



DEEP LEARNING NETWORK FOR ROAD IMAGE ANALYSIS WITH TRAFFIC AND ACCIDENT DETECTION

Gopichand G.¹, Sampath Routu² and Ramesh Lalith Kumar²

¹Department of Information Security, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

²SCOPE, Vellore Institute of Technology, Vellore, India

E-Mail: gopichand.g@vit.ac.in

ABSTRACT

A vast quantity of information about vehicular traffic is logged into the monitoring system that monitors traffic each second. It takes a lot of work to manually monitor this data, and it also necessitates hiring staff who are just responsible for monitoring. Deep learning, also known as convolutional neural networks, is a methodology that has the potential to be used for the purpose of controlling and monitoring traffic. After going through some preliminary processing, the data from the various traffic monitoring systems are included in the training Traffic-Net dataset. Therefore, this work implemented the deep learning convolutional neural network (DLCNN) for analysing road images, which can detect the accident, a fire occurred, normal, and dense traffic classes. Initially, the TrafficNet dataset is divided into eighty percent for training and twenty percent for testing. After that, a dataset preparation procedure is carried out to standardise the complete dataset. During the procedure of image pre-processing, all the pictures are scaled down to the same dimensions. In addition, DLCNN is used for the prediction of traffic status. The simulation results showed that the proposed DLCNN resulted in superior traffic analysis performance measurement as compared to other methods.

Keywords: traffic monitoring system, deep learning convolutional neural network, accident, fire occurred, normal, dense traffic.

Manuscript Received 17 August 2023; Revised 21 November 2023; Published 30 December 2023

1. INTRODUCTION

The pace of urbanisation has speeded up, which has resulted in a major rise in the amount of traffic in urban areas. A similar situation has also emerged on motorways that are linked to metropolitan regions [1]. Real-time monitoring of traffic on motorways might offer vehicles with advanced traffic information, allowing them to choose other routes to escape excessive congestion. In addition, comprehensive records of traffic monitoring over extended periods of time will be useful in the process [2] of formulating effective transportation plans and strategies for both urban and suburban locations. At the current moment, the most popular techniques that are used for the goal of monitoring traffic information are closed circuit television (CCTV) or detecting devices [3]. Dedicated short range communication (DSRC), image detectors, and radar detectors, loop detectors are the components that make up the detecting equipment. Most of the time, closed-circuit video cameras are installed in fixed places, and as a result, they are able to maintain a vigilant eye on the region around the highway at all hours of the day and night. Since CCTV [4] can only monitor specific areas, a significant number of CCTV circuits are necessary in order to monitor a diverse network of roads and motorways. Nevertheless, there is a significant cost connected with the installation as well as the continuous maintenance of the many CCTV connections [5]. Because security cameras record highways from an off-angle viewpoint, it is also difficult to automatically identify autos in CCTV video due to the overlap that occurs between individual vehicles. This is a challenge due to the fact that surveillance cameras often record motorways [6].

Recently, video collection approaches that make use of unmanned aerial vehicles (UAVs) have been

deployed to overcome the restrictions that are associated with gathering information about traffic via the use of CCTV [7]. In contrast to CCTV, UAVs can monitor an expansive region of roadways by varying their height or altering their position. Additionally, UAVs are directed to a particular spot to keep an eye on unforeseen events, such as car accidents [8]. In addition, since a UAV looks at the motorways in a direction that is perpendicular to them, the automobiles in the films that are filmed do not overlap. However, people are now responsible for monitoring the footage captured by any installed CCTV or remotely controlled UAVs. As a result, growth in the number of CCTV circuits and UAVs necessitates the employment of more human resources [9]. In addition, we are unable to eliminate the possibility of human mistakes, and it is very challenging to analyse real-time recordings to accurately monitor traffic information.

Rest of the article is organized as follows: Section 2 contains the detailed analysis of existing surveys. Section 3 provided the details of the proposed method. Section 4 contains the details of the results, followed by a comparison. Finally, section 5 concludes the article.

