International Journal of Emerging Science and Engineering (IJESE)
ISSN: 2319–6378 (Online), Volume-13 Issue-3, February 2025

Advanced Zebra Crosswalk Detection Using Deep Learning Techniques for Smart Transportation Systems

Md. Muktar Ali, Tariqul Islam

Abstract*: The growing prevalence of road traffic accidents poses significant challenges to vulnerable populations, particularly visually impaired individuals. To enhance safety and accessibility within smart transportation systems, this study presents an advanced detection system for zebra crosswalks, vehicles, and pedestrians utilizing deep learning techniques. The system leverages the Single Shot MultiBox Detector (SSD) model with Transfer Learning for rapid convergence and high accuracy. The dataset, derived from real-world street scenarios, includes nine object classes, enabling the system to provide real-time detection and monitoring. Experimental results demonstrate high precision and recall, underscoring the system's potential to improve road safety and assist traffic management. Future developments will focus on integrating this system into portable devices for broader application.

Keywords*: Single Shot MultiBox Detector (SSD), Traffic Monitoring, Visually Impaired Assistance, Object Detection, Pedestrian Detection.

I. INTRODUCTION

Gradually, the number of automobiles is increasing so fast that we cannot imagine. As a result, the number of road accidents is also rising. However, when it comes to visually impaired or blind people, they take the highest life risk to travel from one place to another. Traffic intersections are among the riskiest places that blind and visually impaired traveler's encounter. They cannot cross any road without another person's help [1]. Sometimes if they try to cross any road on their own, then it gets riskier. Sometimes this decision takes their life. If an accident occurs, we cannot know what happened, how, when, or where. To learn everything takes a hard way and much time. If we monitor them, we could take proper action based on their problem if something happened to them. If some lousy accident occurrence happens, then we can save their lives [2]. We are now working on a zebra crosswalk detection using deep learning, which also can detect vehicles. If someone monitors them, they can get alert, take proper steps, and save their lives. If someone needs to be informed, they can quickly notify them and tell the nearest hospital and ambulance to get some help [3]. Pedestrians are the most important on the road, although it is very dangerous for them to cross any road. For that reason, we saw many people, NGOs, and governments take many steps. However, some visually impaired people can't get any benefits from those steps. Their life is as valuable as us or normal people [4]. Based on WHO (World Health Organization), annually there are more than 270,000 Pedestrian deaths on the road.

Another census confirms that more than 57,877 pedestrians were injured or killed on Philippine roads from 2005 up to 2015. Of this number, 1,859 or 3.2% were killed, while 56,018 or 96.7% were injured [5]. In the current decays independent mobility for the visually impaired has been constantly an active research topic. Road traffic deaths have increased in many countries in recent times while no sign of decrease was seen in low-income countries and only a minor decrease rate in middle and high- income [6]. Zebra crosswalks are painted in black and white alternating stripes that give pedestrians some rights over vehicles while crossing the street. It is more visible to drivers than two-line crosswalks which can prevent serious injuries [7]. Zebra crossing detection is a constantly challenging task in computer vision because of crosswalk type, alternating stripes that are very similar to other structures, viewing angles, occlusion by other objects, diverse and unpredictable lighting conditions, and so on. Although blind or visually impaired people can navigate in outdoor environments with white canes, trained pets, auditory cues from moving vehicles, and other assistance, crossing busy street intersections is much more difficult without any visual signs from these aids [8]. In this paper, we will make a system that can detect pedestrians, crosswalks, and different types of vehicles.

A. Motivation and Objective of this Research

It is tough to travel from one place to another for a blind or impaired person, and when it comes to crossing any road on their own, it is more complex and dangerous for them. For that reason, we will make software that will make roads safer for blind and impaired people. They have the same rights as us, and however, they cannot do ordinary things as we do in our daily life. This study detects any object or person using image processing and learns how to classify any moving object on any highway. We also identified pedestrians, cars, motorcycles, bicycles, rickshaws, trucks, buses, CNG, etc. We also trained our models to identify traffic lights, traffic signs, auto-rickshaws, vans,

Manuscript received on 05 January 2025 | First Revised Manuscript received on 10 January 2025 | Second Revised Manuscript received on 17 January 2025 | Manuscript Accepted on 15 February 2025 | Manuscript published on 28 February 2025.

