# Real-Time Drowsiness Detection System for Student Tracking using Machine Learning

Dilipkumar A. Borikar, Himani Dighorikar, Shridhar Ashtikar, Ishika Bajaj and Shivam Gupta
Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Many studies on fatigue detection have been carried out that were focused on experimention over different technologies. Machine vision based driver fatigue detection systems are used to prevent accidents and improve safety on roads. We propose the design of an alerting system for the students that will use real time video of a person to capture the drowsiness level and will signal alert to the student when the student is in that state of fatigue. A device, if enabled with the system, will start the webcam and track the person. An alert will be generated based on the set frame rate when a continuous set of frames are detected as drowsy. The conventional methods cannot capture complex expressions, however the vailability of deep learning models has enabled a substantial research on detection of states of a person in real time. Our system operates in natural lighting conditions and can predict accurately even when the face is covered with glasses, head caps, etc. The system is implemented using YOLOv5 (You Look Only Once) is an extremely fast and accurate detection model.

Keywords: Drowsiness detection, Deep learning, Convolutional Neural Network, YOLO models,
Online Monitoring, OpenCV.

## 1. INTRODUCTION

Drowsiness is a state of sleepiness that may arise due tovaried reasons such as environmental conditions like rainy seasons, lack of rest, laziness, etc. Post COVID-19 pandemic scenario has seen adoption of online mode of interaction as a de-facto practice these days. During the two-and-a half year long pandemic, the academic activities, especially teaching instructions and evaluations have happened majorly using online medium over the internet. After the pandemic, online mode has taken a new drift. Students are learning online and working for hours in order to complete a task. Longer exposure to online medium might lead to the student falling prey to habit of sudden unconscious sleep which may result in missing out on important deadlines or poor performance due to non-attention. To avoid such states, a bot may be designed that detects the unconscious sleep and alerts the student. This can be useful in improving performance and facilitating students to relax well ahead of the state of drowsiness; in turn helping them to achieve targets. This alert system detect tiredness in real time by observing the eye closure and yawning during the monitoring session and triggers an alarm that intensifies at different intervals until the student takes corrective action and comes out of the drowsiness. A neural network forms the core of any deep learning algorithm. The artificial neural networks, convolutional neural networks (CNN) and recurrent neural networks (RNN) have been used to design solution using deep learning approaches. The CNNs have been found well-suited for extracting patterns in the images. The "You Look Only Once" (YOLO) models are CNN models that are known for their fast and accurate results. In this paper, YOLO models are used to find the patterns in images to detect and classify them as drowsy or awake. Our work uses YOLOv5, a faster and more accurate model to get the best predictions. Other versions of YOLOv5 model have also been experimented. The prediction relies on analysis of set of consecutive frames before finalizing the state as drowsy.

## 2. LITERATURE REVIEW

Researches have been directed towards incorporating CNN models and other approaches using image processing technologies towards benefitting the mankind. Helping people, especially drivers in avoiding accidents due to drowsiness and helping students to get timely alert before losing connect with the context of instruction delivery during online mode are among the few areas explored extensively.

Manikandan V. M., et al. [2021] has introduced the behavioural-based approach that analyses the frontal features. The authors have implemented the Viola-Jones algorithm to detect frontal faces from the image frames. The image frames were pre-processed by using RGB-to-Grey transformations and their division into set of grids. The support vector machine (SVM) classifier is used to track the movement of the eyes. Similarly, the head movement is tracked by continually computing the Euclidean distance against the defined threshold.

Navya Kiran V. B., et al. [2020] has used PERCLOS drowsiness metric to detect driver's drowsiness level. PERCLOS is the percentage of eyelid closure over the pupil for a duration and is sensitive to slow eyelid closures than normal eye blinks. In context of the vehicular drowsiness detection systems, the PERCLOS is one of the promising real-time measures of alertness. The approach considered parameters such as face perception, eye position, state of eyes and computes eye trail and face trail alongwith percentage closure of eyelids to confirm a drowsy state. They have used a mean shift method incorporating Cam-Shift algorithm and Viola Jones Algorithm that uses AdaBoost filtering coupled with HAAR cascades while processing the image frames.