2. LITERATURE SURVEY

Deep learning is an area of machine learning that attempts to simulate the way in which people naturally learn by observing the behaviours of others and modelling their own behaviour after what they've observed. Because of this, the conclusions produced by deep learning are more trustworthy [10]. Deep learning is a type of machine learning that involves providing training to a computer framework for it to be able to directly categorise tasks based on the documents that are available to it in the form



of text, images, or sounds. This type of machine learning is also known as "supervised learning." The convolution neural network is one of the most prevalent types of DLCNN, and it is used for a broad variety of different problems [11]. This network offers an automated approach for feature extraction in contrast to the machine learning methodology, which extracts the features manually by learning the features directly from the photos or the text. In comparison, the machine learning methodology extracts the features manually. These characteristics may be gleaned immediately either from the photos or from the text. Deep learning, also known as DLCNN [12], is a kind of machine learning that, in most cases; uses the neural network in order to carry out correct categorization. This paves the way for deep learning neural networks to have an accuracy that is not only on par with the most recent advancements in the field but also significantly outperforms the work done by people [13]. As a result, the reason for doing this research is to conduct a survey on the deep learning neural network designs that are employed in a range of applications for the goal of having a correct classification along with automated feature extraction. The flow of traffic exhibits varied intensities of temporal patterns, including those that are short-term [14] (daily and weekly) and those that are long-term (annual and multi-annual) (monthly and yearly). The vast bulk of the research that has been done so far on attempting to anticipate the flow of road traffic has focused on short-term trends. There has been a limited amount of investigation on the influence that a variety of long-term trends have on the ability to forecast road traffic flow [15]. It's likely that employing a variety of temporal data segments, in addition to giving additional information about the temporal context, can lead to improved prediction results. This is a point that should be kept in mind. In this research, we studied various magnitudes of temporal patterns, such as short-term and long-term, via the use of diverse temporal data segments to understand how contextual temporal data might enhance prediction [16]. These magnitudes include short-term, medium-term, and long-term patterns. Patterns of these magnitudes are broken down into three categories: short-term, medium-term, and long-term. In addition, in order to make the process of dynamically learning temporal patterns easier, we have designed a one-of-a-kind online framework for dynamic temporal context neural networks [17]. A wide range of temporal data segments are used as input characteristics and are used by the framework. During the process of online education, the updating scheme makes a real-time assessment of how useful a particular segment of temporal data (including both short-term and long-term temporal patterns) is for prediction. It then assigns an appropriate weight to the segment so that it can be incorporated into the regression model [18]. Because of this, the framework is in a position to include not just essential short-term patterns but also key long-term patterns in the regression model, which eventually leads to improved prediction outcomes [19]. We conducted a full experimental assessment utilising an actual dataset that

had daily, weekly, monthly, and yearly data segments. The study comprised all of these types of data. According to the results of the study, both short-term and long-term temporal patterns contributed to an improvement in the level of accuracy of the predictions made [20]. In addition to this, the outputs of prediction were 10.8% more accurate when compared with a deep gated recurrent unit model, and this is all because of the online dynamical framework that was constructed.

For the intelligent transportation system's (ITS) route guidance to be effective in halting the growth of traffic jams, it is necessary to identify regions that experience high volumes of vehicular traffic. Even though the surveillance system has been utilised on the highway for years [21], it is difficult to automatically detect and report traffic congestion in intricate transportation scenes due to the many different types of lighting, weather, and other disturbances. This is the case even though the system has been used. The detection process that depends on the human eye is time-consuming and difficult since the accuracy of machine detection is not high enough to meet the standards of practical applications [22]. DLCNN models were used in this research to construct a unique classifier that produces four TrafficNet based on two championships of ILSVRC, including AlexNet and VGGNet (CNN) [23]. In both AlexNet and VGGNet, the design of the CNN is followed by the implementation of a support vector machine, commonly known as an SVM. This is done rather than utilising layers that are totally coupled to one another [24]. With the assistance of this new framework, lessons and tests is conducted using photographs that include congestion as well as those that do not feature congestion. The current traffic surveillance footage is used to extract an image database that contains more than 30,000 photographs, and then associated labels are added manually [25]. Using video from previously collected traffic surveillance, an image database was developed. These TrafficNet are trained and assessed by utilising the database, and their detection accuracy as well as the amount of time it takes to train are compared. The findings of the experiments show that the accuracy of the suggested method could reach up to 90%, which is a significant improvement over the traditional method, which is based on feature extraction but does not use deep learning. The conventional method is able to achieve a 90% accuracy rate.