Correspondence Author(s)
Md. Muktar Ali, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Email: muktar15-2298@diu.edu.bd, ORCID ID: 0000-0002-9367-4340.
Tariqul Islam*, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Email: tariqul15-2250@diu.edu.bd, ORCID ID: 0000-0002-2619-921X.

etc. Our country's growth is developing day by day, our education system is improving continuously, and our safety precautions are developing. We will contribute to the sustained growth of our education system and safety and security by using this zebra crosswalk detection. By using this system, we can secure any blind or visually impaired public safety and security. This study aims to avoid any accidents by using this system and solve the safety issue for blind people.

II. LITERATURE REVIEW

Already there is much research done in this object detection field in Deep Learning. However, very few people worked with crosswalk detection in Deep Learning, and they are using different algorithms in our research. Our paper study thinks no one completed any research using SSD (Single Shot MultiBox Detector) to detect crosswalks. Moreover, some people researched different kinds of object detection by using SSD, such as pedestrian detection, vehicle detection, sign language detection, traffic sign detection, vehicle number plate detection, etc.

Mon Arjay Malbog's tried to analyze the zebra crosswalk from real-time video using 500 data and has only one class in his paper. They are using MASK R-CNN for pedestrian crosswalk detection and instance segmentation. In that research, he only detects zebra crosswalks, nothing more than that. In his project, his main goal was detecting crosswalks for pedestrians auto driving, or smart vehicles [9].

Md. Masud Haider's paper used Faster R-CNN to detect zebra crosswalk region and localization based on Deep Convolutional Neural Network. They only focused on the single class- crosswalk, training the network with images of their dataset combined with extracted image frames. They are also done the same type of work as like previous one. Their goal is close to similar to ours but not the same because of its limited classes [10].

In Mihály Radványi's paper, they used CNN to research Advanced Crosswalk Detection for the Bionic Eyeglass. In their study, they used enhanced color preprocessing through mean shift segmentation to detect crosswalks. They also implement their research in hardware. Their main goal was to detect pedestrian crosswalks based on eyeglasses [11].

In Vedat TÜMEN and Burhan ERGEN journal, they have researched Intersections and Crosswalk Detection Using Deep Learning and Image Processing Techniques by using CNN. In their research, it can detect road separations, intersections, and crosswalks for drivers and autonomous vehicles. Their proposed method is used for driver assistance and a compelling structure [12].

Hojun Son and his team tried to analyze the Crosswalk Guidance System for Blind people. Their project has been researched using CNN, which can detect crosswalks, and implemented their research in smart glass. Their research is to create a completely self-contained detection and guidance system on a commercial, wearable HMD. Their goal is very similar to Mihály Radványi's paper [13].

Matteo et al. paper used GPS images to detect roads and crosswalks based on smartphones. Their main goal is crosswalk detection and localization for visually impaired pedestrians. Their approach is self-localization requires imagery taken by the user to be matched to a template of the traffic intersection. It is based on a computer vision smartphone system developed to provide guidance to blind and visually impaired travelers at traffic intersections [14].

Rânik Guidolini and her team's paper used CNN to research Handling Pedestrians in Crosswalks Using Deep Neural Networks in the IARA Autonomous Car. It only can detect pedestrians and crosswalks for autonomous vehicles. Their main goal is to detect pedestrians and crosswalks to ensure pedestrian safety [15].

Hyok Song and his team's paper used SSD structure and mobilenets for faster processes and higher detection ratio like us [16]. They researched Vulnerable pedestrian detection and tracking using deep learning [17]. In their project, their only approach is to detect pedestrians and cars [18], a tracking and action recognition system using deep learning using video streams that come from CCTVs [19] installed at Seoul National University [20].

