scores were 0.95 for bike, 0.88 for bus, and 0.87 for both car and truck. Training curves show mild overfitting on the validation split, while the separate test partition confirms balanced four-class performance. The study is limited to a VGG16 baseline; comparisons with residual and transformer architectures, together with robustness and latency tests, remain future work. Downstream effects on emergency response or smart city operations were not measured. The contribution is a transparent VGG16 baseline for four-class vehicles typing, a task-aware literature synthesis, and a clear agenda for subsequent refinement.

Index Terms— Convolutional neural networks, deep learning, transfer learning, vehicle classification, VGG16, Visual recognition

I. INTRODUCTION

Road traffic crashes remain a serious public health concern and a leading cause of death among children and young people aged 5 to 29 years. The World Health Organization (WHO) estimated about 1.19 million road traffic deaths in 2021, together with many millions of non-fatal injuries [1]. As traffic volumes grow, clear documentation of crashes matters for emergency response, insurance and legal work, and road safety policy. Computer vision systems are now widely used for traffic

monitoring and for reviewing events after they occur, and deep learning is the main tool behind most of these image recognition tasks.

In this setting, the type of vehicle involved in a crash is useful information. Heavy vehicles, passenger cars, and bikes are linked with different injury patterns, infrastructure needs, and reporting practices. A system that can label a still image as a given vehicle type can therefore support later accident analysis. Convolutional neural networks (CNNs), especially transfer learning from large backbones pretrained on ImageNet such as the Visual Geometry Group 16-layer network (VGG16) [2], remain a practical starting point when labeled crash images are limited. Newer models, including residual networks [3], compound-scaled convolutional models such as EfficientNet [4], and Vision Transformers (ViTs) [5], now lead many general vehicle-recognition benchmarks. Accident-detection studies likewise use EfficientNet, You Only Look Once (YOLO) detectors, and transformer ensembles [6]–[9].

Even so, most published models either recognize vehicles in ordinary traffic scenes or detect whether an accident is present. They often do not identify the vehicle type shown in an accident image. Accident scenes are also harder than routine traffic views: vehicles may be deformed, inverted, partly burned, or heavily occluded; lighting may be poor; and several vehicles may share one frame. A model trained only on clean, centered vehicles may therefore look stronger than it is for crash documentation. This study focuses on that need by treating vehicle type classification as its own task, separate from binary accident detection and from object detection.

The aim of this study is to develop and evaluate a transfer-learning classifier that assigns a still image of a road vehicle to one of four types (bike, bus, car, or truck) using the VGG16 architecture. The specific objectives are: (i) to assemble and manually annotate a four-class vehicle-image dataset; (ii) to adapt a VGG16 backbone pretrained on ImageNet to the four-class task through transfer learning; (iii) to evaluate the



classifier using overall accuracy together with class-wise precision, recall, and F1-score (the harmonic mean of precision and recall); and (iv) to provide a clear baseline for later comparison with modern architectures and robustness tests.

The study is guided by four research questions (RQs). RQ1 examines the performance of the VGG16 transfer-learning model for classifying the four vehicle types on the available test images. RQ2 considers comparative evaluation with modern convolutional and transformer-based architectures under an identical data split. RQ3 concerns model robustness under challenging visual conditions, including low illumination, occlusion, and deformation, while RQ4 addresses the trade-off between classification performance and computational cost. The present study experimentally addresses RQ1. The remaining questions define planned extensions of the current baseline and are therefore discussed as directions for future comparative, robustness, and efficiency analyses rather than as outcomes of the present experiment.

The work has both scientific and practical value. Scientifically, it reports a VGG16 vehicle-type classifier with class-wise metrics and shows that coarse four-class recognition is feasible with a conventional backbone. Practically, a reliable vehicle-type label could later support accident records, insurance or law-enforcement review, and traffic-monitoring pipelines. Because only the classification task was measured, these uses are presented as potential applications rather than experimentally demonstrated improvements in response time, legal outcomes, or operational performance.

This paper reports the completed experiment: four-class vehicle-type classification with a VGG16 transfer-learning pipeline. On 350 held-out test images, the model reached 89% overall accuracy and a macro-averaged F1-score of 0.89. A separate YOLO-based detector trained on the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) benchmark [13] is not used as evidence for the classification task because KITTI covers ordinary driving scenes and detection mean Average Precision (mAP) is not equivalent to classification accuracy.

