Driver Alert System

Aditya Khursale Bhushan Mahajan Anurag Pawar
University of Maryland Baltimore County
CMSC 691 Final Report
Spring 2022 Group 11

Abstract-

Distracted driving causes about 920,000 total accidents and approximately 3,000 deaths per year (Coleman, 2022). Driving a vehicle requires the driver's undivided attention. Distracted driving includes any activity that takes your attention away from the road when you are behind the wheel (Coleman, 2022). Another cause of severe road accidents is weariness, in which the driver unknowingly loses his concentration due to sleep deprivation. Modern vehicles on the other hand, are growing smarter to integrate modern technologies and algorithms from Computer Vision, Machine Learning, and Artificial Intelligence to reduce these mishaps and ensure passenger safety. In this project, we are combining Computer Vision and Deep Learning techniques to monitor the driver, identify distraction, and generate voice alarms. The proposed system incorporates several monitoring approaches like eye blinking rate, eye gaze tracking, head tilt, mouth openness for yawning, driver activity monitoring in pre-defined classes, and recognizing emotion of the driver from facial expression to track cognitive distractions. The system proposed is unique in terms of fusing and supporting multiple features together to improve the capability of contemporary co-pilot systems.

1. INTRODUCTION

Modern technology, such as cell phones and interactive car dashboards, have increased driver distractions, and there has been an alarming increase in road accidents. As a result, in-car perception is a hot issue for study and development these days. We investigated some of the topics in this domain and proposed a complete system that includes the most important applications from this study. This project is effective and unique since it includes warnings based on gaze tracking, tiredness detection, eye openness, yawn detection, head pose estimate, activity monitor, and emotion identification in a single system.

The proposed system provides a wide range of real-world application and scalability. Beyond a semester, the project might be followed as a thorough study topic or even as a business endeavor. To mention a few possible extensions, the project may be developed for smart seat adjustment based on recorded faces or for commercial infotainment control via gestures. This system may also be investigated further for problems such as maintaining / increasing the system's functionality whenever cameras adjust themselves with auto exposure whenever the car enters a tunnel, or how to validate the capabilities of the system in an automated fashion. As a result, the project has a lot of room for growth.

We think that by researching and executing this project, we will gain hands-on experience with a variety of algorithms in the field of computer vision, which will allow us to conduct more research, develop the proposed system, and explore fresh initiatives in the domain.

2. RELATED WORK

We went through various research works to decide the scope of the problem, determine existing bottlenecks of the driver monitoring systems, and decide on what and how should we implement the system. Deng and Wu (2019) proposed a system called DriCare which uses a commercial camera vehicle device for capturing the driver and sending the captured images to a cloud server for further processing and finally receives a response from the server if the driver is drowsy or not. In another work, Dua et al. (2018) mounted a mobile phone on the windshield of the driver and used approximately 30 facial landmarks, and a ratio-based technique to detect the drowsiness of the driver. Whereas Devi and Bajaj (2008) used intensity shifts and the ratio between forehead and eyelids. We think, sending images to the cloud is not scalable for a real time system and brings additional failure points like network communication, latencies, etc. to the system. Hence, we plan to perform all computation at the edge systems, additionally, we employ 68 facial landmarks, and track only eyelid ratio along with the combination of head tilt, and mouth ratio for the drowsiness detection part.

For emotion recognition, we studied Derisknet, a driver monitoring system based on a deep convolutional neural network that identifies seven emotion states of the driver, and crawls the web to import and play sounds to change the mood of a driver. (Wu et al., 2018) We are employing a similar technique however we are using a different dataset, also we don't think crawling the web in a real time system is a good idea. However, we don't plan to play music to enlighten the mood in the current deliverable of the project. We think the best way to do this is to maintain a static database, also to add dynamicity, the web crawling can be implemented as a scheduled job every night or new music can be pushed as over-the-air updates onto the system. But all of this is out of the scope for the current term project.