A. Comparative Analysis and Summary

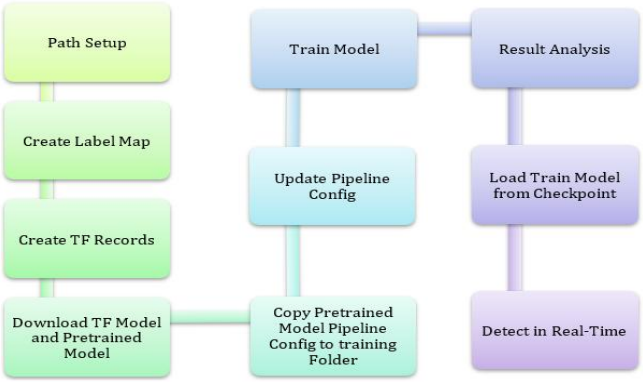
In our study, we proposed to find an easy solution for Zebra Crosswalk Detection using Deep Learning to avoid any accident occurrence for blind or visually impaired people. Our country is developing day by day. However, in our country, many people's economic conditions are not that good. That is why all blind or visually impaired people cannot afford advanced high-priced products to use in their regular lives. We can implement our research on any device that has a camera module and some computing power. This research aims to create a system that can extract features from a picture to identify zebra crosswalks, pedestrians, and different kinds of vehicles. This system needs to be automated so that we can have minimum human interaction in the operation of this system. This development makes any blind or visually impaired person's life easier, safer, and more secure in this modern era. If any blind or visually impaired person uses this system, they can travel to any place without any risk or help.

III. RESEARCH METHODOLOGY

We used Transfer Learning which is based on the Deep Learning model. We are working with SSD or Single Shot MultiBox Detector by using Transfer Learning to complete this research. Transfer learning is an ML (Machine Learning) method where a model is developed for a project, and it is reused as the starting point for a model on a second project. This approach is prevalent in Deep Learning. Pre-trained models are used as the initial point on computer vision and natural language processing tasks given the vast computing and time resources required to develop neural network models on these problems and from the enormous jumps in skill they provide on related issues. In this research, we have used MobileNet V2 FPNLite 320x320 as the pre-trained model. We will take classes from our dataset as we train our model with SSD. To prepare our dataset, we used LabelImg. Using it, we can mark our region of interest to create an XML (Extensible Markup Language) file for that image.

This research project establishes a pedestrian, vehicle, and zebra crosswalk detection based on SSD (Single Shot MultiBox Detector) in Deep

Learning. We are using Transfer Learning based on this architecture. We are only working to detect zebra crosswalks, pedestrians, and different types of vehicles in this research project. We will implement it in a device in the future that gives direction when crossing any road or when not crossing any route to a blind or visually impaired person. The fundamental step to train any model is that we need many image collections in a dataset. Our dataset was extracted from a street video that was taken by our smartphone. The smartphone camera resolution is 12 megapixels, and we shot the video in Mirpur, Dhaka, Bangladesh. We took almost all the images from the sight of pedestrians standing by our smartphones.



[Fig.1. Overall Research Methodology]

A. Data Collection and Preprocessing

We collected our data from a video which is shot by our smartphone. The video footage is taken from Mirpur, Dhaka, Bangladesh. From that video, we extracted the frames as images and selected some of the frames or pictures because all images were not usable for our dataset. This dataset contains nine classes and 63 photos, and they are mainly images of zebra crosswalks, pedestrians, and different types of vehicles. In the future, we will add more classes and images to detect more things accurately.

Our dataset contains many images because our video duration is 20 minutes and has 60 frames per second. We extracted all photos from that video, but some are usable, and some are unusable. Moreover, we cannot annotate/label all the images in a short amount of time, and it will take a very long time to label or annotate all the pictures from that video. We also used the SSD model for training our dataset. In proportion to the number of images we need, our dataset is too small, and we know that very well. That is why we are adding a few pictures to our model frequently. To solve this problem, we used Transfer Learning to get more accuracy.

B. Statistical Analysis

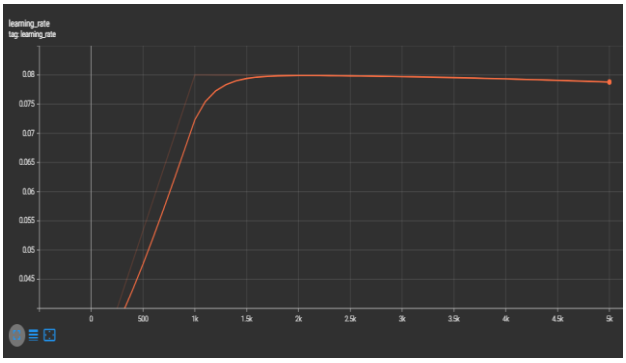
We studied many neural networks and learned so many things. Our study discovered that present-day SSD (Single Shot MultiBox Detection) and YOLO (You Only Look Once) are very popular for their speed and accuracy. Compared to other neural networks, these two have higher speed and accuracy. That is why these two neural networks can detect any object in real-time. After analyzing the difference between SSD and YOLO, we can say SSD is better than YOLO for our research. For that reason, we used SSD MobileNet V2 FPNLite.