The contribution of this study is not the introduction of a new neural-network architecture. Instead, the study provides a domain-specific experimental baseline for vehicle-type classification in accident-related imagery. Its contributions are threefold. First, it formulates accident-scene vehicle recognition as a four-class image-classification problem involving bike, bus, car, and truck, explicitly distinguishing this task from binary accident detection, object detection, and vehicle-damage classification. Second, it provides a transparent transfer-learning implementation using an ImageNet-pretrained VGG16 backbone and reports overall as well as class-wise precision, recall, and F1-score on a held-out test partition. Third, it documents the dataset construction, preprocessing, augmentation, training configuration, and evaluation procedure in sufficient detail to establish a reproducible baseline for subsequent comparison with modern architectures and accident-specific robustness evaluations.

Accordingly, the study should be interpreted as a baseline and feasibility investigation rather than as a novel deep-learning

architecture. Its value lies in defining and evaluating a specific accident-imagery classification task and establishing a measurable reference point for future work on model comparison, event-level generalization, visual robustness, and computational efficiency.

II. LITERATURE REVIEW

This section reviews work close to the present study: image-based vehicle recognition on modest datasets, transfer learning with convolutional networks, and vision systems used in road safety or surveillance. The discussion moves from vehicle type classification to accident-related detection and then to the remaining gaps. Large fine-grained benchmarks with very high accuracy on many classes [10], [11] are noted only briefly, because their labels, corpus sizes, and viewing conditions differ from the four-class setting used here.

A. Vehicle Classification with Convolutional Networks

Deep learning has substantially improved image-based vehicle recognition, with transfer learning remaining a practical approach when task-specific labeled data are limited. Earlier convolutional architectures such as VGG16 established strong image-classification baselines, while residual networks and more recent architectures such as EfficientNet and Vision Transformers have provided improved representation capacity and computational alternatives in a range of vehicle-recognition settings. Existing studies have demonstrated successful classification across different numbers of vehicle categories, datasets, and levels of label granularity. However, these results are predominantly reported on conventional traffic imagery in which vehicles are relatively visible, upright, and captured under normal operating conditions.

Transfer learning from CNNs pretrained on ImageNet is still a practical choice when labeled vehicle images are limited. VGG16, built from stacked 3x3 convolutions and fully connected layers [2], remains a common baseline. Residual networks [3], EfficientNet [4], and ViTs [5] have improved many recognition benchmarks, and recent surveillance and fine-grained studies continue to compare these families under transfer learning [8], [12], [15], [20], [22]. At the same time, several applied studies still rely on compact datasets and classical CNN backbones.

Zhang and Feng combined Residual Network 34 (ResNet34) with a vehicle-oriented augmentation scheme and raised test accuracy from 75.42% to 80.0% [16]. The gain shows that careful data design helps when vehicle types look similar, yet the accuracy remains below the four-class result reported in this paper. Ismail et al. tested Single Shot Detector (SSD), CenterNet, and EfficientDet on 1,263 images from five vehicle classes and obtained a best F1 score of 0.731 with SSD [17]. That setting is closer in scale to the present study than large make and model corpora, and its F1 score is lower than the macro averaged F1 score of 0.89 reported below. Beg et al. trained YOLOv8 on about 1,000 Malaysian roadside images and reported an F1 score of 71%, with precision 88.2% and recall 84% [18]. Their goal was driving assistance rather than post event vehicle typing. Hasan et al. used Residual Network

50 (ResNet-50) with transfer learning on 10,440 images of 13 native vehicle types and compared it with AlexNet and VGG-16 [20], which again shows that VGG style networks remain useful baselines. Ul Khairi et al. compared MobileNet-v2, Inception-v3, VGG-19, and ResNet-50 for fine grained vehicle recognition [22], while Mani et al. studied hybrid machine learning pipelines for real time vehicle classification [23].

The distinction between conventional traffic imagery and accident-scene imagery is important. Accident scenes can contain damaged or deformed vehicles, unusual orientations, partial visibility, severe occlusion, poor illumination, background clutter, and multiple vehicles within a single frame. These characteristics can alter the visual features available for classification and may reduce the reliability of models developed primarily from clean or routinely captured traffic images. Consequently, high performance on conventional vehicle-recognition benchmarks cannot be assumed to represent equivalent performance on accident-related imagery.

B. Accident-Related Visual Recognition

A second group of papers addresses road crashes directly, but usually as an event detection problem rather than as vehicle type classification. Tamagusko et al. applied EfficientNetB1 and MobileNetV2, including synthetic accident images, and reported mAP of 0.89 and 0.88 on roadside camera data [6]. Zahid et al. fine-tuned several CNNs on simulated accident frames and reported an 80% true positive rate for AlexNet on real surveillance footage [19]. Zhang and Sung combined ViT frame features with convolutional trajectory features for accident versus non accident classification and reported about 97% accuracy on frames from the CCTV Accident Detection and Prediction (CADP) resource [7]. CADP remains a major public accident video resource [9].