Elena Magan, et al. [2022] has proposed two phased approach for detecting fatigue. In the first phase deep learning is used to analyse a sequence of images of the driver. In the second phase a combination of AI and deep learning techniques is used to extract the important features from the image. The results and features from both phases are introduced to a fuzzy inference system that evaluates whether the driver is drowsy or not. Umang Lahoti, et al. [2020] implemented the drowsiness detection system which uses Ear Aspect Ratio and Mouth Aspect Ratio in their implementation that carries validation using the ResNet50 pre-trained model. Image-based methods are used to capture the eyes using Haar based cascade classifier and blink detection is done using histogram of oriented gradient (HOG) based features along with SVM classifiers. The classification is done based on the percentage of eye closure (PERCLOS). Dweepna Garg, et al. [2018] has used YOLO models for better accuracy in face detection process. Their study compares the accuracy of detecting against the traditional approaches and significant efficiency in the same has been observed. Belmekki G. A., et al. [2020] proposed PerStat method that determines the most frequent state. The acquired states are fed to a multi-layer perceptron in estimating the driver state. The Elman RNN receives the output generated from the MLP and produces a sequential feedback which links the states with time dimension thereby establishing the state evolution trail.

Feng You, et al. [2020] experimented on the WIDER FACE data set and has used YOLOv3 models to detect the face of the driver. To obtain the face features they have used Dlib toolkit with YOLO. Sukrit Mehta, et al. [2019] have preferred non-intrusive methods to prevent distraction to the driver that is caused due to fixing sensors on the drivers body. Instead measures like an eye aspect ratio (EAR) and the eye closure ratio (ECR) with adaptive threshold is used to discover the state of drowsiness or fatigue. Their system gives 84% accurate predictions with random forest classifiers. A neuro-fuzzy approach incorporating particle swarm optimization (PSO) and SVM has been proposed for detecting drowsiness inside the vehicle ( S. Arefnezhad, et al. [2019] and K. Mutya et al. [2019]). The system captures the steering wheel images of the driver. An approach using CNN algorithms over steering wheel images capture dynamically is seen to be sensitive to false drowsy detection rates ( K. Mutya et al. [2019]). R. Jabbar, et al. [2020] proposed a CNN approach with ML algorithm to detect micro sleep and drowsiness which acquires images using an overhead camera and computes the facial features.

Jamuna Murthy et al. [2022] has proposed a framework that provides real-time object detection with optimal speed and assists the driver with accuracy. The framework uses the state-of-the-art YOLOv5 algorithm in implementing the three major phases, namely, extraction, detection, and visualization. Redmon et al. [2016] have used YOLO — A one-time convolutional neural network for the prediction of the frame position and classification of multiple candidates. It achieves an end-to-end target detection. It uses a regression for object detection. A single end-to-end system constitutes an ensemble of modules that generates classification from the input images. The extensions to use of bounding box prediction and the YOLO based approach to feature extraction has significant impact on our approach.

## 3. METHODOLOGY

### 3.1 Dataset

For the proposed approach, the data of more than 3000 subjects is collected which includes Closed Eyes in the Wild (CEW) face images, DSM (Driver status monitor) HD videos where frames are generated at random intervals from the videos, human faces including different expressions are taken out from different sources on Google and some videos are recorded by the team members capturing different gestures with different cases like dim light, bright light, different angles, etc. to make the model learn as much gestures and situations as possible. Under drowsy state, yawning and eye closeness is considered (Iason Katsamenis, et al. [2022]).

The states include whether the person is drowsy or awake if the person is present, otherwise the person's absense. Two states are included in the dataset: drowsy and awake. The data collected is then labelled with either drowsy or awakened states. Figure 1 describes the process of custom dataset generation. There are a total of 6066 images of which 2734 are in drowsy state and 3332 are awaken state. The dataset is divided into three parts: training, validation and testing in the ratio of 7:3:1. The training set has 4k images, the validation set has 1.2k images and the testing part has 606 images.