C. Model Evaluation

Here we used two models, 1st one is our proposed model which is SSD MobileNet V2 FPNLite 320x320. And second one we used to compare our model which is Faster R-CNN ResNet50 V1 640x640. We trained our dataset based on that pre-trained model. After completing our training, we saw some losses, learning rates, and steps per second based on training. To train 5000 steps with batch 4, it took approximately 3 hours in both cases and then we got,

Table 1: Model Training Losses

Class Name	Value
Classification Loss	0.0565
Localization Loss	0.01632
Regularization Loss	0.1261
Total Loss	0.1989



[Fig.2. Training Learning Rate and Steps per Second]

Based on training losses, learning rate (0.07878), and steps per second (0.7056), we can say that our training is good. After completing our training dataset, we must evaluate our model to see the model result or performance. After evaluating 5000 steps, we can see Detection box precisions, Detection box recalls, and this model's losses. The precision results are given below:

Table 2: Precision/mAP

Name	Value
DetectionBoxes_Precision/mAP	0.6769
DetectionBoxes_Precision/mAP (large)	0.7146
DetectionBoxes_Precision/mAP (medium)	0.3156
DetectionBoxes_Precision/mAP (small)	-1
DetectionBoxes_Precision/mAP@.50IOU	0.8675
DetectionBoxes_Precision/mAP@.75IOU	0.718

Table 3: Recall/AR

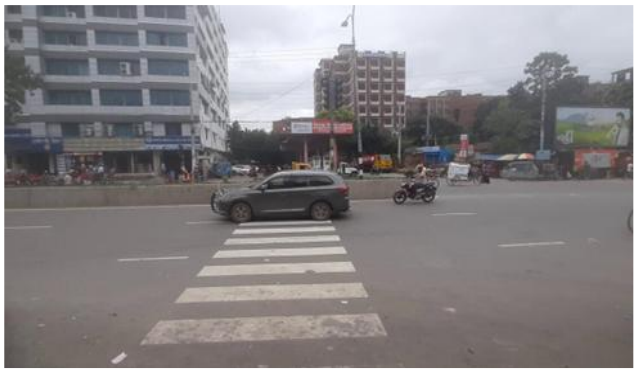
Name	Value
DetectionBoxes_Recall/AR@1	0.6974
DetectionBoxes_Recall/AR@10	0.7575
DetectionBoxes_Recall/AR@100	0.7575
DetectionBoxes_Recall/AR@100 (large)	0.7907
DetectionBoxes_Recall/AR@100 (medium)	0.4063
DetectionBoxes_Recall/AR@100 (small)	-1

After analyzing our model's precision, recall, and losses, we can say that SSD is a good choice for our model. It performs good results although it has a small amount of data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

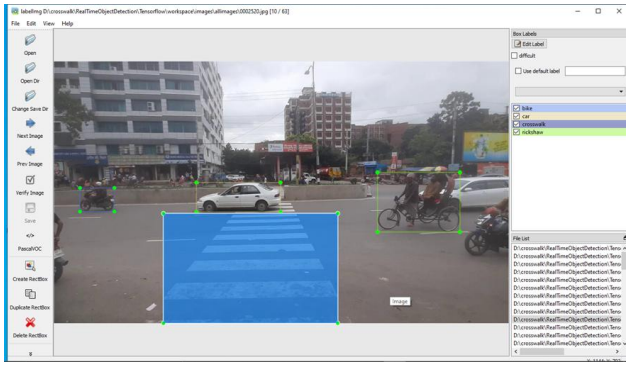
We set up *LabelImg* to annotate or label any image to create a dataset for object detection. After, we collected some images from a video recorded by our smartphone

from the pedestrian's viewpoint. Then we extracted the frames as images and annotated or labeled our images to create a custom dataset. From that dataset, we separated images as 70% and 30%. Then 70% of the image was copied to our training folder and 30% to our testing folder.



[Fig.3. Collected Sample Image]

After code execution, an application opens on our computer, detecting objects from real-time video frames that we want to detect. We started a video from our smartphone for experimental results, and then we saw our detecting results. The results are very satisfying, but sometimes it detects wrong things, and sometimes it cannot detect some object we want. Some experimental results are given below,



[Fig.4. Using LabelImg, Annotating/Labeling Images for a Custom Dataset]

[Fig.5. Crosswalk, CNG, Bike, Rickshaw, Car, Pedestrian, and Bus Detection]

After analyzing the experimental results, our model has good results but has some issues, although our model's dataset is tiny. If we increase the number of images in our dataset, we will get more accurate predictions in our custom model. If we also increase the number of batch sizes and steps, the accuracy and prediction improve. We have already planned; will increase the number of images in our dataset. By adding more data and more training, our model performance will become better.