We have studied Ahlstrom et al. (2013) paper to narrow down the effectiveness of gaze tracking in distracted driver detection. The paper demonstrated the effectiveness of the system by conducting an extensive field investigation of seven drivers who traveled an average of 4351 km in a realistic setting. The study showed that this system helped to reduce visual time-sharing for glances from 9.94 to 9.20 seconds. Furthermore, we studied the research of driver activity recognition, where authors trained the behavioral recognition CNN model and presented a comparative analysis against the standard model. Xing et al. (2019) Our architecture for the activity classifier is inspired by this paper, however, we plan to train a model from scratch on a Distracted Driver Dataset. Abouelnaga et al. (2017).

3. METHODOLOGY

The project is divided in to three primary blocks: Sense, Detection/Algorithms, and Visualization. Figure 1 is the flow diagram between these blocks, and each of this block is described in detail below.

3.1. SENSE

The driver is essentially captured by the Sense module. Commercial infotainment or in-car perception systems typically include one to four cameras. For this project and the features proposed in this scope, the system requires only one camera configuration, which is directly focused on the driver's face. The front camera will be utilized largely to detect the driver's face and facial traits,

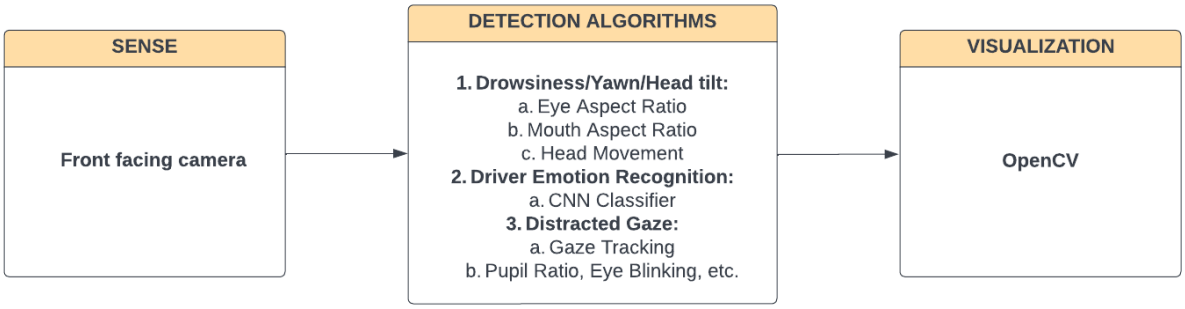


Figure 1: Driver Alert System

such as eye blinking, gaze tracking, mouth opening, head movement, and so on. For the demonstration purpose, we have used a 720p FaceTime HD Camera which is shipped with Macbooks. We are using video related APIs from the imutils package of python to grab and process the frames captured by web camera. Once the frames are captured basic preprocessing like converting frames into gray or cropping the non-facial area of the image etc. is done within this block using the OpenCV APIs.

3.2. DETECTION/ALGORITHMS

This is the system's heart, processing sensory data (camera frames recorded) and predicting whether or not the driver is paying attention.

To achieve the aim and alert the distracted driver, the system first recognizes the driver's face using the Histogram of Oriented Gradients and Linear Support Vector Machines. In this method, we sample positive and negative training sets where positive samples corresponds to faces whereas negative samples corresponds to non-faces, and then extract HOG descriptors for both of these. Based on this data linear SVM is trained, and hard-negative mining is applied. Hard-negative mining is computing hog descriptors and applying a classifier on each image and each possible scale of the image in a sliding window fashion. After hard-negative mining, false positives are ordered based on their confidence and retrained the classifier using the hard-negative samples. This classifier is finally used to recognize the face from the captured image.

Once the faces are detected, we have used a pre-trained model which trained on the iBUG-300W dataset Sagonas et al. (2016, 2013a,b). This model identifies key shapes from the face like eyes, nose, mouth, etc., and finds 68 facial landmarks for each of the faces found in the frame. Visualization of these 68 facial landmarks is shown in Figure 2. We have used these landmarks to detect weariness and inattentive driving using the methods outlined below.