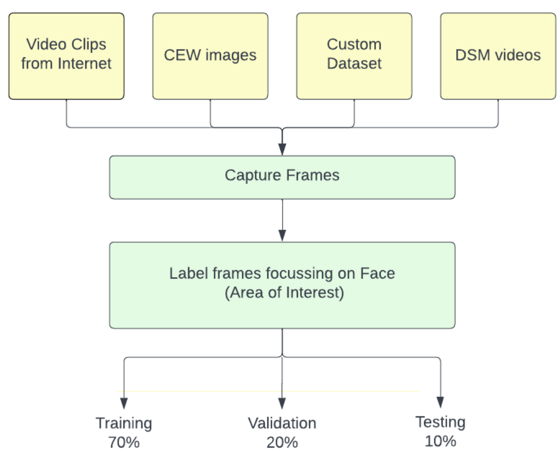


Figure 1. Dataset collection and distribution

The images are labelled using software that focus on the area of interest (capturing face area) as shown in Figure 2. In the training phase the model will process the face region and extract features like closed eyes, normal state, yawning, and others.

### 3.2 Face Detection and Recognition based in YOLOv5 Models

The YOLOv5 architecture, which is a single stage object detector, is used to extract important features from the input images using regions of interest (ROI). It requires labelled images for training on the features. Therefore, the custom dataset is annotated with labels accordingly, focussing on the face region to achieve the weights. It works on a single face for detecting the

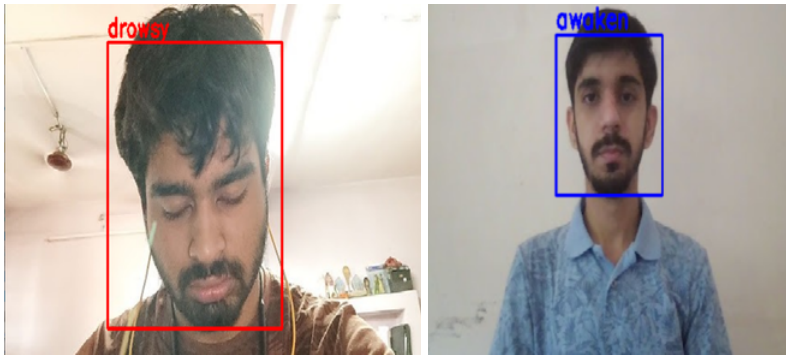


Figure 2. Data Labelling

states whether drowsy or awakened. The frames are extracted from the videos and are given to the framework.

The proposed architecture is a fusion of Darknet and Cross Stage partial network (CSP). The grids do the feature extraction and target information annotation from the input images. During the face identification stage the image is divided into N x N grids. If the target image centre lies in the grid, that grid is accountable for the identification. The confidence that the grid contains the target is calculated as:

$$C_{ij} = P_{ij} * IOUScore \tag{1}$$

$C_{ij}$ denotes the confidence score that the $j^{th}$ bounding box is present in the $i^{th}$ grid. $P_{ij}$ is the probability that there exists a target in the grid, intersection over union (IOU) between the true and the estimated box. The IOU score must be high for the confidence to be more precise (Akhil Kumar, et al. [2020]). Figure 3 illustrates a generic approach to drowsiness detection system.