3. PROPOSED SYSTEM

Each second, a vast quantity of information about vehicular traffic is logged into the monitoring system that monitors traffic. In addition to being a time-consuming process, monitoring this data with a human eye also requires the hiring of staff members whose only responsibility is to perform monitoring duties. It's possible that the technique of deep learning, which is also referred to as a DLCNN, will be used for the monitoring and control of traffic. The training dataset is built using the data from the traffic surveillance systems after they have been pre-processed.

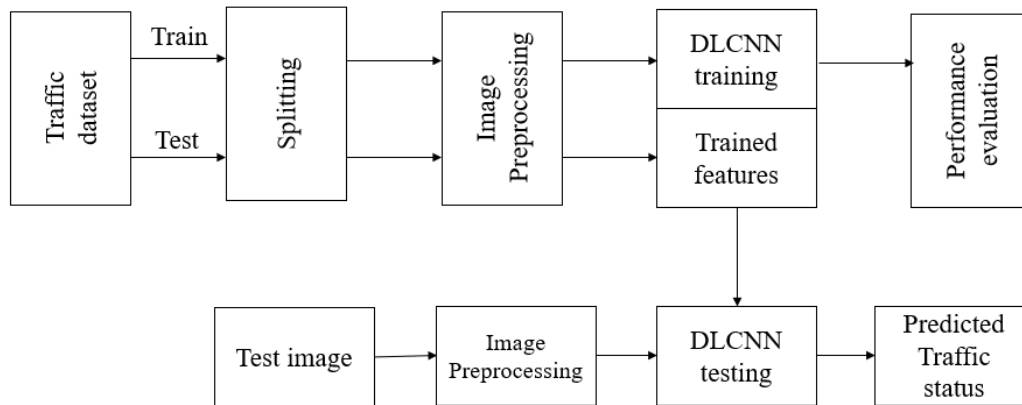


Figure-1. Proposed Method.

The network is then retrained using a self-established data set after being sent to the traffic applications where it is used in the construction of the Traffic net. This Traffic net has the potential to be used in large-scale applications for regional detection. In addition to this, its use is not limited to any one sector. The efficiency is impressively exhibited by the speed of discovery in the high accuracy presented in the case study. It is possible that the preliminary analysis will result in an efficient implementation of the system into a traffic monitoring system and may in the future have potential enrichment for intelligent transportation systems.

The suggested process is shown as a block diagram in Figure-1. At first, the TrafficNet dataset is divided into eighty percent for training and twenty percent for testing. After that, a dataset preparation procedure is carried out in order to standardise the complete dataset. During the procedure of image preprocessing, all of the pictures are scaled down to the same dimensions. In addition, DLCNN is used for the prediction of traffic status, such as high traffic, low traffic, accidents, and fires that have happened based on the test sample. The purpose of the performance assessment is to demonstrate the superiority of the suggested approach.

3.1 Traffic Condition Dataset

The disparate classifications that make up the input dataset are culled from the World Wide Web. The evaluation of the output class is written next to the dataset that was collected. In order to facilitate training and ensure the reliability of the system, 900 photographs have been taken and are now being kept in each of the following four folders: scarce traffic, crowded traffic, fire, and accidents. The name of the folder stands in for the class value when the output is categorised.

3.2 Image Pre-Processing

The term "digital image processing" refers to the practice of applying certain image processing procedures to digital photographs via the use of specific computer algorithms. Analogue image processing is superior to its digital equivalent, digital image processing, in a number of respects, including the fact that digital image processing is

a subset of digital signal processing. It makes it possible to apply a considerably greater range of algorithms to the data that is being entered. The goal of digital image processing is to make the image data (features) better by eliminating undesired distortions and/or increasing certain important picture characteristics. This is accomplished via a combination of the two. Because of this, our artificial intelligence and computer vision models will be able to make greater use of the upgraded data that they are working with. It is necessary for the size of our photographs to match to the size of the network's input in order to successfully train a network and afterwards make predictions based on newly collected data. In the case that we need to adjust the size of the images to make them compatible with the network, we may either rescale or crop the data in order to acquire the necessary size. This will depend on which option will provide the best results.