These studies are useful for alerting systems, yet they leave a clear gap for the present purpose. Knowing that an accident has occurred is not the same as knowing whether the vehicles in the frame are bikes, buses, cars, or trucks. Damage-classification work, such as that of Lee et al. on broken, crushed, and undamaged cars [14], also assumes that the object is already known to be a car. Darji et al. studied transfer-learning CNNs for identifying vehicles in hit-and-run settings [24], which is closer to post-event documentation but still differs from a transparent four-class type classifier with class-wise reporting. In short, accident detection and damage typing answer related questions, but they do not replace a dedicated vehicle-type classifier for crash imagery.

C. Detection Versus Classification

Object detectors such as YOLO and SSD locate and categorize objects in one pass and are typically scored with mAP. Image classifiers assign a single label to an image or crop and are scored with accuracy, precision, recall, and F1 score. These two families of metrics are not interchangeable. Benchmarks such as KITTI [13] further show the distinction: they support detection research on ordinary driving scenes and are not, by themselves, accident scene classification corpora. For that reason, detector results are treated here as related

context rather than as direct substitutes for the four-class classification experiment.

D. Research Gap

Taken together, the literature points to three linked gaps. First, strong vehicle-recognition results are often obtained on ordinary traffic imagery, either with few classes and modest accuracy [16]-[18] or with many fine-grained classes and large corpora that are not crash-oriented [10], [11], [20], [22]. Second, studies that use accident imagery usually detect the event or the damage state, not the vehicle type [6], [7], [14], [19], [24]. Third, few applied papers combine a transparent transfer-learning baseline, an approximately balanced four-class test set, and class-wise precision, recall, and F1-score for bike, bus, car, and truck [23], [25]. The present study addresses this middle ground by adapting VGG16 to a four-class vehicle-type task, evaluating it on 350 held-out images, and reporting both overall and class-wise metrics. A numerical comparison with selected previous studies is given later in Table IV.

III. METHODOLOGY

A. Task Definition

The task is four-class vehicle-type classification from a still image. The classes are bike (labeled Bike in the experimental files), bus, car, and truck. The model assigns one class to each image. The study is not a binary accident-versus-non-accident detector, an object detector, or a damage classifier. Results from those other tasks are not used to support the classification claims.

B. Dataset

The dataset comprises 1,750 vehicle images representing four vehicle categories: bike, bus, car, and truck. All images were acquired from publicly accessible web sources through a manually supervised web-scraping and image-retrieval process conducted over approximately six months. No images were collected directly from CCTV systems, road-surveillance cameras, or other private surveillance infrastructure. The scraping/retrieval tool was used to facilitate image acquisition, whereas image screening, class assignment, quality control, and dataset organization were performed manually. The collected images were reviewed individually and assigned to one of the four target categories. Where required, vehicles were manually isolated from the original scene before classification, meaning that the reported experiment evaluates vehicle-type classification on prepared image regions rather than automatic vehicle localization in complete accident scenes. Images that were corrupted, incorrectly labeled, or unsuitable for the target classification task were removed during quality control. Where necessary, the target vehicle was manually cropped from the source image to reduce surrounding scene clutter and ensure that the vehicle represented the dominant visual content presented to the classifier. The cropped region was then resized and processed in the same manner as the remaining images before class assignment and model input. This preparation step was applied only to isolate the vehicle of interest; therefore, the reported results relate to vehicle-type classification from

prepared image regions rather than automatic localization or end-to-end recognition within complete accident scenes.

The dataset was divided into training, validation, and test partitions containing 1,260, 140, and 350 images, respectively. The test partition contained 81 bike images, 88 bus images, 89 car images, and 92 truck images, providing an approximately balanced evaluation across the four classes. The complete dataset contained 405 bike images, 440 bus images, 445 car images, and 460 truck images before partitioning. The training and validation subsets were formed from the remaining images after allocation of the held-out test samples. Near-duplicate images from the same capture session were reduced during quality control; however, an explicit incident-level identifier was not retained. Consequently, the possibility that visually related images from the same incident may occur across partitions cannot be completely excluded and is treated as an important limitation of the evaluation.

Because the dataset combines multiple public image sources, source composition should be considered when interpreting generalization. The experiment evaluates performance on the available curated corpus rather than claiming generalization to all road-accident imagery, geographic regions, camera systems, or accident conditions (Table I).