V. CONCLUSION

This research addresses the critical need for enhancing road safety for visually impaired individuals by developing an advanced detection system for zebra crosswalks, pedestrians, and vehicles using deep learning techniques. The system employs the Single Shot MultiBox Detector (SSD) model, optimized with Transfer Learning, to achieve real-time object detection with high accuracy. Our findings demonstrate the system's potential to significantly improve traffic monitoring, enforce compliance with road rules, and enhance mobility for vulnerable populations. Despite limitations such as a small dataset, the proposed approach shows promising results, paving the way for future developments. These include scaling the dataset, improving accuracy, and implementing the system on portable devices such as Raspberry Pi or smartphone applications. Ultimately, this project contributes to the evolution of smart transportation systems by making roads safer and fostering independent mobility for visually impaired individuals.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests**: Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it has been conducted without any external sway.
- **Ethical Approval and Consent to Participate**: The data provided in this article is exempt from the requirement for ethical approval or participant consent.

▪ **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
▪ **Authors Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Shangguan, Longfei, et al. 'CrossNavi: Enabling Real-Time Crossroad Navigation for the Blind with Commodity Phones'. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Association for Computing Machinery, 2014, pp. 787–98. ACM Digital Library, DOI: https://doi.org/10.1145/2632048.2632083.
2. Li, Xiang, et al. 'Cross-Safe: A Computer Vision-Based Approach to Make All Intersection-Related Pedestrian Signals Accessible for the Visually Impaired'. Advances in Computer Vision, edited by Kohei Arai and Supriya Kapoor, Springer International Publishing, 2020, pp. 132–46. Springer Link, DOI: https://doi.org/10.1007/978-3-030-17798-0_13.
3. Karacs, Kristof, et al. 'Bionic Eyeglass: An Audio Guide for Visually Impaired'. 2006 IEEE Biomedical Circuits and Systems Conference, 2006, pp. 190–93. IEEE Xplore, DOI: https://doi.org/10.1109/BIOCAS.2006.4600340.
4. Trinh, Thong Duc, et al. 'Robust Crosswalk Detection Using Deep Learning Approach'. Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2021, edited by Aboul Ella Hassanien et al., Springer International Publishing, 2022, pp. 62–69. Springer Link, DOI: https://doi.org/10.1007/978-3-030-89701-7_6.
5. Shioyama, Tadayoshi. 'Computer Vision Based Travel Aid for the Blind Crossing Roads'. Proceedings of the 8th International Conference on Advanced Concepts For Intelligent Vision Systems, Springer-Verlag, 2006, pp. 966–77. ACM Digital Library, DOI: https://doi.org/10.1007/11864349_88.
6. Coughlan, J., and R. Manduchi. 'A Mobile Phone Wayfinding System for Visually Impaired Users'. Assistive Technology from Adapted Equipment to Inclusive Environments, IOS Press, 2009, pp. 849–849. ebooks.iospress.nl, DOI: https://doi.org/10.3233/978-1-60750-042-1-849.
7. Browne, Geoffrey R., and Jonathan Flower. 'Zebra Crossings at T-Intersections: Likelihood of Unintended Negative Consequences for Safety and Walkability'. Transportation Research Part F: Traffic Psychology and Behaviour, vol. 95, May 2023, pp. 510–20. ScienceDirect, DOI: https://doi.org/10.1016/j.trf.2023.05.005.
8. Radvanyi, Mihaly, et al. 'Crosswalks Recognition through CNNs for the Bionic Camera: Manual vs. Automatic Design'. 2009 European Conference on Circuit Theory and Design, 2009, pp. 315–18. IEEE Xplore, DOI: https://doi.org/10.1109/ECCTD.2009.5274981.
9. Malbog, Mon Arjay. 'MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation'. 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2019, pp. 1–5. IEEE Xplore, DOI: https://doi.org/10.1109/ICETAS48360.2019.9117217.
10. Haider, Md. Masud, et al. 'Zebra Crosswalk Region Detection and Localization Based on Deep Convolutional Neural Network'. 2019 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON), 2019, pp. 93–97. IEEE Xplore, DOI: https://doi.org/10.1109/RAAICON48939.2019.41.