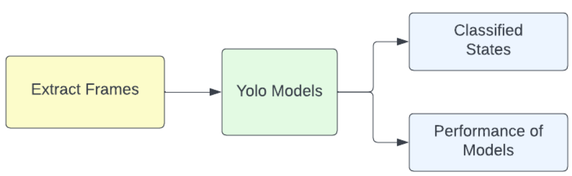


Figure 3. Proposed methodology

The system specifications used for training the model are mentioned in Table I.

| **Parameter** | **Specification** |
|---|---|
| Operating system | Windows 10 Pro |
| CPU | Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz |
| RAM | 16 GB (usable) |
| GPU | NVIDIA@, CUDA@ |
| Frameworks | PyTorch, OpenCV |

Table I: System Specification

### 3.3 Drowsiness Alert

Figure 4 outlines the system flow of the process of drowsiness detection. The process starts from the data collection of various images showing different expressions mainly at drowsy state and normal state. Our model detects the faces and classifies them into two classes: drowsy and awakened. The recognition is based on eye status and yawning state. It involves determining the number of consecutive frames. Based on the states recognized in consecutive frames the alert will be given that will be progressive in nature i.e., increasing alarm intensity if not responded on time. The person in front of the webcam will be continuously monitored. The real time frames will be generated using OpenCV. The trained model will identify the state of each frame and recognize the drowsy state. It will check for 20-50 consecutive frames. If all consecutive frames are drowsy, then an alert will be generated. However, the system continues to monitor the person and gives progressive alarm if not responded on time.

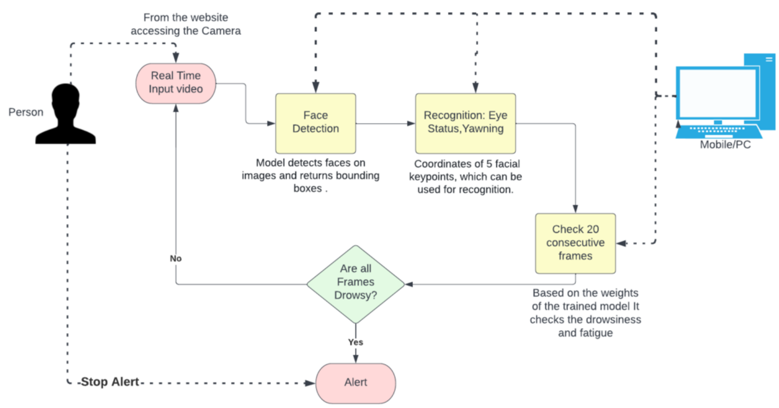


Figure 4. System Flow for Drowsiness Alert

The flow of the utility for the drowsiness detection system is described in Figure 5. When the detection is enabled, it will ask for the webcam permission. Automatically, it starts capturing images using OpenCV and a trained model will detect the real time images. Rate of frame creation is 30 fps. Hence, when contigous 30-60 frames detected as a drowsy, the system will invoke the alarm. The alarm would last a couple of seconds and again the detection is monitored. If the drowsy state is continued the alarm intensifies its effect and turns on.

## 4. RESULTS AND DISCUSSION

The datasets contain images of varied scenarios like talking, screaming, gazing, etc. and cases like drowsy, opening and closing of eyes, yawning, nodding are taken into account in order to train the system for all possible instances. The resolution of frames was different for different clips. However, for implementation and pre-processing the images captured in real time were transformed into image frames with 640 x 480 resolution. The faces in the dataset have significant transformation and perceptions. Hence the dataset is appropriate to the performance in practical scenarios.

As the proposed model is expected to work on real time capturing using webcam, the custom videos were taken using the webcams with the resolution of 1280 x 640. The model is trained on different versions of the YOLOv5 model. The models are trained and various measures are observed and compared as shown in Table II.

The YOLOv5 model is implemented as core architecture with different configurations to make the model learn desired states. The model is trained for 100 epochs. Figure 6 shows the precision, the recall and the mean average precision (mAP) at 0.5 IOU when the model is trained at 100