TABLE I

DATASET INVENTORY FOR THE FOUR-CLASS EXPERIMENT

Vehicle class	Original	Cleaned	Train	Validation	Test
Bike	405	405	292	32	81
Bus	440	440	317	35	88
Car	445	445	320	36	89
Truck	460	460	331	37	92
Total	1,750	1,750	1,260	140	350

C. Preprocessing and Augmentation

Images were resized to $224 \times 224 \times 3$ and rescaled by $1/255$, consistent with the VGG16 input convention [2]. Augmentation was applied only to the training split using TensorFlow ImageDataGenerator with rotation up to 20 degrees, horizontal flipping, a zoom range of 0.2, and a brightness range of [0.8, 1.2]. Validation and test images were left unaltered apart from resizing and rescaling. No ablation of augmentation was performed in this study.

D. VGG16 Baseline

VGG16 contains 13 convolutional layers and three fully connected layers [2]. The network was loaded with ImageNet weights and without its original top layer (include_top=False), and the convolutional base was frozen. A custom head was then added: global average pooling, a dense layer with 256 units and Rectified Linear Unit (ReLU) activation, dropout with rate 0.5, and a four-unit softmax output trained with categorical cross entropy. Figure 1 shows the standard ImageNet topology; the 1000-way head was replaced by the four-class head described above.

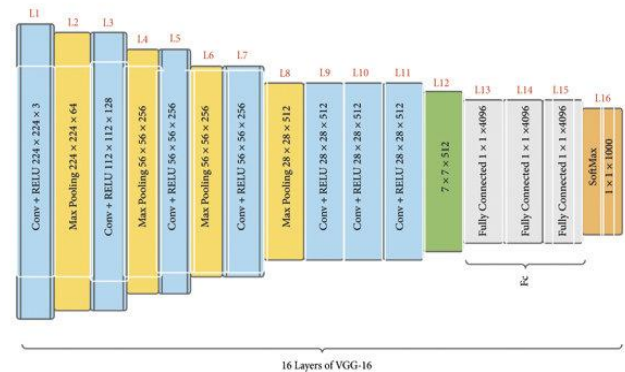


Fig. 1. Standard VGG16 topology. For this study, the ImageNet 1000-way softmax was replaced by a custom head (global average pooling, dense ReLU with 256 units, dropout 0.5, and a four-way softmax) on a frozen convolutional base.

E. Training Protocol

The classifier was implemented using TensorFlow 2.12 and the Keras API. VGG16 was initialized with ImageNet-pretrained weights, its convolutional base was kept frozen, and only the newly added classification head was optimized during training. The classification head consisted of GlobalAveragePooling2D, a 256-unit fully connected layer with ReLU activation, dropout with a rate of 0.5, and a four-unit softmax output layer. Categorical cross-entropy was used as the loss function and Adam was used as the optimizer with a learning rate of 1×10^{-4} . Training was performed for eight epochs using a batch size of 32.

All input images were resized to 224×224 pixels and rescaled by $1/255$. Data augmentation was applied dynamically to the training partition only, using rotations of up to $\pm 20^\circ$, horizontal flipping, zoom augmentation of 0.2, and brightness variation within the range [0.8, 1.2]. Validation and test images were not augmented and were subjected only to the required resizing and rescaling operations. A fixed random seed of 42 was used for the reported experimental run. Early stopping was monitored using validation loss but was not activated within the eight-epoch training schedule.

The test partition was kept separate from model training and validation and was used only for final performance evaluation. The reported results therefore correspond to a held-out test evaluation rather than performance measured on the training data. The experimental configuration, including the model architecture, optimization settings, data split, augmentation parameters, software environment, hardware, and random seed, is reported to facilitate replication (Table II).

TABLE II
TRAINING CONFIGURATION FOR THE VGG16 TRANSFER-LEARNING EXPERIMENT

Setting	Value
Backbone	VGG16, ImageNet-pretrained, include_top=False, base frozen
Custom head	GlobalAveragePooling2D, Dense (256, ReLU), Dropout (0.5)
Output classes	4 (Bike, Bus, Car, Truck)

Output activation / loss	Softmax / categorical cross-entropy
Trainable layers	Custom head only
Optimizer	Adam
Learning rate	1×10^{-4}
Data split	Train 1,260 / Val 140 / Test 350
Augmentation	Rotation up to 20 deg, horizontal flip, zoom 0.2, brightness [0.8, 1.2] (train only)
Batch size	32
Epochs	8
Framework	TensorFlow 2.12 / Keras
Hardware	NVIDIA Tesla T4 GPU (Google Colab)
Random seed	42

F. Evaluation Metrics

The classifier is evaluated using overall accuracy and per-class precision, recall, and F1-score. Macro-averaged and weighted-averaged summaries are also reported. Because the test set is nearly balanced, the macro-averaged F1-score is an appropriate headline metric alongside accuracy. Inference latency and model size were not the focus of this experiment and are left for future measurement.