To assess the driver's tiredness, we have integrated three approaches: Measurement of the head tilt angle, Eye openness, and Mouth openness. To detect the openness of eyes and mouth we relied on simple aspect ratios. These are just the horizontal and vertical ratios of the euclidean distance between the facial landmarks identified. E.g. to detect whether the left eye is open or not, we calculate the horizontal ratio between landmark numbers 37 to 40 and further ratio it with vertical landmarks of the left eye i.e., 38, 39, 41, and 42. Based on some pre-set thresholds for these ratios we conclude if the eyes or mouth is open or not. To add more robustness to the alerts, we further monitor consecutive frames and put an additional threshold on the number of

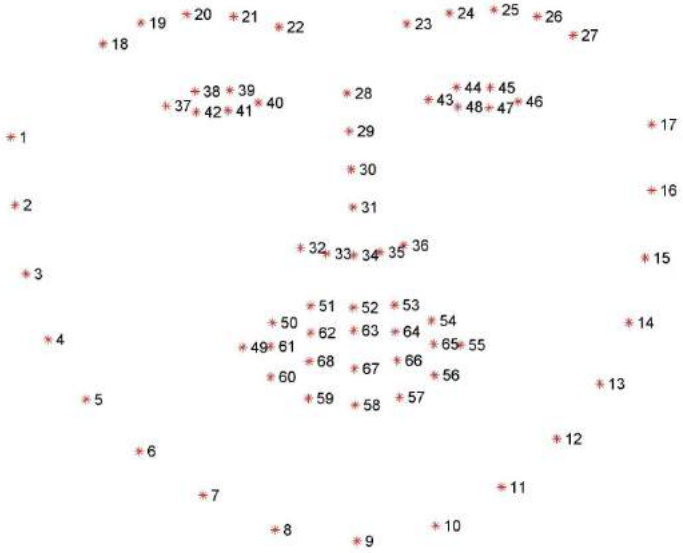


Figure 2: 68 Facial Landmarks

consecutive frames for which eyes are closed or mouth is open. Ratios are further used in the Gaze tracking functionality as well which is a little different from the drowsiness detection but geared towards detecting if the driver is attentive or not. For the gaze tracking, we identify the pupil locations and then only consider the horizontal ratios to determine if the driver is looking left or right, and not focused. Falling on either extreme of these ratios raises an alert. Apart from that, head tilt is calculated using the solvePnp API of opencv and by converting the rotation matrices to Euler angles. solvePnp basically calculates the pose estimation based on the coordinates of the key features of the face.

In addition to the methods described above, we have trained a model from scratch that will classify the driver's emotions into pre-defined categories. We believe this feature is particularly significant because various studies have shown that the driver's mood might impair his or her ability to drive safely. While developing this model, our major focus was on increasing the accuracy compared to a number of parameters. Reducing parameters is typically important to build a small CNN which will alleviate us from slow performances on hardware constrained systems especially when we explore the practicability of the project i.e., to deploy in the car. Hence, we picked Google's XCeption architecture as the base architecture.

The model which we built has the separable convolutional layers which separate the feature extraction and combination within a layer. In addition to the separable convolutional layers, the model also integrates the residual layers which reduce the depth complexity of the model by skipping the layers. And, finally, we have used a global average pooling. Categorical cross-entropy is the loss function we have used for this model. The categorical cross-entropy for M classes is given by:

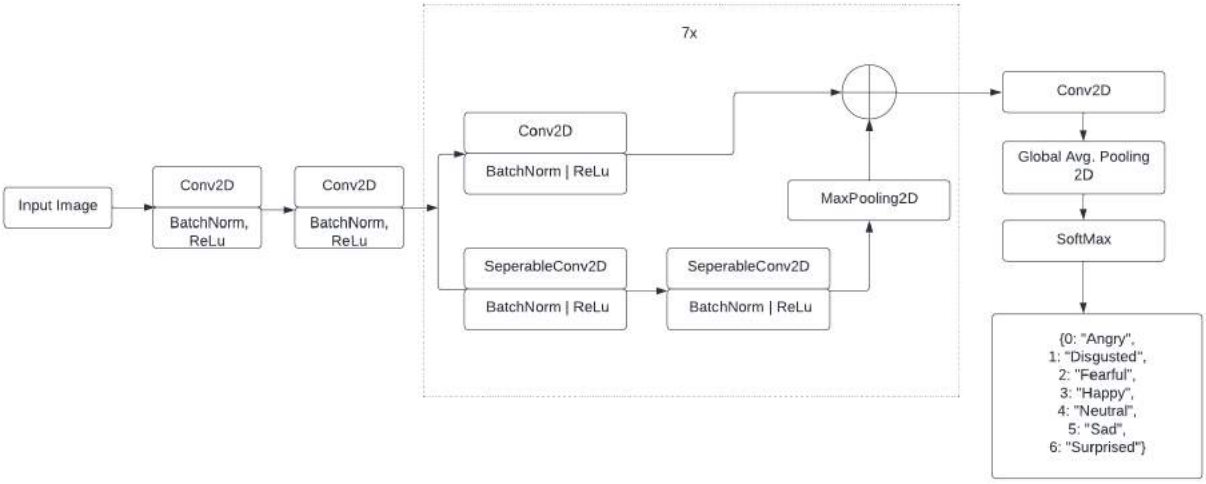


Figure 3: Emotion Recognition Model Architecture

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

Figure 3 shows the summarized architecture of the model we built. We have used FER-2013 dataset to train the model which contains near about 28709 train images and 7178 test images, labelled into seven classes namely Angry, Surprise, Sad, Neutral, Disgust, Fear, and Happy.

3.3. VISUALIZATION

This is the system's last block, significant component or difficult work in this module is to annotate the frame with bounding boxes and textual information detailing the outputs of the detection/algorithms' block. The goal of this module is to demonstrate the system and develop credibility toward the system. For the time being, we have utilized simple OpenCV APIs to render bounding boxes and textual information on the frames captured. This further can be extended to build a more sophisticated user experience with modern UX frameworks, AR engines, and gamified experience can be built. Furthermore, alerts generated are also integrated into this block, we have kept a couple of static mp3 audios in the resources of our project, and used a playsound library of python to play these audios whenever the driver is sleepy or yawning.

3.4. DEMONSTRATION AND EVALUATION

As described in a few of the above sections, we have integrated all of the above features into a single User Interface (UI) that will display a short overview of each reading, ratios, driver emotions, eye/mouth's open/close status, head pose, and so on computed on each frame. Furthermore, UI highlights if the driver is found to be distracted or drowsy. Aside from the UI, a couple of alarms are generated depending on the severity of the driver's inattention.

From the evaluation perspective, we have demonstrated the accuracy and loss over the epochs for the emotion recognition model which we have built. Aside from that, there are no