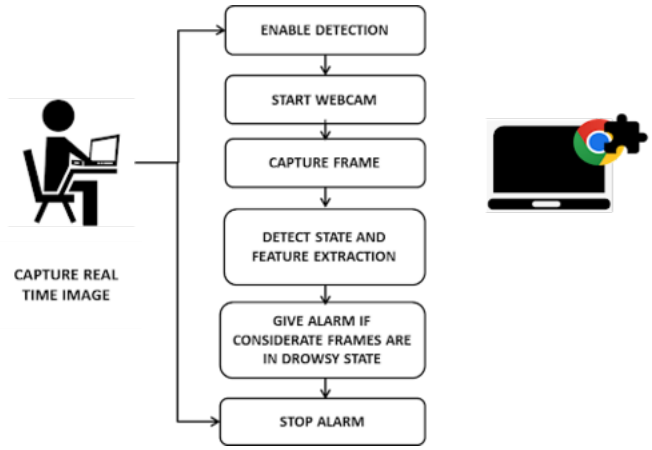


Figure 5. Working of real-time detection system

| Sr. no. | Model | Precisionl | Recall | mAP |
|---|---|---|---|---|
| 1 | YOLO v5s | 0.951 | 0.945 | 0.960 |
| 2 | YOLO v5m | 0.941 | 0.957 | 0.975 |
| 3 | YOLO v5x | 0.944 | 0.948 | 0.979 |

Table II: Performance of YOLO Models on Custom Drowsiness Dataset

epochs. The confusion matrix for the YOLOv5 $x$ model is depicted by Figure 7. The model was found to yield faster detections on real images for the frame rate of 30 fps.

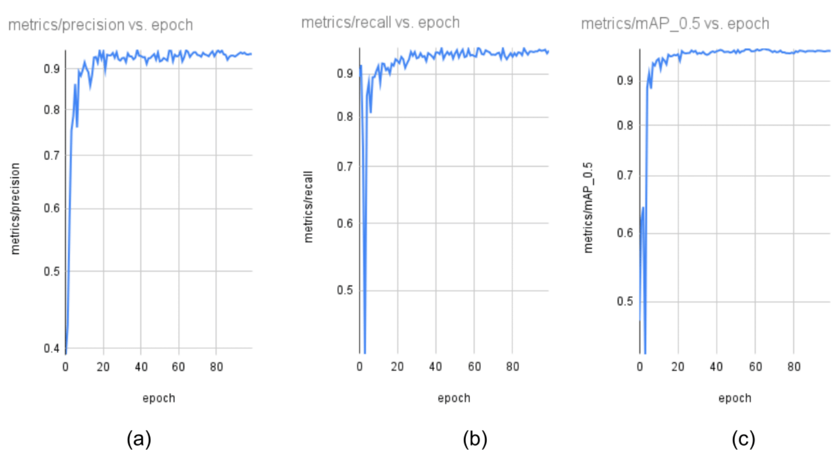


Figure 6. Model Performance for 100 epochs (a) precision (b) recall (c) mAP

Figure 8 shows the real time webcam tests showing the state and the confidence achieved by the model on a live video frame.

## 5. CONCLUSION

In this paper, we have proposed a drowsiness detection system that detects whether the student is in a drowsy state or awaken state. Additionally, depending on the confidence level it alerts the user. The system uses YOLOv5 pre-trained CNN architecture for efficient detection and