G. Scope of the Present Experiment

The present study is designed as a controlled baseline evaluation rather than an exhaustive comparison of contemporary deep-learning architectures. The completed experiment evaluates a single VGG16 transfer-learning configuration under a fixed four-class dataset split. Accordingly, the reported results establish the performance of the selected baseline on the available test corpus but do not establish that VGG16 is superior to alternative convolutional or transformer-based architectures. Similarly, the experiment does not quantify the independent contribution of augmentation or fine-tuning, nor does it establish performance under systematically controlled image degradation or computational constraints. These factors are therefore treated as limitations of the present evaluation rather than as conclusions of the study.

IV. RESULTS AND DISCUSSION

A. Four-Class Test Performance

Table III summarizes the held-out test performance of the VGG16 classifier on 350 images. The model achieved an overall accuracy of 0.89, with macro-averaged precision, recall, and F1-score of 0.90, 0.89, and 0.89, respectively. The corresponding weighted averages were 0.89 for precision, recall, and F1-score. Among the four classes, bike achieved the highest performance, with precision of 0.97, recall of 0.93, and F1-score of 0.95. Bus, car, and truck obtained F1-scores of 0.88, 0.87, and 0.87, respectively. Car had the lowest precision (0.84) but a comparatively high recall (0.91), indicating that some samples from other classes were assigned to the car category. The confusion matrix in Fig. 3 provides a class-level view of these cross-class errors, while Fig. 2 presents the corresponding

precision, recall, and F1-score values across the four vehicle categories.

The reported accuracy of 0.89 is interpreted in light of the dataset-partitioning constraints described in Section III-B. Although the test images were excluded from model training and validation, no incident-level identifier was retained during dataset construction. Consequently, visually related images originating from the same accident event cannot be completely ruled out across different partitions. The reported accuracy therefore reflects performance on the present held-out image partition and does not constitute an event-independent estimate of generalization to entirely unseen accident incidents.

TABLE III
CLASSIFICATION REPORT FOR THE FOUR-CLASS VGG16 EXPERIMENT

Class	Precision	Recall	F1-score	Support
Bike	0.97	0.93	0.95	81
Bus	0.90	0.86	0.88	88
Car	0.84	0.91	0.87	89
Truck	0.87	0.87	0.87	92
Accuracy	-	-	0.89	350
Macro average	0.90	0.89	0.89	350
Weighted average	0.89	0.89	0.89	350

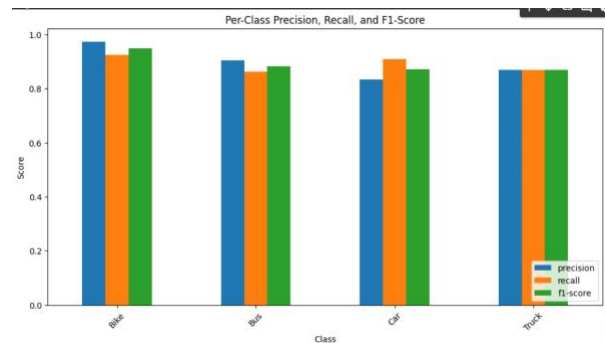


Fig. 2. Per-class precision, recall, and F1 score for bike, bus, car, and truck.

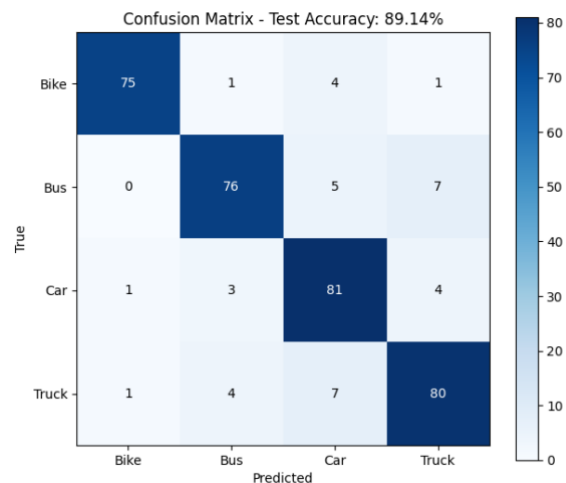


Fig. 3. Four-class confusion matrix.