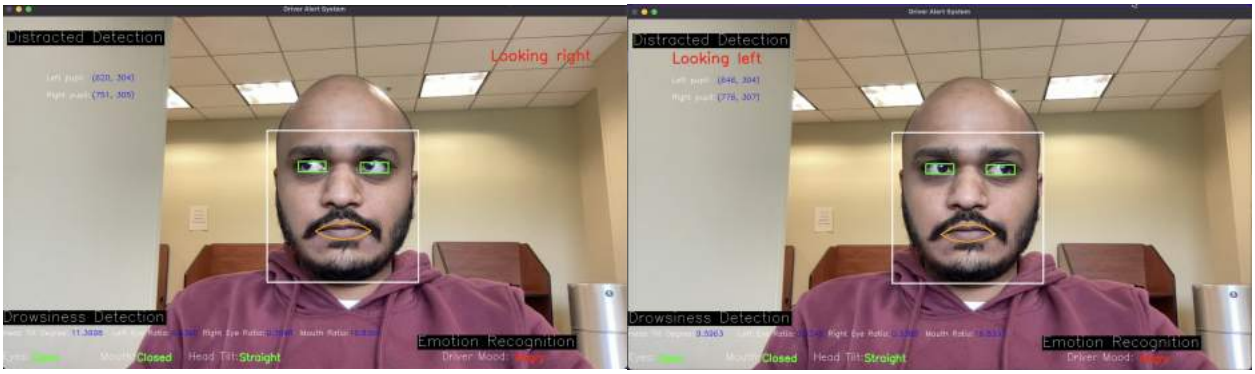
standard Key Performance Indicators (KPIs) for the ratio based algorithms, however, use case demonstrations where the system works and fails will be demonstrated in the final report. Additionally, We have investigated how the integration of multiple features influences these results. Another consideration in quantitative implementation is that when we combine many features, serial processing of frames would certainly impede the system's real-time responsiveness, and the frame rate will decrease. We will exhibit this performance analysis concerning frame rate as well as the effectiveness of the system in the final report.

4. RESULTS

Results of the driver alert system can be observed in Figure 4. All of our features: Drowsiness Detection, Distracted Detection, and Driver Mood Recognition have been fully integrated in to driver alert system.

Figure 4: Driver Alert System User Interface

Figure 5a and Figure 5b shows the gaze tracking feature of the system. The actual gaze and text seems reversed as these are the mirror images on the UI, however, left and right gaze of the eyes are correctly detected, and can be observed from the figures 5a and 5b.



(a) Looking Right *(b) Looking Left*

Figure 5: Distracted Driver

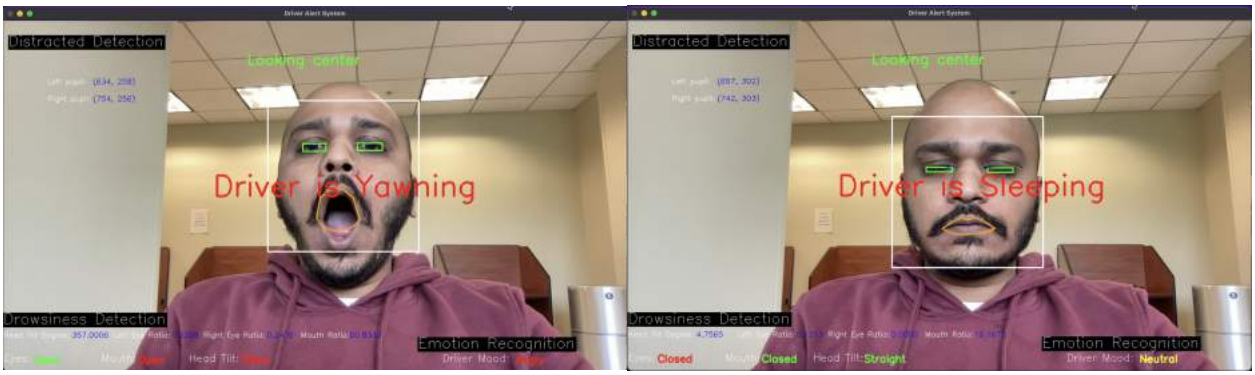
Figures 6a, 6b, and 6c demonstrates the eye and mouth open/closeness and head tilt which

are important aspects for the drowsiness detection part of the system. The white and blue textual information shows the actual readings, and ratios calculated by the detection and algorithm block of the system. These ratios determine the status open/close and tilted/straight of the eye, mouth, and head respectively. As can be seen in those figures, these status are updated correctly, if the mouth is open for few consecutive frames, it will lead to the yawning, and if eyes are closed it might lead to sleepy driver, similarly tilted head is not recommended for the alert driver, and hence, those are marked in red.