We can effectively increase the overall amount of training data by using the technique of randomised augmentation to the data. A further benefit of augmentation is that it enables networks to be trained to be invariant to aberrations in the data that they receive from images. For instance, to make a network insensitive to the presence of rotation in input photographs, we might rotate such pictures in a random fashion before feeding them into the network. This would make the network less sensitive to the presence of rotation. An enhanced Image Datastore provides an easy way to apply a limited number of augmentations to two-dimensional photos, which is used for the purpose of resolving classification challenges.

3.3 Proposed DLCNN

DLCNN are increasingly being used in the process of determining traffic conditions. The DLCNN is a type of artificial neural network that is designed to simulate the human brain by employing learnable parameters to stand in for the connections that are made by neurons. The structure of biological neural networks is used as a model for the construction of DLCNN as shown in Figure-2. A convolutional neural network is the name of one of the most common types of structures that can be found in DLCNN. The feed forward neural network category includes this structure as a subset. The



achievements of the deep learning network model provided more support for the relevance of the DLCNN model. Since then, DLCNN has seen significant advancements and has found extensive use in a wide range of situations, including financial supervision, text and

speech recognition, smart home, medical diagnostics, and other fields. Table-1 shows the different layers of DLCNN, which contain the SoftMax, convolution, ReLU, and MaxPooling layers.

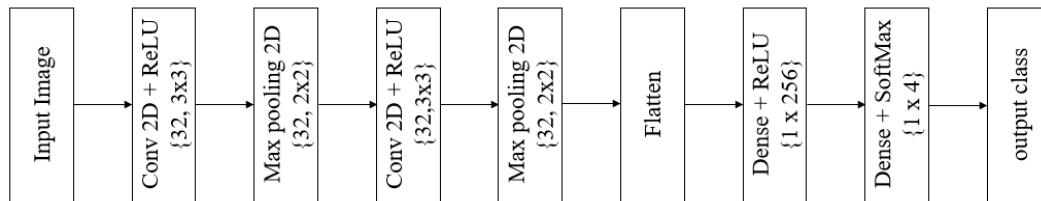


Figure-2. Proposed DLCNN.

Table-1. Layers description.

| Layer Names | No. of filters | Kernel size | Feature size |
|-----------------|----------------|-------------|--------------|
| Conv 2D +ReLU | 32 | 3 x 3 | 62x62x32 |
| Max pooling 2D | - | 3 x 3 | 31x31x32 |
| Conv 2D+ReLU | 32 | 3 x 3 | 29x29x32 |
| Max pooling 2D | - | 3 x 3 | 14x14x32 |
| Flatten | - | 1x6272 | 1x6272 |
| Dense +ReLU | - | 1 x 256 | 1 x 256 |
| Dense + SoftMax | - | 1 x 4 | 1 x 4 |

In most cases, DLCNN are made up of three distinct components. Convolutional layer, used for separating out features. Feature selection is the primary use of the convergence layer, which is often referred to as the pooling layer. By lowering the total number of features, the number of parameters is brought down. The attributes are summarised and output by the entire connection layer, which is responsible for this function. One component of a convolution layer is a convolution process, while the other component is a nonlinear activation function called ReLU. The standard framework of a DLCNN model for crop Figure-2 illustrates the process of traffic condition identification.

The image on the left is known as the input layer, and the computer reads this picture as the data that goes into a number of different matrices. The next layer is the convolution layer, which employs ReLU as its activation function. This layer comes after the previous one. Within the pooling layer, there is not an activation function that can be found. The combination of the convolution and pooling layers is built in several different ways. There are various possibilities. When developing the model, the possible combinations of convolution layers and convolution layers, as well as convolution layers and pool layers, are extremely freely combined and are not restricted in any way. On the other hand, the most frequent form of DLCNN is a mix of numerous layers of convolution and pooling. In the end, but certainly not least, there is a full connection layer that functions as a

classifier and transfers the recently learned feature representation to the sample label space.