11. Radványi, Mihály, and Kristóf Karacs. 'Navigation through Crosswalks with the Bionic Eyeglass'. 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010), 2010, pp. 1–2. IEEE Xplore, DOI: https://doi.org/10.1109/ISABEL.2010.5702881.
12. Tümen, Vedat, and Burhan Ergen. 'Intersections and Crosswalk Detection Using Deep Learning and Image Processing Techniques'. Physica A: Statistical Mechanics and Its Applications, vol. 543, Apr. 2020, p. 123510. ScienceDirect, DOI: https://doi.org/10.1016/j.physa.2019.123510.
13. Son, Hojun, et al. 'Crosswalk Guidance System for the Blind'. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 3327–30. IEEE Xplore, DOI: https://doi.org/10.1109/EMBC44109.2020.9176623.
14. Poggi, Matteo, et al. 'Crosswalk Recognition Through Point-Cloud Processing and Deep-Learning Suited to a Wearable Mobility Aid for the Visually Impaired'. New Trends in Image Analysis and Processing ICIAP 2015 Workshops, edited by Vittorio Murino et al., Springer International Publishing, 2015, pp. 282–89. Springer Link, DOI: https://doi.org/10.1007/978-3-319-23222-5_35.
15. Guidolini, Ranik, et al. 'Handling Pedestrians in Crosswalks Using Deep Neural Networks in the IARA Autonomous Car'. 2018 International Joint Conference on Neural Networks (IJCNN), July 2018, pp. 1–8. Semantic Scholar, DOI: https://doi.org/10.1109/IJCNN.2018.8489397.
16. Song, Hyok, et al. 'Vulnerable Pedestrian Detection and Tracking Using Deep Learning'. 2018 International Conference on Electronics, Information, and Communication (ICEIC), 2018, pp. 1–2. IEEE Xplore, DOI: https://doi.org/10.23919/ELINFOCOM.2018.8330547.
17. Sajan, A. Z., & King, G. R. G. (2021). Wireless Speed Control of Vehicles with Detection of Person & Zebra Crossing. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 10, Issue 2, pp. 61–67). DOI: https://doi.org/10.35940/ijrte.b6116.0710221
18. Shobana, G., & Suguna, Dr. M. (2019). Road Traffic Monitoring System. In International Journal of Engineering and Advanced Technology (Vol. 8, Issue 6s3, pp. 1249–1251). DOI: https://doi.org/10.35940/ijeat.f1215.0986s319
19. S., D., T. Sri, G., S. N M Vamsi, K., & SK., M. (2023). Intelligent Traffic Signal Control using RF Technology for Emergency Vehicles. In International Journal of Inventive Engineering and Sciences (Vol. 10, Issue 4, pp. 1–5). DOI: https://doi.org/10.35940/ijies.b3906.0410423
20. Adinarayana, B., & Shafi Mir, Dr. M. (2019). Non-Motorized Transport (NMT) User Safety Guidelines in Urban Areas of Developing Countries. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 2, pp. 4436–4453). DOI: https://doi.org/10.35940/ijitee.b7555.129219

AUTHOR'S PROFILE

Md. Muktar Ali earned his B.Sc. degree in Computer Science and Engineering from Daffodil International University in 2023. He has experience as both a researcher and practitioner at Daffodil International University, where he authored multiple peer-reviewed publications in journals and presented at international conferences. He has actively participated in various international scientific conferences and seminars. His research interests span data science, cybersecurity, big data, cloud computing, computer vision, and image processing.

Tariqul Islam received a B.Sc. degree in Computer Science and Engineering from the Daffodil International University. He is currently a lecturer in the Computer Science and Engineering Department under the Faculty of Science and Information Technology (FSIT) at the Daffodil International University, Dhaka, Bangladesh. He has published 16 research articles in reputed National and International journals and more than 10 research papers in International Conferences. His research interests include Network Security, Information Security, Cryptography, Cyber Security, Wireless Networks, Mobile Networks, Artificial Intelligence, Machine Learning, Deep Learning, Software Security, Satellite Networks, Image Processing, Software Testing, and Neural Networks. He serves as a reviewer for various reputed journals and conferences annually. He was attending different International scientific conferences and many seminars.