(a) Tilted Head *(b) Open Mouth* *(c) Closed Eyes*

Figure 6: Drowsiness Detection: Eyes, Mouth, Head Status

As described earlier, if the driver's eyes are closed for few consecutive frames i.e. beyond, an alarm for sleepy driver is raised. And, if the driver's mouth is open for few consecutive frames an alarm for the yawning is raised. This are voice alerts and can not be displayed in this document, however, the textual notifications for the same can be observed from Figures 7a and 7b.



(a) Driver is Yawning *(b) Sleepy Driver*

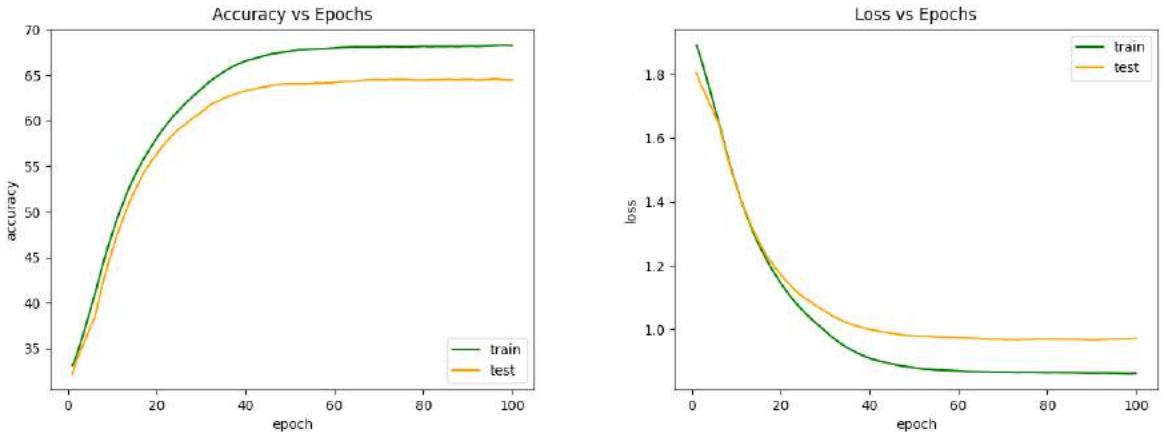
Figure 7: Drowsiness Detection

Also, the feature of emotion recognition of a driver can be observed from Figure 8 and all other figures. Figure 8 specifically shows the happy driver which is different from the other images presented in this result. Other images also can be observed where most of the time driver is neutral, however, there are instances where driver might be wrongly classified into angry as opposed to the neutral, and that is mostly caused because of the imbalance of training data.

Figure 9a and Figure 9b shows the accuracy and loss respectively of the emotion recognition model over 100 epochs with a learning rate of 0.001. We have tried number of different configurations and hyperparameters during the training, and were able to achieve approximately 69% of training and 64% of test accuracy. It might seem that the accuracy is very low, however,

Figure 8: Happy Driver

there was a competition ran on Kaggle to build a classifier based on this dataset, and the winner of the competition achieved near about 71% accuracy. Considering that and the architecture we have adopted by reducing the number of trainable parameters to improve the speed, we have achieved a decent score with three hours of training using default Neural Cores of the Apple Silicon M1 system.



(a) Training and testing accuracies *(b) Training and testing loss*

Figure 9: Evaluation of Emotion Recognition model

Figure 10 shows the improvements achieved by introducing multi-threading in the system for features and alert system. It can be observed that multi-threading improved the FPS by almost 1.6 times, this can be further improved by different hardware and software accelerators based on what is available in the real time setting.