The DLCNN is mostly effective in solving the two challenges. The issue that arises from their being an excessive amount of parameters. It is presumed that the size of the picture being submitted is 50 by 50 pixels on a 3-pixel border. When placed in a feedforward network that has all of its links established, there are 7500 connections to the hidden layer that are entirely separate from one another. In addition to this, every link has its very own one-of-a-kind weight parameter that relates to it. When more layers are added to the model, the size of the parameters likewise increases dramatically. This is because each parameter represents a layer in the model. On the one hand, it is quite probable that it will result in the problem of over-fitting happening more often. On the other hand, the neural network is too intricate, which will have a large and detrimental influence on the efficiency of the training. Sharing parameters in DLCNN is a procedure that enables the exact same parameters to be used in many functions of a model. In addition, each individual element that makes up the convolutional kernel will perform its operations on a specific position inside each local input. The neural network just has to learn a collection of parameters; it is not necessary for it to maximise its learning for each individual parameter of each location.

Consistency of the image Picture stability is a local invariant characteristic that shows that the natural image will not be altered regardless of whether the image size is scaled, translated, or rotated. This property is



referred to as image invariance. This is because image stability is the property that is locally invariant. In deep learning, improving the data is often required to increase performance, and fully linked feedforward neural networks are notoriously difficult to train in order to guarantee that a picture will keep its local invariance over time. The convolution process, which is part of a DLCNN, is used to tackle this issue. Finally, the SoftMax classifier identifies the different classes in the dataset.

4. RESULTS AND DISCUSSIONS

This section gives a detailed analysis of simulation results implemented using “python environment”. Further, the performance of the proposed method is compared with existing methods using the same dataset. The disparate classifications that make up the input dataset are culled from the World Wide Web. The evaluation of the output class is written next to the dataset that was collected. To facilitate training and ensure the reliability of the system, 900 photographs have been taken and are now being kept in each of the following four folders: scarce traffic, crowded traffic, fire, and accidents. The name of the folder stands in for the class value when the output is categorised.

Totally, the dataset contains 3600 images, where 80% were used for training and 20% were used for testing. Figure-3 shows the different layers of the proposed DLCNN model, where different layers were presented. Each layer is responsible for filtering photographs of varied dimensions, such as 62 x 62, 31 x 31, and so on. Figure-4 shows the classified outcomes for various test images. Here, the DLCNN classifier predicted low traffic in Figure-4(a), classifier predicted heavy traffic in Figure-4(b), classifier predicted accidents that occurred in Figure-4(c), classifier predicted fire accidents in Figure-4(d).

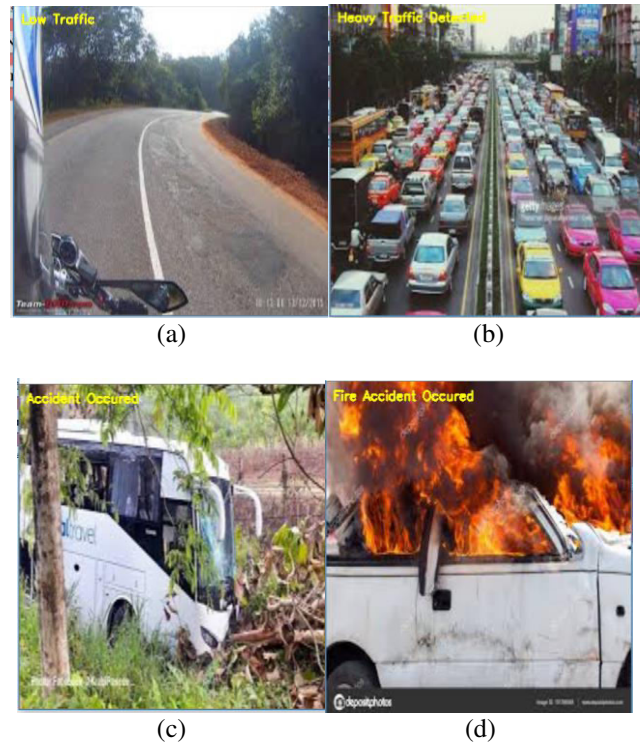


Figure-4. Predicted outcomes, (a) Classifier predicted as low traffic, (b) Classifier predicted as heavy traffic, (c) Classifier predicted as accident occurred, (d) Classifier predicted as fire accident.