B. Training and Validation Behavior

Fig. 4 shows the learning curves that accompany Table III. Training accuracy rises from about 0.48 to about 0.81 over eight epochs, while validation accuracy follows a non-monotonic path and ends near 0.72-0.73. Training loss decreases steadily, whereas validation loss decreases, then rises, and then falls again, indicating mild overfitting on the validation split. The final validation accuracy is therefore lower than the held-out test accuracy of 0.89. This difference is reported without treating the higher test score as evidence of superior generalization. The validation set contains 140 images and was used for model monitoring, whereas the test set contains 350 separate images and was used only for final evaluation; differences in sample composition between these fixed partitions can affect the observed accuracies. Because the study used a single split and did not retain incident-level identifiers, the 0.89 test result is interpreted strictly as performance on the present held-out partition rather than as an event-independent estimate of generalization.

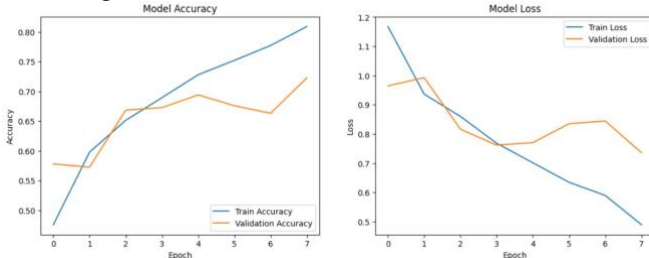


Fig. 4. Training and validation accuracy (left) and loss (right) over eight epochs. The held-out test accuracy corresponding to Table III is 0.89.

C. Error Analysis

The class-wise results indicate that bike achieved the strongest performance, whereas bus, car, and truck produced more closely grouped F1-scores. The relatively lower precision observed for the car class, combined with its comparatively high recall, indicates that some test samples from other categories were predicted as car. The confusion matrix in Fig. 3 substantiates the direction and frequency of these cross-class errors, with the off-diagonal entries concentrated mainly among bus, car, and truck. However, the matrix does not by itself establish the visual causes of the errors. Similarities in vehicle shape, partial visibility, viewpoint variation, background clutter, or image degradation remain plausible explanations that would require inspection of individual misclassified samples for confirmation.

D. Comparison with Previous Studies

Table IV places the present results beside selected previous studies that are close in dataset scale, class count, or road safety context. Direct numerical ranking is avoided where the tasks differ. The purpose of the table is to show where the present four-class classifier sits relative to related work.

TABLE IV
COMPARISON WITH PREVIOUS STUDIES

Study	Task	Dataset / classes	Method	Reported result	Relation to this study
-------	------	-------------------	--------	-----------------	------------------------

Zhang and Feng [16]	Vehicle classification	Vehicle images; multi-type	ResNet34 + augmentation	Accuracy 80.0% (from 75.42%)	Lower accuracy ; ordinary classification setting
Ismail et al. [17]	Vehicle recognition	1,263 images; 5 classes	SSD (best among tested)	F1 0.731; P 0.721; R 0.741	Smaller F1; detection-oriented pipeline
Beg et al. [18]	Roadside detection	~1,000 images ; vehicles, bikes, lamps	YOLOv8	F1 71%; P 88.2%; R 84%	Lower F1; driving assistance focus
Zahid et al. [19]	Accident detection	Simulated + surveillance frames	Fine-tuned CNNs (AlexNet best)	80% true positive rate	Event detection, not vehicle typing
Tamag usko et al. [6]	Accident detection	Roadside cameras + synthetic	Efficient NetB1 / MobileNetV2	mAP 0.89 / 0.88	Event detection n metric; not 4 class typing
This study	Vehicle type classification	4 classes; n_test = 350	VGG16 transfer learning	Accuracy 0.89; macro F1 0.89	Class-wise report for Bike, Bus, Car, Truck

Relative to Zhang and Feng [16], who reported 80.0% accuracy after augmentation with ResNet34, the present VGG16 classifier reaches a higher overall accuracy of 89% on a four-class test set. Relative to Ismail et al. [17] and Beg et al. [18], whose best F1-scores were 0.731 and 0.71 on modest roadside or traffic corpora, the present macro-averaged F1-score of 0.89 is higher, although those studies emphasize detection rather than single-label classification. Accident-detection work [6], [19] remains complementary: it answers whether a crash is present, whereas Table III addresses which vehicle type is present. Large fine-grained benchmarks that exceed 96% accuracy on many classes [10], [11] are not treated as direct competitors here because their corpora and label spaces differ from the four-class problem addressed in this paper.