We further explored our system for the different use cases described in table 1. It can be observed that system is independent of genders, however, there are some scenarios where the system fails, and it is expected since, the system relies on the front face pictures captured through

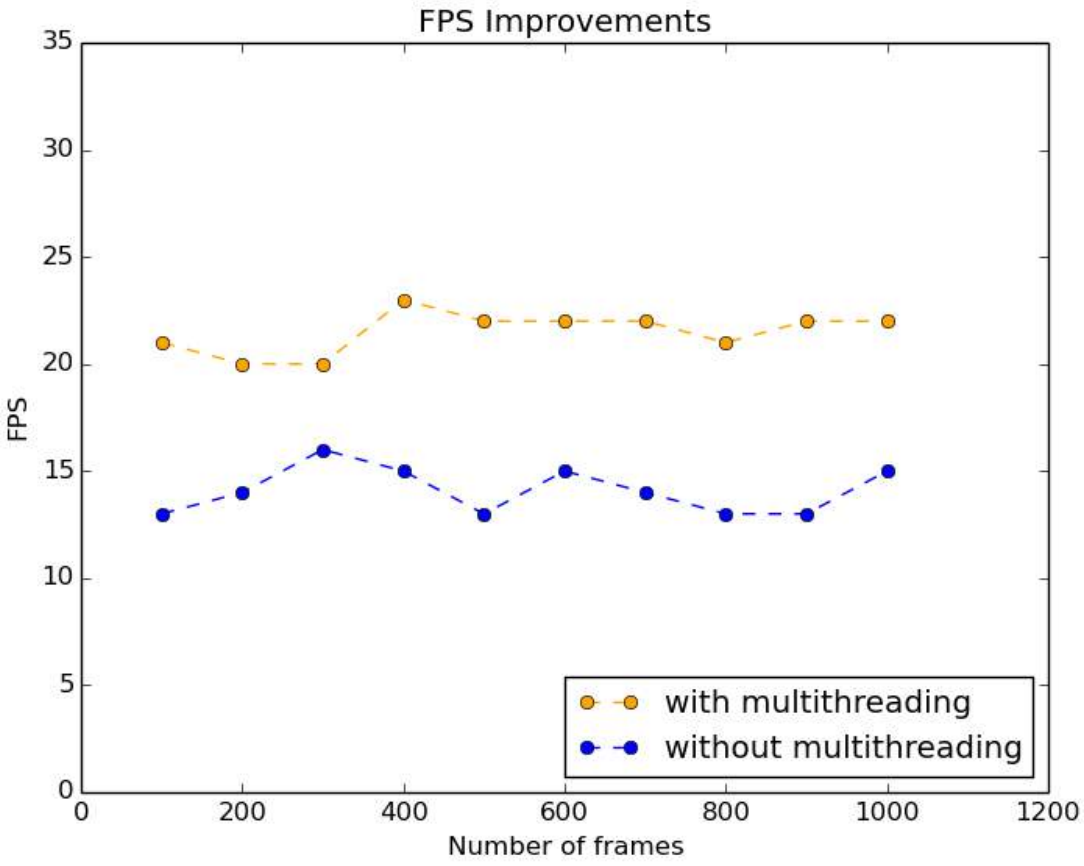


Figure 10: Evaluation of FPS Improvements

cameras. We observed that the imbalance in training data as well inability to identify certain facial features are the main reasons of the failure scenarios. It'd be interesting study to improve the system in order to integrate these scenarios. To test out these scenarios, we requested our friends, and in some cases like Man with turban, we couldn't test system the comprehensively as we held photographs in front of camera to test the system.

Usecase	**Gaze Tracking**	**Yawning**	**Sleepiness**	**Head Tilt**	**Emotion Recognition**
Women	Works	Works	Works	Works	Works
Man without beard	Works	Works	Works	Works	Works
Man with beard	Works	Works	Works	Works	Works
Women with cap	Works	Works	Works	Works	Works
Man with cap	Works	Works	Works	Works	Works
Man with turban	Not tested	Not tested	Works	Not tested	Works
Man with tinted sunglasses	Fails	Works	Fails	Works	Low accuracy
Women with tinted sunglasses	Fails	Works	Fails	Works	Low accuracy
Women with transparent sunglasses	Very Low Accuracy	Works	Low accuracy	Works	Medium accuracy
Man with transparent sunglasses	Very Low Accuracy	Works	Low accuracy	Works	Medium accuracy
Man with mask	Fails	Fails	Fails	Fails	Fails
Women with mask	Fails	Fails	Fails	Fails	Fails