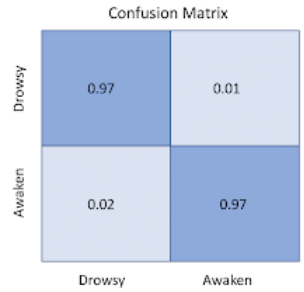


Figure 7. YOLOv5 $x$ Confusion Matrix

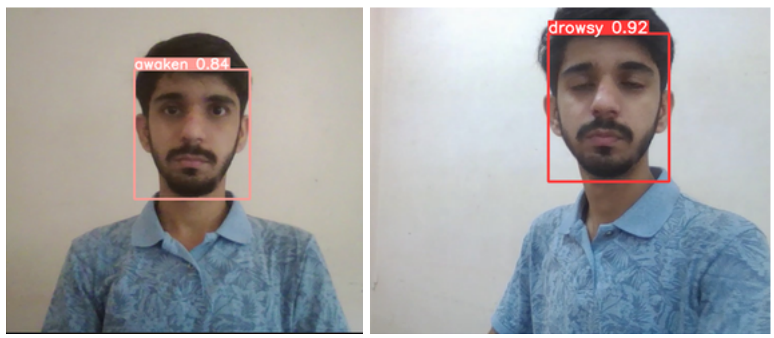


Figure 8. Class Predictions

recognition. Test runs were conducted and analyses were carried out using YOLO models with $s$, $m$ and $x$ versions. The mAP scores of 0.98, 0.975 and 0.979 were obtained for the respective models. However, the accuracy when tested on YOLOv5 $x$ is observed to be higher and therefore was considered for further experimentation. The precision and recall score were observed to be 0.944 and 0.948 respectively. The trained model can be integrated in the extension to ease the utilization of the system. It was noticed that at times there may exist identical expressions with different emotions. The impact created by such expressions is based on usual human behavior. For example, screaming loud and yawing are two different gestures, but the drowsiness detection system might classify both these as drowsy. Future improvements call for proper detection in situations as discussed.

## References

Arefnezhad, S., Samiee, S., Eichberger, A., and Nahvi, A. 2019. Driver Drowsiness Detection based on Steering Wheel Data Applying Adaptive Neuro-Fuzzy Feature Selection. In: *Sensors, Vol. 19, No. 4,* DOI: https://doi.org/10.3390/s19040943.

Belmekki, G. A., Mekkakia, M. Z., and Pomares, H. 2020. Driver Drowsiness Detection and Tracking based on YOLO with Haar Ccascades and ERNN. *International Journal of Safety and Security Engineering, Volume 11, No. 1*, pp. 35-42. DOI: https://doi.org/10.18280/ijsse.110104.

Garg, D., Goel, P., Pandya, S., Ganatra, A., and Kotecha, K. 2018. A Deep Learning Approach for Face Detection using YOLO. *IEEE PuneCon* , pp. 1-4, DOI: 10.1109/PUNECON.2018.8745376.

Jabbar, R., Shinoy, M., Kharbeche, M., Al-Khalifa, K., Krichen, M., and Barkaoui, K. 2020. Driver Drowsiness Detection Model using Convolutional Neural Networks Techniques for Android Application. In: *Proceedings of 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) IEEE.*

Katsamenis, I., Karolou, E. E., Davradou, A., Protopapadakis, E., Doulamis, N., Doulamis, A., and Kalogeras, D. 2022. TraCon: A Novel Dataset for Real-Time Traffic Cones Detection using Deep Learning, In: *Krouska, A., Troussas, C., Caro, J. (eds) Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022). NiDS 2022. Lecture Notes in Networks and Systems, vol 556. Springer, Cham.* DOI: https://doi.org/10.1007/978-3-031-17601-2_37.

Kumar, A., Kalia, A., and Sharma, A. 2020. Object Detection: A Comprehensive Review of the State-of-the-Art Methods. *International Journal of Next-Generation Computing, Vol. 11, No. 1.*

Lahoti, U., Joshi, R., Vyas, N., Deshpande, K., and Jain, S. 2020. Drowsiness Detection System for Online Courses. *International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, Issue 2.*

Magan, E., Sesmero, M. P., Alonso-Weber, J. M., and Sanchis, A. 2022. Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images. *Applied Sciences 2022, Volume 12, Issue 3,* pp. 1145, DOI: https://doi.org/10.3390/app12031145.

Mehta, S., Dadhich, S., Gumber, S., and Jadhav Bhatt, A. 2019. Real-Time Driver Drowsiness Detection System Using Eye Aspect Ratio and Eye Closure Ratio. *International Conference on Sustainable Computing in Science Technology and Management, Elsevier SSRN.* .

Murthy, J. S., Siddesh, G. M., Lai, W. C., Parameshachari, B. D., Patil, S. N., and Hemalatha, K. L. 2022. ObjectDetect: A Real-Time Object Detection Framework for Advanced Driver Assistant Systems Using YOLOv5. *Wireless Communications and Mobile Computing, Volume 2022,* DOI: https://doi.org/10.1155/2022/9444360.