Table-2 compares the performance of the proposed method with existing methods. Here, the Proposed CNN resulted in superior accuracy as compared to the existing Naïve bayes (NB) [17] and support vector machine (SVM) [19].

Table-2. Performance comparison.

| Method | NB [17] | SVM [19] | Proposed CNN |
|-----------------|---------|----------|--------------|
| Accuracy values | 89.5 | 90 | 99.3 |

```

Model: "sequential_1"
Layer (type)                Output Shape                Param #
-----
conv2d_1 (Conv2D)           (None, 62, 62, 32)         896
max_pooling2d_1 (MaxPooling2 (None, 31, 31, 32)         0
conv2d_2 (Conv2D)           (None, 29, 29, 32)         9248
max_pooling2d_2 (MaxPooling2 (None, 14, 14, 32)         0
flatten_1 (Flatten)         (None, 6272)                0
dense_1 (Dense)             (None, 256)                 1605888
dense_2 (Dense)             (None, 4)                   1028
-----
Total params: 1,617,060
Trainable params: 1,617,060
Non-trainable params: 0
None
    
```

Figure-3. Layers of DLCNN model.

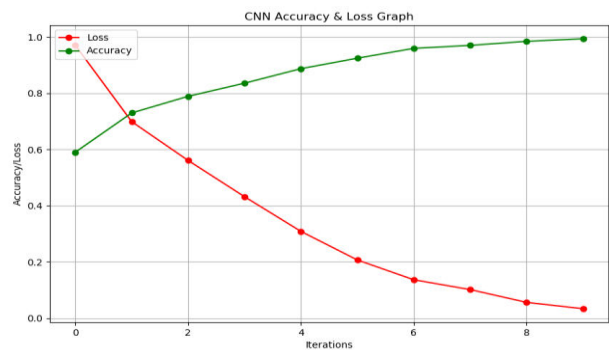


Figure-5. Accuracy and loss graph.

Figure-5 shows the accuracy and loss graph for various iterations. The line depicting loss in the graph is coloured red, whereas the line representing accuracy is



coloured green. The number of iterations, also known as EPOCH, is shown along the x-axis, while the y-axis represents accuracy or loss. The graph on the right shows that with each iteration of the DLCNN model, the accuracy becomes higher, and the loss and error rates get lower. This indicates that the DLCNN model is becoming better with each iteration.

5. CONCLUSIONS

The convolution neural network is a system that eliminates the requirement for human intervention in the process of evaluating the level of traffic congestion on highways. It is anticipated that this will make it possible for deep learning to be used in a range of various roles that are more realistic. The proposed DLCNN will be put through its paces in terms of training and validation, and it will be analysed as a multiclass issue. Finally, this work implemented the DLCNN for analysing the road images, which can detect the accident, a fire occurred, normal, and dense traffic classes. The simulation results show that the proposed model improved performance over conventional methods. Analysing films of traffic in real time will allow the present traffic circumstances to be assessed, which is one possible step toward improving the situation. Using a video splitting methodology that locates each frame and shows the current traffic conditions is one possible method for accomplishing this goal. Real-time traffic identification on video is now a highly relevant academic issue, particularly for emerging nations such as India.