Table 1: Evaluation of system with different use-cases

5. CONCLUSION

Distracted driving is a problem that leads to an alarming number of accidents globally. In addition to other initiatives to address this issue, we believe that smart vehicles would contribute to a safer driving experience. In this paper, we presented a vision-based driver alert system that recognizes distracted and drowsy drivers and alerts them. We fused multiple algorithms like eye-blinking rate, eye gaze tracking, head tilt, and mouth openness in a single system. In addition, we developed a CNN classifier that recognizes cognitive distractions by classifying emotions from

the driver's facial expressions. Our best model utilizes Google XCeption architecture to achieve a 69% accuracy. We illustrated all of the features and how they operate in the results, as well as the instances in which this system fails. Finally, we created a threaded system that alerts the driver. In a real-time situation, threaded implementation reduces performance overhead. As a result, the system presented in this paper is a real-time system with high operating speed as well as significant accuracy and reliability.

6. DISCUSSION

The driver alert system presented in this paper incorporates numerous algorithms, however the majority of these methods rely on facial landmark ratios. We haven't thoroughly tested these algorithms on drivers of various ethnicities and with varying face characteristics. Thresholds employed in the system may require changes or some form of mechanism to modify these thresholds adaptively. Furthermore, there was an imbalance in the data, which limited the accuracy of the emotion detection model to 69%; alternative datasets might be studied to enhance the accuracy. Furthermore, we were unable to finish activity monitor for integration into this system. We envision having robust activity monitor is very useful feature for this kind of system. In addition to the features implemented and proposed in this paper, this study can be extended and evaluated to incorporate a variety of other ideas, such as how the system performs in tunnel when auto exposure of cameras is activated, or extending the system for smart seat adjustment based on recorded faces or commercial infotainment control via gestures, and so on.

REFERENCES

Abouelnaga, Y., Eraqi, H. M., and Moustafa, M. N. (2017). Real-time distracted driver posture classification. *CoRR*, abs/1706.09498.

Ahlstrom, C., Kircher, K., and Kircher, A. (2013). A gaze-based driver distraction warning system and its effect on visual behavior. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):965–973.

Coleman (2022). Distracted driving statistics 2022. https://www.bankrate.com/insurance/car/distracted-driving-statistics/.

Deng, W. and Wu, R. (2019). Real-time driver-drowsiness detection system using facial features. *IEEE Access*, 7:118727–118738.

Devi, M. S. and Bajaj, P. R. (2008). Driver fatigue detection based on eye tracking. In *Proceedings of the 2008 First International Conference on Emerging Trends in Engineering and Technology*, ICETET '08, page 649–652, USA. IEEE Computer Society.

Dua, H. K., Goel, S., and Sharma, V. (2018). Drowsiness detection and alert system. *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pages 621–624.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J.,

Xie, J., Romaszko, L., Xu, B., Chuang, Z., and Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In *Neural Information Processing*, pages 117–124, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sagonas, C., Antonakos, E., Tzimiropoulos, G., Zafeiriou, S., and Pantic, M. (2016). 300 faces in-the-wild challenge: database and results. *Image and Vision Computing*, 47:3–18.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., and Pantic, M. (2013a). 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *2013 IEEE International Conference on Computer Vision Workshops*, pages 397–403.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., and Pantic, M. (2013b). A semi-automatic methodology for facial landmark annotation. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 896–903.

Wu, Y.-L., Tsai, H.-Y., Huang, Y.-C., and Chen, B.-H. (2018). Accurate emotion recognition for driving risk prevention in driver monitoring system. In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pages 796–797.

Xing, Y., Lv, C., Wang, H., Cao, D., Velenis, E., and Wang, F.-Y. (2019). Driver activity recognition for intelligent vehicles: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 68(6):5379–5390.