E. Interpretation and Practical Implications

The completed experiment shows that a conventional VGG16 transfer-learning pipeline can separate four coarse vehicle types with balanced class-wise performance on the

available test images (Table III and Figs. 2-3). Bike images were the easiest to recognize in this corpus, while bus, car, and truck remained closer in performance. The learning curves in Fig. 4 show mild overfitting on the validation split, while the held-out test accuracy of 89% reflects performance on the specified test partition. In practical terms, a four-class vehicle tag may later support accident documentation, archive filtering, or human review in insurance and policing workflows. These examples represent potential downstream applications only; insurance, policing, emergency-response, and smart-city outcomes were not evaluated experimentally, and no claim is made that the present classifier improves operational performance in those settings.

F. Limitations

Several limitations should be kept in view when reading Table III and Table IV. The present experiment uses a single VGG16 baseline under one split and does not yet include architecture comparisons, augmentation ablations, or labeled robustness subsets for low light, blur, occlusion, or damage. Inference latency and model size were not measured. These points bound the contribution to a carefully reported four class baseline that can be extended in future work.

V. CONCLUSION

This study evaluated a VGG16-based transfer-learning classifier for four-class vehicle-type recognition involving bikes, buses, cars, and trucks in a curated corpus of accident-related imagery. On the held-out test partition of 350 images, the model achieved 0.89 overall accuracy and a macro-averaged F1-score of 0.89, with class-wise F1-scores of 0.95 for bike, 0.88 for bus, and 0.87 for both car and truck. These results demonstrate the feasibility of using a conventional transfer-learning backbone for coarse vehicle-type classification within the evaluated dataset.

The findings are interpreted within the experimental scope of the study. The model was evaluated under a single dataset partition and without direct comparison against modern architectures, controlled robustness subsets, or measured computational cost. In addition, vehicles were manually isolated before classification and an explicit incident-level identifier was not retained, limiting the extent to which the results can be generalized to complete accident scenes or unseen accident events. Therefore, the study does not establish end-to-end accident detection, real-time deployment capability, or improved emergency-response performance.

The primary contribution is a transparent baseline that defines and evaluates vehicle-type classification as a distinct task within accident-related visual analysis. Future work should strengthen this baseline through event-level data partitioning, comparative evaluation of modern and lightweight architectures, controlled robustness testing, repeated experimental runs, and computational-efficiency analysis. Such extensions would provide stronger evidence regarding generalization and practical applicability.

VI. FUTURE WORK

Several directions remain open. Stronger event-level splitting would further reduce the risk of leakage from near-duplicate or same-incident frames. Under an identical split, the VGG16 baseline could be compared with a residual or lightweight network and, where feasible, a transformer. Controlled studies of augmentation and fine-tuning, together with labeled low-light, blur, occlusion, and damage subsets, would clarify robustness. Measuring inference latency and model size on stated hardware would support later discussion of practical deployment. These proposed studies correspond to RQ2-RQ4 and remain future work; they are not part of the experimental claims reported in the present paper.

FUNDING STATEMENT

The author(s) received no specific funding for this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest to report regarding the present study.

AUTHOR CONTRIBUTIONS

Conceptualization, M.J. and A.R.; methodology, M.J.; software, M.J.; validation, M.J. and A.R.; writing: original draft preparation, M.J.; writing: review and editing, A.R.; supervision, A.R. data curation, A.R., and M.J.; investigation, M.J.; visualization, R.A. and A.R.; formal analysis, M.J.; resources, A.R., and R.A.;

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

Data is available on reasonable request.

CONSENT FOR PUBLICATION

All authors have read and approved the final version of the manuscript and consent to its publication.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING

For preparing this manuscript, the authors have only made use of generative artificial intelligence systems to enhance language proficiency, grammar, and readability.

REFERENCES

- [1] World Health Organization, Global status report on road safety 2023. Geneva, Switzerland: WHO, 2023. [Online]. Available: <https://www.who.int/publications/i/item/9789240086517>
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.
- [3] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778), doi: 10.1109/CVPR.2016.90.