Murthy, S. K., Siddineni, B., Kompella, V. K., Kondaveeti, A., Boddupalli, H., and Manikandan V. M. 2021. An Efficient Drowsiness Detection Scheme using Video Analysis. *International Journal of Computing and Digital Systems, Volume 11, Issue 1*, DOI: https://dx.doi.org/10.12785/ijcds/110146.

Mutya, K., Shah, J., McDonald, A. D., and Jefferson, J. 2019. What Are Steering Pictures Are Worth? Using Image-based Steering Features to Detect Drowsiness on Rural Roads. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 63, No. 1,* DOI: https://doi.org/10.1177/1071181319631220.

Navya Kiran, V. B., Raksha, R., Rahman, A., Varsha K. N., and Nagamani, N. P. 2020. Driver Drowsiness Detection. *International Journal of Engineering Research and Technology (IJERT), NCAIT-2020, Volume 8, Issue 15.*

Redmon J., Divvala S., and Farhadi A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 779-788, DOI: 10.1109/CVPR.2016.91.

You, F. , Gong, Y., Tu, H., Liang, J., and Wang, H. 2020. A Fatigue Driving Detection Algorithm Based on Facial Motion Information Entropy. *Journal of Advanced Transportation, Volume 2020*, DOI: https://doi.org/10.1155/2020/8851485.

**Prof Dilipkumar A Borikar** is an Assistant Professor in the Department of Computer Science and Engineering at Shri Ramdeoebaba College of Engineering and Management (RCOEM), Nagpur since 2014. Prior to joining RCOEM he worked as Head, Department of Information Technology at KITS Ramtek, Nagpur. He earned degrees in Bachelor of Engineering (B.E.) and Master of Business Administration (M.B.A.) from Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur. He has completed his M. Tech.(IT) from the School of Information Technology at Indian Institute of Technology, Kharagpur. Prof. D. A. Borikar is a recipient of the prestigious Institute Silver Medal of IIT Kharagpur. He persues research interests in informaion processing, NLP, image and video processing and data analytics. He is a life member of ISTE and has 20+ publications in refereed journals and international conferences.
E-mail: borikarda@rknec.edu

**Himani Manoj Dighorikar** is pursuing her undergraduate engineering programme (B.E. CSE) in the Department of Computer Science and Engineeering of Shri Ramdeoebaba College of Engineering and Management (RCOEM), Nagpur. She is currently in her final year. She has been working on her interest areas which includes Natural Language processing(NLP), Data Analysis, and Deep Learning. Himani has a publication in a peer reviewed refereed journal.
E-mail: dighorikarhm@rknec.edu

**Shridhar Ashtikar** is pursuing his undergraduate engineering programme (B.E. CSE) in the Department of Computer Science and Engineeering of Shri Ramdeoebaba College of Engineering and Management (RCOEM), Nagpur. He is currently in his final year. He has been working on image and video analysis, and machine learning. Shridhar has a publication in a peer reviewed refereed journal on NLP. His research interest areas include NLP, Image Analysis and Data Science.
E-mail: ashtikarsa@rknec.edu

**Ishika Bajaj** is pursuing her undergraduate engineering programme (B.E. CSE) in the Department of Computer Science and Engineeering of Shri Ramdeoebaba College of Engineering and Management (RCOEM), Nagpur. She is currently in her final year. She has been working on image and video analysis, natural language processing and machine learning. Ishika has a publication in a peer reviewed refereed journal.
E-mail: bajajij@rknec.edu

**Shivam Gupta** is pursuing her undergraduate engineering programme (B.E. CSE) in the Department of Computer Science and Engineeering of Shri Ramdeoebaba College of Engineering and Management (RCOEM), Nagpur. He is currently in his final year. He has been working on image and video analysis, and machine learning. Shivam has a publication in a peer reviewed refereed journal.
E-mail: guptasb_1@rknec.edu