REFERENCES

- [1] Bashar A. 2019. Survey on Evolving Deep Learning Neural Network Architectures. *Journal of Artificial Intelligence*. 1(02): 73-82.
- [2] Zoe Bartlett *et al.* 2019. A Novel Online Dynamic Temporal Context Neural Network Framework for the Prediction of Road Traffic Flow. *IEEE Access*. Vol.7.
- [3] H. Lei *et al.* 2018. A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning. *Pattern Recognit*. 79: 290302.
- [4] P. Wang, L. Li, Y. Jin and G. Wang. 2018. Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNNarchitecture trafficNet. in *Proc. 13th IEEE Conf. Ind. Electron. Appl.*, May/June. pp. 1134-1139.
- [5] T. Pamula. 2018. Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks. *IEEE Intel. Transp. Syst. Mag.* 10(3): 1121.
- [6] Hagler Jr, Donald J., *et al.* 2019. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage*. 202: 116091.
- [7] Xu M., Li C., Zhang S. & Le Callet P. 2020. State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing*. 14(1): 5-26.
- [8] Ghosh Swarnendu, *et al.* 2019. Understanding deep learning techniques for image segmentation. *ACM Computing Surveys (CSUR)*. 52.4: 1-35.
- [9] Chen Junde, *et al.* 2020. Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture*. 173: 105393.
- [10] Jiang Weiwei and Jiayun Luo. 2022. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*. 117921.
- [11] Lavanya Y., *et al.* 2023. Road Accident Detection and Indication System. *ICDSMLA 2021: Proceedings of the 3rd International Conference on Data Science, Machine Learning and Applications*. Singapore: Springer Nature Singapore.
- [12] Basheer Ahmed, Mohammed Imran, *et al.* 2023. A real-time computer vision based approach to detection and classification of traffic incidents. *Big Data and Cognitive Computing*. 7.1: 22.
- [13] Basheer Ahmed, Mohammed Imran, *et al.* 2023. A real-time computer vision based approach to detection and classification of traffic incidents. *Big Data and Cognitive Computing*. 7.1: 22.
- [14] Samadzadegan F., *et al.* 2023. Automatic Road Crack Recognition Based on Deep Learning Networks from Uav Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 10: 685-690.
- [15] Khatri Ankit, *et al.* 2023. Pavement Distress Detection Using Deep Learning Based Methods: A Survey on Role, Challenges and Opportunities. *Computing, Communication and Learning: First International Conference, CoCoLe 2022, Warangal, India, October 27-29, 2022, Proceedings*. Cham: Springer Nature Switzerland.
- [16] Azimjonov Jahongir, Ahmet Özmen and Metin Varan. 2023. A vision-based real-time traffic flow



monitoring system for road intersections. Multimedia tools and applications: 1-20.

- [17] Aljuaydi Fahad, Benchawan Wiwatanapataphee and Yong Hong Wu. 2023. Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. Alexandria engineering journal. 65: 151-162.
- [18] Jain Jinit, *et al.* 2023. Helmet Detection and License Plate Extraction Using Machine Learning and Computer Vision. Cognition and Recognition: 8th International Conference, ICCR 2021, Mandya, India, December 30–31, 2021, Revised Selected Papers. Cham: Springer Nature Switzerland, 2023.
- [19] Qibtiah Raja Mariatul, Zalhan Mohd Zin, and Mohd Fadzil Abu Hassan. 2023. Artificial intelligence system for driver distraction by stacked deep learning classification. Bulletin of Electrical Engineering and Informatics. 12.1: 365-372.
- [20] Medina Juan and Xiaoyue Cathy Liu. 2023. Network Effects of Disruptive Traffic Events.
- [21] Badidi Elarbi, and Dhanya Gopinathan. 2023. On the CPU Usage of Deep Learning Models on an Edge Device. Data Science and Algorithms in Systems: Proceedings of 6th Computational Methods in Systems and Software 2022, Vol. 2. Cham: Springer International Publishing. 209-219.
- [22] Kamath B., Nikhil, *et al.* 2023. TAKEN: Traffic Knowledge-Based Navigation System for Connected and Autonomous Vehicles. Sensors. 23.2: 653.
- [23] Samo Madiha, Jimiama Mosima Mafeni Mase and Graziela Figueredo. 2023. Deep Learning with Attention Mechanisms for Road Weather Detection. Sensors. 23.2: 798.
- [24] Cao Fengyun, *et al.* 2023. Traffic Condition Classification Model Based on Traffic-Net. Computational Intelligence and Neuroscience 2023.
- [25] Song Wei, Guangde Zhang and Yicheng Long. 2023. Identification of dangerous driving state based on lightweight deep learning model. Computers and Electrical Engineering. 105: 108509.