- [4] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
- [5] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [6] Tamagusko T, Correia MG, Huynh MA, Ferreira A. Deep learning applied to road accident detection with transfer learning and synthetic images. *Transportation research procedia*. 2022 Jan 1;64:90-7. doi: 10.1016/j.trpro.2022.09.012.
- [7] Y. Zhang and Y. Sung, "Hybrid traffic accident classification models," *Mathematics*, vol. 11, no. 4, p. 1050, 2023, doi: 10.3390/math11041050.
- [8] Z. Seghir et al., "A real-time consensus-free accident detection framework for Internet of Vehicles using vision transformer and EfficientNet," *AI*, vol. 7, no. 1, p. 4, 2026, doi: 10.3390/ai7010004.
- [9] Shah AP, Lamare JB, Nguyen-Anh T, Hauptmann A. 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) 2018 Nov 27 (pp. 1-9). IEEE. doi: 10.1109/AVSS.2018.8639160.
- [10] Ma S, Yang JJ. Image-based vehicle classification by synergizing features from supervised and self-supervised learning paradigms. *Eng.* vol. 4, no. 1, pp. 444–456, 2023, doi: 10.3390/eng4010027.
- [11] Hayee S, Hussain F, Yousaf MH, Yasir M, Ahmad S, Viriri S. A benchmark dataset and methodology for fine grained vehicle make and model classification. *Frontiers in Computer Science*, vol. 7, p. 1561899, 2025, doi: 10.3389/fcomp.2025.1561899.
- [12] Hanif MS, Nawaz Z, Malik MK. Smart Vision Traffic Surveillance: Vehicle Re-Identification and Tracking Using Vision Transformer, *Vehicles*, vol. 8, no. 2, p. 36, 2026, doi: 10.3390/vehicles8020036.
- [13] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition 2012 Jun 16 (pp. 3354-3361). IEEE. doi: 10.1109/CVPR.2012.6248074.
- [14] Lee D, Lee J, Park E. Automated vehicle damage classification using the three-quarter view car damage dataset and deep learning approaches" *Heliyon*, vol. 10, no. 13, p. e34016, 2024, doi: 10.1016/j.heliyon.2024.e34016.
- [15] Lima GE, Nascimento V, Santos E, Nascimento Jr E, Laroca R, Menotti D. Toward unified fine-grained vehicle classification and automatic license plate recognition. *arXiv preprint arXiv:2604.05271*. 2026 Apr 7.
- [16] K. Zhang and X. Feng, "A vehicle classification method based on deep learning and data augmentation," in *Proc. SPIE*, vol. 12634, 2023, Art. no. 1263408, doi: 10.1117/12.2678655.
- [17] Ismail A, Ismail AR, Shaharuddin NA, Puzi AA, Awang S. Investigation of Convolutional Neural Network Model for Vehicle Classification in Smart City. *International Journal of Advanced Computer Science & Applications*. 2025 Apr 1;16(4):905.
- [18] M. S. Beg, M. Y. Ismail, and M. S. U. Miah, "Evaluating the performance of a visual support system for driving assistance using a deep learning algorithm," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 34, no. 1, pp. 38–50, 2024, doi: 10.37934/araset.34.1.3850.
- [19] A. Zahid, T. Qasim, N. Bhatti, and M. Zia, "A data-driven approach for road accident detection in surveillance videos," *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 17217–17231, 2024, doi: 10.1007/s11042-023-16193-0.
- [20] M. M. Hasan, Z. Wang, M. A. I. Hussain, and K. Fatima, "Bangladeshi native vehicle classification based on transfer learning with deep convolutional neural network," *Sensors*, vol. 21, no. 22, p. 7545, 2021, doi: 10.3390/s21227545.
- [21] Butt MA, Khattak AM, Shafique S, Hayat B, Abid S, Kim KI, Ayub MW, Sajid A, Adnan A. Convolutional neural network based vehicle classification in adverse illuminous conditions for intelligent transportation systems," *Complexity*, vol. 2021, Art. no. 6644861, 2021, doi: 10.1155/2021/6644861.
- [22] D. ul Khairi, F. Ayaz, N. Saeed, K. Ahsan, and S. Z. Ali, "Analysis of deep convolutional neural network models for the fine-grained classification of vehicles," *Future Transp.*, vol. 3, no. 1, pp. 133–149, 2023, doi: 10.3390/futuretransp3010009.
- [23] P. Mani, P. R. G. Komarasamy, N. Rajamanickam, R. Alroobaea, M. Alsafyani, and A. Afandi, "An efficient real-time vehicle classification from a complex image dataset using eXtreme gradient boosting and the multi-objective genetic algorithm," *Processes*, vol. 12, no. 6, p. 1251, 2024, doi: 10.3390/pr12061251.
- [24] Darji K, Ramoliya F, Trivedi C, Gupta R, Kakkar R, Tanwar S, Garg D. Transfer learning and cnn-based vehicle identification approach for hit-and-run cases. In 2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE) 2023 Nov 23 (pp. 898-902). IEEE.
- [25] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015, doi: 10.1109/TITS.2015.2402438.